Statistical Engineering

Statistical Engineering

An Algorithm for Reducing Variation
in Manufacturing Processes

Stefan H. Steiner and R. Jock MacKay

ASQ Quality Press
Milwaukee, Wisconsin

American Society for Quality, Quality Press, Milwaukee 53203

Printed in the United States of America

12 11 10 09 08 07 06 05 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Steiner, Stefan H., 1964–
Statistical engineering : an algorithm for reducing variation in manufacturing processes / by Stefan H. Steiner and R. Jock MacKay.
p. cm.
Includes bibliographical references.
ISBN 0-87389-646-7
1. Production management. 2. Manufacturing processes. I. MacKay, R. Jock. II. Title.

TS155.S773 2005
658.5—dc22

2004030676

Publisher: William A. Tony
Acquisitions Editor: Annemieke Hytinen
Project Editor: Paul O'Mara
Production Administrator: Randall Benson

ASQ Mission: The American Society for Quality advances individual, organizational, and community excellence worldwide through learning, quality improvement, and knowledge exchange.

Attention Bookstores, Wholesalers, Schools, and Corporations: ASQ Quality Press books, videotapes, audiotapes, and software are available at quantity discounts with bulk purchases for business, educational, or instructional use. For information, please contact ASQ Quality Press at 800-248-1946, or write to ASQ Quality Press, P.O. Box 3005, Milwaukee, WI 53201-3005.

To place orders or to request a free copy of the ASQ Quality Press Publications Catalog, including ASQ membership information, call 800-248-1946. Visit our Web site at www.asq.org or http://qualitypress.asq.org.

ASQ
AMERICAN SOCIETY FOR QUALITY™

Quality Press
600 N. Plankinton Avenue
Milwaukee, Wisconsin 53203
Call toll free 800-248-1946
Fax 414-272-1734
www.asq.org
http://qualitypress.asq.org
http://standardsgroup.asq.org
E-mail: authors@asq.org

Printed on acid-free paper

To Anne Marie, Erik, and Emily—S.H.S

To Samm—R.J.M.

Contents

Material on Enclosed CD-ROM

Figures and Tables

Acknowledgments

This book is the product of many years' experience helping manufacturing organizations reduce variation. We want to thank all of our clients and colleagues at the Institute for Improvement in Quality and Productivity (IIQP). The IIQP is a not-for-profit organization, started in 1985, committed to the development, communication, and application of methods for quality and productivity improvement.

Former and current IIQP colleagues include Bovas Abraham, Dennis Beecroft, Jerry Lawless, Jack Robinson, Jim Whitney, and Clif Young, among others. The numerous corporate members of the institute over the years included AT Plastics, Bell Canada, BF Goodrich Canada, BUDD Canada, Campbell Soup Company, Canada Post Corporation, C-I-L, Continental Can Canada, Dofasco, Epton Industries, GSW Water Products Company, General Motors Canada, Lear Seating Canada, Lear Seigler Industries, Metal Koting Continuous Colour Coat, Nortel Networks, Quebec and Ontario Paper Co., Research in Motion (RIM), Stelco Steel, The Woodbridge Group of Companies, Uniroyal Goodrich Canada, Wescast Industries, and Xerox Canada. Thanks also to other companies we have worked with, including Fisher and Paykel, Focus Automation Systems, Ford Motor Company, Imperial Oil of Canada, Seagram Americas, Stackpole, Toyota Motor Manufacturing Canada, and many others. We learned from you all.

We have also benefited from helping to organize and attending the Continuous Improvement Symposiums held by General Motors of Canada each year since 1987. The symposiums provide an opportunity for employees and suppliers to showcase process improvement success stories. Some of the examples in the book come from these symposiums.

We want to single out Mike Brajac, Pete Peters, and Mark Smith at General Motors of Canada, with whom we have had a long-standing relationship. You are the best problem solvers we know. We also thank Wescast Industries and especially Harry Schiestel, who taught us a lot about foundries and how to solve problems.

We have had outstanding relationships with our colleagues at the University of Waterloo who have contributed, perhaps unknowingly, in many ways to the book. We especially thank Winston Cherry and Wayne Oldford for their penetrating questions and stimulating discussions, and Dennis Beecroft, who taught us there is more to variation reduction than statistics.

Also, our appreciation goes to the undergraduate and graduate engineering and statistics students taking the inaugural offering of the University of Waterloo course STAT 435/835, Statistical Methods for Process Improvement, in the Winter 2004 term. The students used a draft of the book and provided many helpful comments and suggestions to improve the presentation.

In the early stages of writing this book, we were fortunate to be on sabbatical. During fall 2003, Jock received support from CSIRO in Melbourne, Australia, with thanks to Richard Jarrett and Geoff Robinson. Stefan spent a year starting in May 2002 at the University of Auckland, New Zealand, with thanks to Chris Wild.

We were guided through production by the staff of ASQ Quality Press and Kinetic Publishing Services. Special thanks to Paul O'Mara, Laura Varela, and John Ferguson.

Last, but certainly not least, we would like to acknowledge the wonderful support from our families over the long road required for writing this book.

Preface

Reducing the variation in process outputs is a key part of process improvement. If you have picked up this book, you probably do not need to be convinced of the truth of this statement. For mass-produced components and assemblies, reducing variation can simultaneously reduce overall cost, improve function, and increase customer satisfaction with the product. Excess variation can have dire consequences, leading to scrap and rework, the need for added inspection, customer returns, impairment of function, and a reduction in reliability and durability.

We have structured the book around an algorithm for reducing process variation that we call *Statistical Engineering.* The algorithm is designed to solve chronic problems on existing high- to medium-volume manufacturing and assembly processes. The algorithm will not apply to urgent, short-term sporadic problems such as what to do when a defective item is found. Instead, we look at the problem of reducing the frequency of such defective items.

The fundamental basis for the algorithm is the belief that we will discover cost-effective changes to the process that will reduce variation if we increase our knowledge of how and why a process behaves as it does. A key way to increase process knowledge is to learn empirically—that is, to learn by observation and experimentation. We discuss in detail a framework for planning and analyzing empirical investigations, known by its acronym QPDAC (Question, Plan, Data, Analysis, Conclusion). We use the QPDAC framework at many stages of the Statistical Engineering algorithm to help plan, analyze, and interpret the results of appropriate investigations.

Using the algorithm, you are guided through a series of empirical investigations to a cost-effective solution to the problem. The purpose and plan for each investigation depends on:

- The stage of the algorithm
- The accrued knowledge from earlier investigations
- Other engineering and process knowledge

We classify all effective ways to reduce variation into seven approaches. A unique aspect of the algorithm forces early consideration of the feasibility of each of the approaches. Selecting a working approach helps generate effective and efficient solutions. The choice of approach affects the process knowledge required and hence how we proceed.

Some of the variation reduction approaches (but not all) require knowledge of a dominant cause of variation. We present a low-cost strategy for finding a dominant cause based on families of causes and the method of elimination. The method of elimination uses a series of simple investigations, each of which is designed to eliminate a large number of possibilities from those remaining.

We illustrate all aspects of the algorithm with many examples adapted from our experience. In some cases, we have disguised the data to protect confidentiality; in others, we have taken some liberties with what actually happened to make a point.

Throughout the book, we use the statistical software package MINITAB for all calculations and most displays. To apply the Statistical Engineering algorithm, the user requires a software package capable of making basic plots and finding simple numerical summaries. By referring to the procedures in the software, we avoid algebraic expressions for the most part. We describe the calculations with words and let the software deal with the numerical implementation.

We avoid formal or complicated statistical analysis procedures. Whenever possible, we use graphical displays to guide us to the correct interpretation of the data. We assume that the reader has been exposed to basic statistical concepts and tools, such as standard deviations, averages, histograms, run charts, box plots, scatter plots, process maps, and flowcharts. We provide detailed explanations of more sophisticated analysis tools as needed, including multivari charts, analysis of variance (ANOVA), regression, and designed experiments. We include appendices to explain how to use MINITAB to produce the analysis for all of the methods discussed.

Important issues surround the management of process improvement projects. These are the same issues that arise in any project management exercise. Priorities must be set, plans and schedules made, resources provided, and so on. We do not deal with these issues in detail; rather, we focus on the algorithm, the variation reduction approaches, and the tools required to achieve variation reduction.

The Statistical Engineering algorithm is not meant to replace global improvement systems such as Six Sigma. It is focused on and designed for process improvement in high- to medium-volume manufacturing processes. We suggest that the algorithm, strategies, and methods be incorporated into a general improvement system and used where appropriate.

TARGET AUDIENCE

The primary audience of this book is people who are involved in the improvement of manufacturing processes. They include:

- Process engineers with responsibility for reducing variation, decreasing costs, improving quality, and so on
- Six Sigma Green Belts, Black Belts, and Master Black Belts
- Trainers in process improvement methods
- Academics and students interested in quality and productivity improvement
- Teachers and students of courses in engineering statistics

WHY THIS BOOK IS NEEDED AND HOW IT WILL BENEFIT THE READER

This book is unusual because it focuses directly on the goal of variation reduction through the use of the Statistical Engineering algorithm and its associated approaches. The book is not a collection of statistical analysis tools and methods useful in achieving the goal. Having a lot of tools at your disposal does not help if you do not have a good idea what to do next or when to use a particular tool.

This book will benefit the reader in many ways. In particular, it will help you learn:

- A structured way to address variation reduction problems through a series of process investigations, each depending on what has been learned to that point
- The seven variation reduction approaches
- How to conduct empirical investigations in a sound manner to get reliable conclusions
- The method of elimination for finding a dominant cause of variation
- The appropriate use of statistical tools to support the structured algorithm
- How to assess feasibility and implement each of the seven variation reduction approaches

We present numerous examples and three case studies to convince you that the algorithm works.

HOW TO USE THIS BOOK EFFECTIVELY

While reading the text, we suggest you start an improvement project, or at least think about your own process problems. The more analogies you can draw between our examples and your processes, the better you will understand the material.

We advise the reader to try the exercises and explore the data sets to help gain confidence in the use of the approaches and methods.

STRUCTURE OF THE BOOK

This book is structured around the Statistical Engineering variation reduction algorithm. To limit length and cost, we have divided the book into two parts. In the printed text, we present the algorithm and the concepts and tools needed to use it effectively. On the enclosed CD-ROM, we provide chapter supplements, case studies, exercises, data sets, and appendices. In the chapter supplements, we give more technical details, discuss some important complications and competing methods, and give references for further reading. The printed text can stand alone, but we believe you will find the supplements helpful.

There are four major parts in the printed text:

Part I: Setting the Stage—We start by introducing the language of processes, such as outputs, fixed and varying inputs, dominant cause, and so on. Then we present the variation reduction algorithm, the seven variation reduction approaches, and a framework (QPDAC) for learning empirically by investigating the process.

Part II: Getting Started—We look at how to focus, define, and quantify a problem. We also look at methods for assessing the measurement system and for choosing a working variation reduction approach to guide further investigations.

Part III: Finding a Dominant Cause of Variation—We describe the method of elimination for finding a dominant cause that uses families of causes of variation. We provide a number of investigation plans and analysis methods to help eliminate possible causes. We introduce experimental plans to verify that we have found a dominant cause.

Part IV: Assessing Feasibility and Implementing a Variation Reduction Approach—We return to the choice of variation reduction approach in light of the results of a search for the dominant cause or a decision to skip such a search. In separate chapters, we discuss assessing the feasibility of and implementing each approach. We finally consider validating the solution and look at methods for preserving the gains.

The enclosed CD-ROM contains:

- The chapter supplements
- Three case studies
- Exercises with solutions
- All data from the examples and exercises (in both Microsoft Excel and MINITAB worksheet format)
- Appendices that will help you use MINITAB

In summary, the outstanding features of the book are:

- A structured algorithm for reducing variation in processes
- A classification of potential solutions into seven variation reduction approaches
- An emphasis on planning for data collection and simple analysis methods
- Use of the method of elimination to economically find a dominant cause of variation
- Many examples (with more than 100 datasets available) to illustrate all stages of the algorithm
- Separation of the "how to" (main text) from the supplementary material (CD-ROM)

- Demonstration of the use of MINITAB to help with the implementation of statistical tools, allowing greater focus on the interpretation of the data
- More than 65 exercises designed to reinforce the ideas and tools

We encourage readers to send us feedback regarding their use of Statistical Engineering and the related tools and methods. Our e-mail addresses are shsteiner@uwaterloo.ca and rjmackay@uwaterloo.ca.

Stefan Steiner and Jock MacKay
Waterloo, Ontario, Canada
November 2004

1

Introduction

Problems are only opportunities in work clothes.

—Henry J. Kaiser, 1882–1967

This book presents a systematic algorithm for reducing variation. The algorithm is tailored to high- to medium-volume manufacturing processes where it is feasible to measure the values of selected process inputs and outputs.

We use the word *variation* to mean both the deviation of the output from a target value and the changing value of the output from part to part. For example, in a machining process that produces V6 pistons, the target value for the diameter is 101.591 millimeters. The measured diameter in millimeters of three successive pistons is

101.593, 101.589, 101.597

We can see variation in both senses since none of the pistons has the target diameter and all have different diameters. We will formulate problems by defining appropriate performance measures that capture the nature of the variation that we want to reduce. Excessive variation leads to poor performance, low customer satisfaction, scrap and rework, complex downstream control plans, and so on. If we can resolve such problems, we can reduce costs and improve quality and performance.

The fundamental basis for the algorithm is our belief that *by increasing knowledge of how and why a process behaves as it does, we will discover cost-effective changes to the process that will reduce variation.* One way to increase process knowledge is to learn empirically, that is, to learn by observation and experimentation. Statistics is the discipline that teaches us how to learn empirically. Statistics provides the answers to questions such as "How should we plan our process investigation?" and "How do we interpret the data that we have?" The algorithm we propose relies heavily on statistical methods and tools combined with existing engineering knowledge and theory. Using the algorithm, we will plan and carry out one or more investigations to learn about process behavior.

In most cases, to be cost-effective, the proposed changes involve better process control or alterations to process settings rather than fundamental design changes or replacing process equipment. We have classified these low-cost changes into a set of generic variation reduction approaches and have structured the algorithm to force early consideration of an approach.

The specific objectives of the book are to help you to:

- Think strategically about how to achieve cost-effective variation reduction.
- Reduce variation by following a step-by-step algorithm.
- Understand sources of variation and their role in process improvement.
- Learn how to better use empirical methods; that is, learn effective and efficient ways to plan, execute, and analyze the results of a process investigation.

The purpose of this chapter is to provide examples of the types of problems we can address using the proposed algorithm. Here we discuss the problems only, not the path to their solution. We revisit these examples later in the book. We hope these problems will motivate you to read further. If you can draw analogies between your own processes and problems and those described, then we are confident that you can achieve great benefits by applying the algorithm, approaches, and methods found in the rest of the book.

1.1 TRUCK PULL

Front wheel alignment on light trucks is a set of characteristics that affect the handling of a vehicle and the life of its tires. One component of the alignment is called *pull.* Pull is an important characteristic because it indicates how well the truck will track on a standard highway. A driver can feel pull—a value close to target will produce a more drivable vehicle.

Pull is a torque, measured in Newton-meters, and is a function of right and left front wheel camber and caster angles. For the vehicles discussed here, the relation between pull and the alignment angles is

$$\text{Pull} = 0.23*(\text{right caster} - \text{left caster}) + 0.13*(\text{right camber} - \text{left camber})$$

The alignment characteristics are measured on every truck assembled. In order to improve customer satisfaction, the manufacturer decided to reduce variation in pull around the target value. The performance of the process over a two-month period at the start of the project is shown in Figure 1.1. The histogram is based on pull values from 28,258 trucks.

The histogram was created using MINITAB and the data given in the file *truck pull baseline*. You can find this file (in both a MINITAB worksheet and an Excel spreadsheet) on the enclosed CD-ROM. See the appendices for more information on using MINITAB.

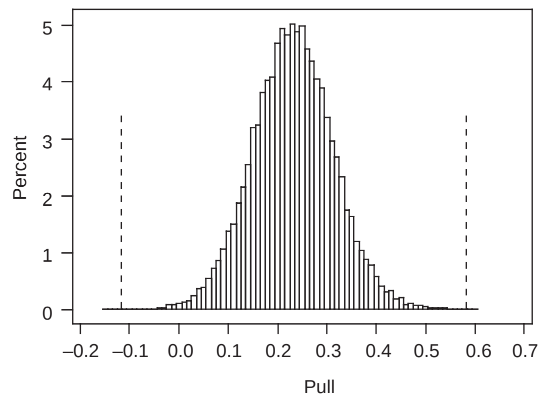


Figure 1.1 Truck pull (as found).

The dashed vertical lines on Figure 1.1 show the specification limits for pull, set at –0.12 and 0.58 Newton-meters. The target value is 0.23 Newton-meters. We see that every truck has pull within the specification limits (in fact, there are a few pull values outside these limits that are not visible in the histogram because of the large number of data points). Any truck with pull that does not meet the specifications is repaired and remeasured before shipment.

The goal of the project was to reduce the variation in pull around the target so that the histogram would look like that in Figure 1.2. If this goal can be achieved, the process will produce a greater proportion of trucks with pull close to the target value; hence, there will be greater overall customer satisfaction. As well, the proportion of trucks needing rework will be smaller, thus reducing cost.

This is a problem in reducing variation in pull from truck to truck. The process is currently centered on the target. As shown by Figure 1.1, adjusting the process center to increase (or decrease) pull on every truck will make the process worse, because then more

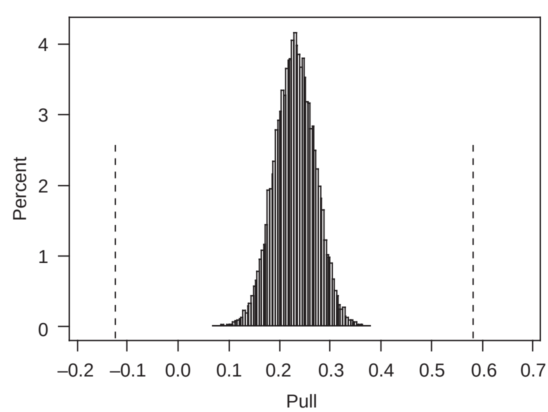


Figure 1.2 Truck pull (as desired).

trucks will have pull outside the specification limits and fewer trucks will have pull near the target value.

We have found that excessive variation from part to part is a widespread problem. For example, many measurement systems exhibit substantive variation if the same part is repeatedly measured, especially over a long period. This variation leads to poor process control, disputes between customers and suppliers, the scrapping of good parts, and so on. Often we understand enough about a process to have an easy method to keep the center of the process near the target value. We rarely have such an adjustment to reduce the variation from one part to the next.

1.2 ENGINE BLOCK LEAKS

A cast-iron foundry manufactured engine blocks, which were then machined at an engine assembly facility. The engine plant pressure tested each block for leakage in the oil and water passages and scrapped leaking blocks. The cost of the scrap, including the wasted machining, was several hundred thousand dollars per year. This cost was assigned to the foundry because it was accepted that the leaks were generated in the casting process. The foundry management established a team to reduce the frequency of block leakers. Figure 1.3 is a run chart of the proportion of leaking blocks over several months' production prior to the start of the project. The team's goal was to reduce the proportion of leakers to less than 1%.

This example is typical of many processes that generate scrap, rework, and returns due to defects. A painting process may produce visual defects such as dirt, craters, and so forth. A molding process can generate defects such as porosity, shrinks, or inclusions; an assembly process can deliver parts that fail to function; and so on. In all these examples, a part either has the defect or it does not. The target value is no defect. The goal of the project is to reduce the frequency of the defect—that is, to reduce the variation in the output by making more parts without defects.

Problems defined in terms of a binary output rather than a continuous output can be difficult because it is harder to learn about the process. In most applications, the defective rate

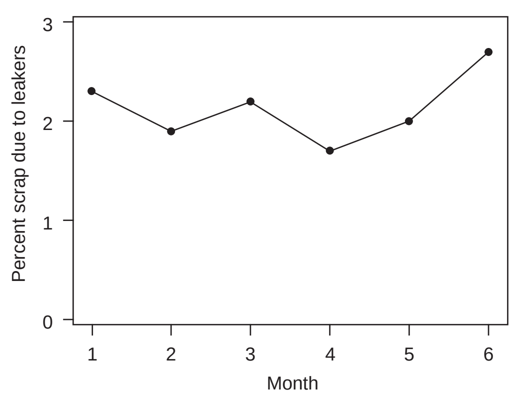


Figure 1.3 Proportion of leaking blocks versus month of production.

at the beginning of the project is likely to be low, so that investigations require large sample sizes to find a few defectives. However, we can sometimes translate a problem defined in terms of a binary output into one with a continuous output. For example, the cause of some of the leaks was the underlying wall thickness, a continuous characteristic. If an internal wall gets too thin, then a leak will occur. For these leaks, we can translate reducing the proportion of leakers into reducing variation in wall thickness about the target value. We expect to be able to learn about wall thickness variation more easily. When possible, we will translate a problem with a binary output into one with a continuous output. This is not always possible. For example, leaks due to sand inclusions cannot (easily) be directly associated with an underlying continuous characteristic, and we are forced to deal with the binary output and large sample sizes.

1.3 CAMSHAFT LOBE RUNOUT

The geometry of the lobes of a camshaft, as shown in Figure 1.4, is critical in the functioning of an engine. The rotation of the camshaft lobes drives the opening and closing of the engine valves. The displayed camshaft has 12 lobes, three of which are indicated by the white arrows.

Viewed from the side, the base of the lobe (±60° from the centerline) is ideally an arc of a circle. Figure 1.5 is a trace of the deviation from ideal, circle (in millimeters) over the base of one lobe.

Of the six measured critical characteristics related to lobe geometry, *base circle runout* was historically the most problematic and thus was chosen as the focus of a variation reduction exercise. Base circle (BC) runout is a positive measure of the deviation (maximum – minimum over the ±60° arc) of the actual lobe geometry from the ideal circle. A value of 0.000 millimeters means that the base is exactly circular. The maximum allowable BC runout is 0.040 millimeters, or 40 microns.

A sample of 108 parts, 9 per day, was collected over 12 days, and the BC runout for each of the 12 lobes was measured on each camshaft. The 1296 runout measurements (recorded in microns) and some other geometric characteristics of the lobes are available in the file *camshaft lobe runout baseline*. A histogram of the runout values over all lobes is shown in Figure 1.6.

The BC runout for all lobes was well below the specification limit, but because of the effects of this critical characteristic on engine performance, management initiated a

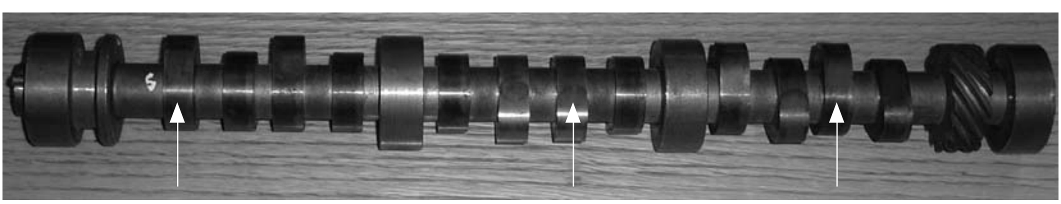


Figure 1.4 Camshaft.

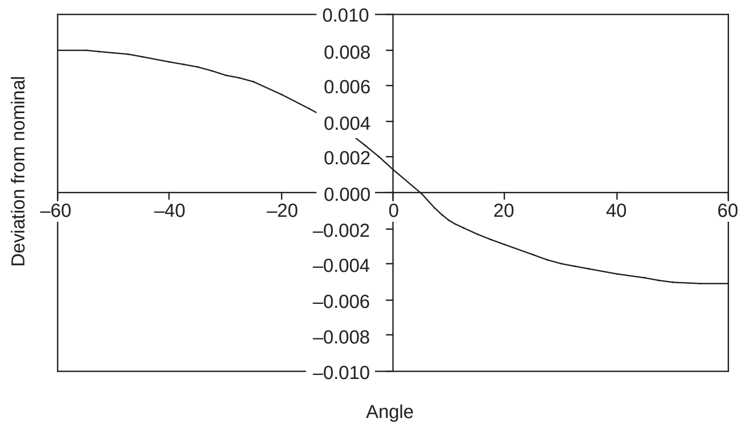


Figure 1.5 Lobe deviation from ideal.

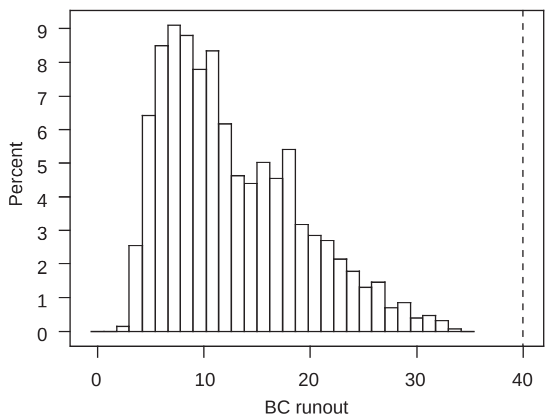


Figure 1.6 Histogram of camshaft lobe BC runout (dashed line gives the specification limit).

project to improve the process. Here the goal was to change the process so that the right tail of the histogram would be shifted to the left and a greater proportion of lobes would have runout close to zero. Because runout must be positive, adjusting the process to lower runout on all lobes was not feasible. In this context, we take variation reduction to mean that we will make a higher proportion of lobes with runout close to the target value of 0 microns.

There are many other process characteristics with a one-sided specification and a physical lower bound. Examples include flatness, porosity, taper, and so on. We can improve the process performance defined in terms of such characteristics by increasing the concentration of values near the target value. That is, we reduce the variation in the process output about the target value.

1.4 SAND CORE STRENGTH

In a cast-iron foundry, there was breakage due to handling of the molded sand cores that create the cavities within the casting. The loss of cores added cost and threatened the financial viability of the overall production of castings. The strength of cores was measured using a destructive measurement system. A sample of 100 cores (five samples spread out over a single day of five consecutive shots of the four-cavity mold) was measured to demonstrate the current process performance. The data are given in the file *sand core strength baseline*. The histograms in Figure 1.7 show the initial performance and the project goal in which the strength would be increased by three pounds on each core.

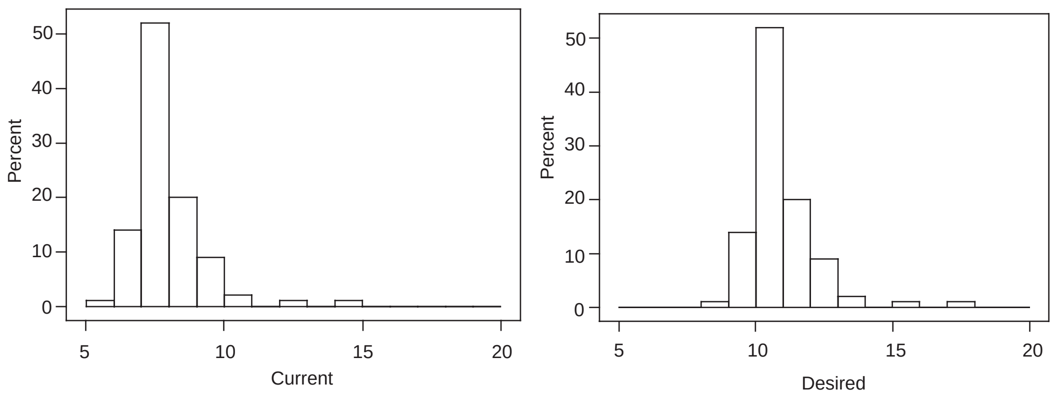


Figure 1.7 Current and initial desired core strengths (pounds).

There was no obvious target for core strength. The specification for the core was that it not be broken. The team planned to increase core strength by increasing the concentration of resin in the core-making process. On first glance, this is an example of a problem where we can improve a process by making a one-time adjustment that increases or decreases the output characteristic on all parts, that is, by changing the process center. Later, the team discovered that overly strong cores caused casting defects. They changed the goal to eliminate cores with low strengths. In other words, the new goal was to reduce variation in strength from core to core.

It is often easy to find a low-cost solution to shift the process center. The challenge is to avoid side effects—in which we replace one problem by another, as in the sand core strength example.

1.5 CRANKSHAFT MAIN DIAMETER

In a process to machine crankshaft main journals, there was excess diameter variation. The histogram of the diameter at the front position of the first main at the start of the project is shown in Figure 1.8. Note the diameter was recorded as the deviation from the target value, measured in microns. The specification limits were ±4 microns.

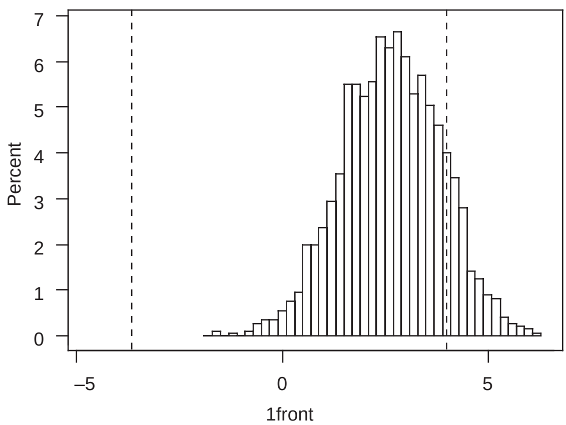


Figure 1.8 Histogram of crankshaft diameters for the front position of the first main.

The process was off target, and it seemed a simple matter to adjust the upstream grinders to reduce the average diameter closer to zero. Again, there were negative consequences. The process operators knew that crankshafts with oversized journals could be reworked but those with undersized journals were scrapped. They had deliberately centered the process above the target to avoid costly scrap. Here the goal is first to reduce piece-to-piece variation and shrink the width of the histogram. Then they could adjust the center of the process to the target without the risk of scrapped parts.

1.6 PAINT FILM BUILD

In a painting operation, there was excessive variation in film build (paint thickness) from vehicle to vehicle at particular locations. As a consequence, to meet the minimum film build specification of 15 thousandths of an inch, the target was kept well above the specification at 17. The film build of five consecutive cars was measured at five locations on the front door every hour for 16 hours. The data are given in the file *paint film build multivari*. A histogram of the film build values is given in Figure 1.9. We see that all film build measurements are above the minimum specification limit. However, running the process above target results in high paint usage and creates visual defects such as runs on occasion. The paint shop management initiated a project with the goal of reducing variation in film build. With lower variation, the film build target could be reduced, resulting in cost savings due to reduced paint usage and rework.

Variation reduction projects are often linked with productivity goals. They may involve cost reduction, as in the paint film build example. In other situations, variation reduction may lead to increased throughput. For example, by virtually eliminating the need for rework due to imbalance in a brake rotor production process, the team was able to increase the daily volume.

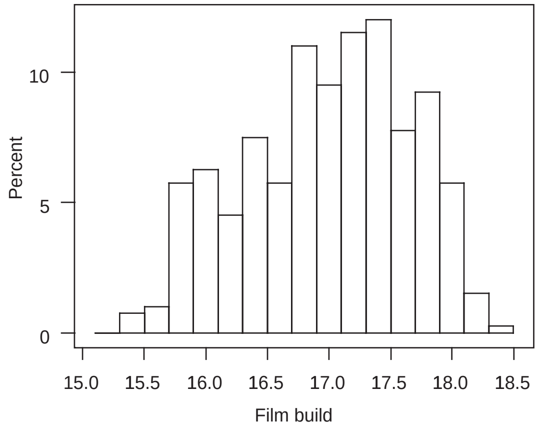


Figure 1.9 Histogram of film build.

1.7 REFRIGERATOR FROST BUILDUP

The manufacturer of a so-called frost-free refrigerator designed for temperate climates expanded its market to several tropical countries. Almost immediately, there were many customer complaints about the frost buildup inside the refrigerator. These complaints had been nonexistent in the traditional market. A team was formed to reduce the number of complaints. The manufacturer feared that the opportunity for market expansion could easily be lost.

There was no clear baseline to specify the problem because there were relatively few refrigerators in service in the new market, and there were difficulties tracking which refrigerators were in use and which were still to be sold. The goal of the team was to deliver the promised frost-free refrigerators to the customers in tropical climates, as measured by a reduction in the number of complaints.

This problem is different in several ways from those described earlier. First, at the start of the project, there was no objective way to measure the output characteristic of interest. Whether frost built up or not could be determined only during customer usage. Second, the circumstances under which the problem occurred were better understood, because there had been no similar complaints of frost buildup in the traditional temperate climate market. In fact, the team was quickly able to identify the likely causes of the problem. The ambient temperature and humidity were higher in the new markets, and also customer usage was different. Some of the new customers opened and closed the refrigerator more often than expected and also added large volumes of warm items to the refrigerator all at once. These suspected causes of frost buildup were not under the control of the manufacturer. They dominated the effects of inputs that changed during manufacturing and assembly.

Many problems have known or identified causes. In an automotive painting operation, process engineers discovered that a visual defect on the roof of the car could be eliminated by reducing the thickness (film build) of the clear coat, the last layer applied in the process.

However, there was a heavy price to pay for this supposed solution due to poorer overall appearance of the painted surface. Using the algorithm given in this book, the engineers found a solution that allowed film build to be increased without the appearance defect occurring.

Key Points

- We can describe different features of variation in the specification of a problem:

  –Part-to-part variation

  –Deviation from target

  –Defect rate
- We can generate low-cost ideas for variation reduction if we better understand how and why a process varies.
- We can gather this knowledge through a series of process investigations supplemented by existing knowledge and theory.

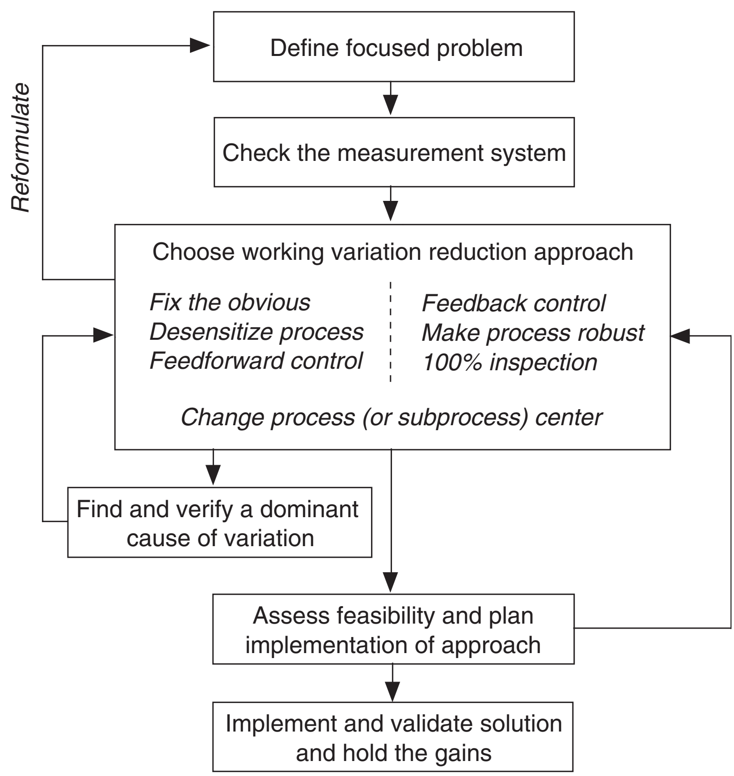
PART I

Setting the Stage

If I had to reduce my message to management to just a few words, I'd say it all had to do with reducing variation.

—W. Edwards Deming, 1900–1993

In this first part of the book, we explore the meaning and consequence of process variation, provide tools for quantifying variation in an understandable way, and discuss various sources of variation. We also present an algorithm for reducing variation, centered around seven variation reduction approaches. We explain the seven variation reduction approaches and give examples. Finally, as the key to improvement is acquiring new process knowledge, we discuss the QPDAC (Question, Plan, Data, Analysis, Conclusion) framework, used to plan and analyze process investigations.



Statistical Engineering variation reduction algorithm.

2

Describing Processes

Failure to understand variation is a central problem of management.

—Lloyd S. Nelson

This book is about how to reduce variation in process output. Increased costs and decreased customer satisfaction are two of the negative consequences of excess variation. In this chapter, we provide some common language and tools to describe processes and variation.

2.1 LANGUAGE OF PROCESSES

First, we introduce some terms to describe a process and its behavior. You will likely be familiar with these terms; we introduce them here so that we can use them without confusion throughout the book. For a running example, we look at a process used to manufacture cast iron exhaust manifolds, as shown in Figure 2.1. In the casting process, scrap iron is melted and doctored to adjust its chemistry and physical properties such as temperature.

Figure 2.1 Exhaust manifold.

Sand molds are formed to determine the external shape of the manifold. Cores are molded to create internal space in the casting. A core is placed in the mold and the molten iron is poured. After cooling, the sand is shaken out, and the result is a rough casting. The casting is finished by machining various surfaces and drilling holes. Throughout, operators make measurements and process adjustments. They also inspect the castings for defects at several points.

A process can be divided into *subprocesses* and is almost always a part of a larger process. In the casting example, the melting of the iron, the creation of the mold, and the core-making are all examples of subprocesses. The manufacturing process for the manifold sits inside a system that includes the design process for both the part and the manufacturing process; the sales, order, and billing processes; the delivery process; and so on.

A *process map* or *flowchart* is a good tool to describe a process, especially if we choose an appropriate level of detail. For example, we can represent the major subprocesses or operations within the casting process by the simple flowchart in Figure 2.2.

This flowchart shows the major subprocesses and the order in which they occur. We can describe the subprocesses in much finer detail if we choose. The flowchart clarifies which operations in the process are parallel and which are sequential. The chart also shows the boundaries that we have selected for the process. We could have chosen to include more upstream or downstream subprocesses as part of the manufacturing process. The selection of process boundaries, like the choice of the level of detail, is driven by a tradeoff between presenting facts and providing information. Too much detail can obscure the information in the chart. We take these issues into consideration when designing a useful flowchart. See Harrington (1987) for a detailed discussion of chart selection, construction, and use.

One important feature of a process is that it is repeatable; each time it operates, the process produces a *unit.* In the example, each finished casting is a unit. For a manufacturing process, we can think of a unit as a part. For a measurement process, a unit is the act of making a measurement.

We use the word *characteristic* to describe a feature or quality of a unit. For the casting process, characteristics of a manifold include its hardness, its dimensions, the pouring temperature of the iron, the properties of the sand in the cores used for that casting, and so on. The *customers* are the people or organizations that use the process units. Here, the customers include the assemblers of the engine and the ultimate users of the vehicles that

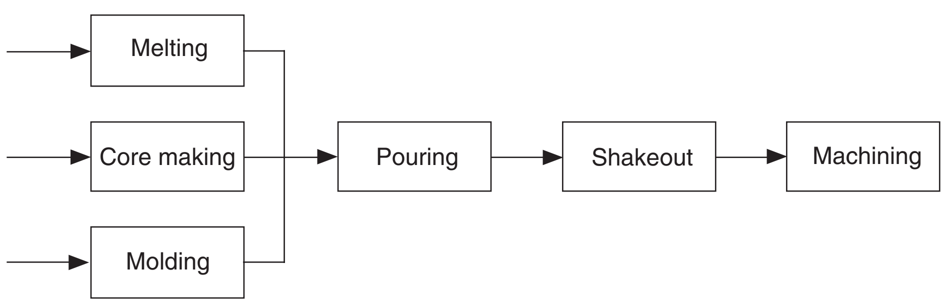


Figure 2.2 Exhaust manifold process map.

incorporate the engines. *Output characteristics,* or more briefly *outputs,* are characteristics of interest to customers. The properties of the casting that describe its performance in the engine and ease of assembly in the engine plant are outputs. We call characteristics of the casting such as the pouring temperature of the iron and the properties of the sand cores *input characteristics,* or *inputs* for short. We also classify characteristics in terms of the values they can take.

Classification	**Description**	**Examples**
binary	only two possible values	defective or not, high or low
discrete	integer values	number of defects
ordinal	ordered categories	score: low, medium, high
categorical	unordered categories	supplier: A, B, C
continuous	a range of real numbers	length, hardness, time

The *suppliers* are the people or organizations that control the subprocesses that determine many of the inputs. In the casting process, some suppliers are the providers of the sand for the cores and other raw materials, the equipment manufacturers, and so on. To complete the picture, we call the people involved in the process its *owners.* In the example, the owners are the process operators and managers. The owners are interested in the cost associated with each unit as well as many other input and output characteristics.

We classify input characteristics as *fixed* or *varying.* An input is fixed if it changes only when we deliberately make it change. For example, the target value for the pouring temperature of the iron is a fixed input. The instruction to the operator to sample five parts every hour, measure the hole locations, and make an appropriate adjustment is a fixed input. The fixed inputs may be changed by the process owners. An input is varying if its value changes from unit to unit or time to time without deliberate intervention. For example, the dimensions of the cores change from casting to casting. Other varying inputs, such as pouring temperature and raw material characteristics, change more slowly over time.

Note that the target value or set point for pouring temperature is a fixed input, but the actual pouring temperature is a varying input. We need to keep this idea clear because sometimes inputs that we consider fixed actually vary. For example, two operators may interpret the same set of instructions in different ways. The instructions are a fixed input, but the implementation of the instructions is a varying input.

2.2 CAUSES OF VARIATION

To reduce variation, we must change one or more fixed inputs of the process. In Chapter 3, we consider some approaches to variation reduction in which we first identify a dominant cause of the variation and then change fixed inputs to reduce the effects of these causes. In this section, we make clear this notion of a dominant cause. We continue with the casting and machining example to illustrate the points.

In Chapter 1, we saw that most problems involved reducing the unit-to-unit variation. We define a cause with respect to this kind of variation.

A *cause of variation in a process output* is a varying input with the property that if all other (varying) inputs were held constant, then the output changes when the input changes. Part of this definition is conceptual since it is not possible to hold all other inputs constant for all units.

In the example, suppose the output is the hardness of the casting measured at a particular location. Hardness varies from casting to casting. The concentration of carbon in the iron when it is poured into the mold is a cause of hardness variation. When this concentration changes, all other inputs being constant, the hardness of the casting will change. There are many causes of hardness variation.

Now for our first controversial statement. A fixed input cannot be a cause of variation. For example, the design of the product or process is not the cause of variation since the design is a fixed input. Since we are interested in problems defined in terms of variation, we will never say that the design is the cause of the problem. As you will see, we will change one or more fixed inputs to solve the problem, but these are not the causes.

We use a simple mathematical model to describe a cause by specifying the values of all varying inputs (again, this is only conceptually possible for a real process) for any one unit produced by the process. We write the functional model

$$output = f\,(input1,\ input2,\ ...)$$

With this model, *input1,* for example, is a cause if the *output* changes when the value of *input1* changes while all other inputs are held constant. Note that the function f ( ) depends implicitly on the values of the fixed inputs.

The model is useful because it helps us understand what is meant by the effect of a cause. The *effect* of the cause (or varying input) is the change in the output produced by a change in the input. The effect depends on the size of the change in the input, the initial value of the input, and perhaps the values of the other inputs. A cause has a *large effect* if a relatively small change in the input produces a relatively large change in the output. We define a small change in the input and a large change in the output relative to the variation we see in these characteristics under regular operation of the process. We simplify the language if we call all varying inputs *causes,* even those with no effect.

For any process output, there are likely to be a large number of causes, each with an effect. We assume that the Pareto principle[1] (Juran et al., 1979) will apply and that only a few causes will have large effects. We call these causes *dominant.* We base our strategies and approaches to reducing variation on the assumption that there will be only one or two dominant causes. We justify this focus on dominant causes more fully in Section 2.4.

We can model the effect of a dominant cause as

$$output = f\,(dominant\ cause) + noise,$$

where f (*dominant cause*) is a function that captures the effect of the dominant cause and the term *noise* captures the effect of all the other inputs. For a dominant cause, the range in f (*dominant cause*) is greater than the output variation due to the noise.[2]

A dominant cause can be a single input or involve two or more inputs in a variety of ways. Figure 2.3 shows a dominant cause for a binary output that is either good (G) or bad (B). The horizontal line represents the normal range of values of the input. Small values of the input to the left of the dotted line correspond to good output, large values to bad output.

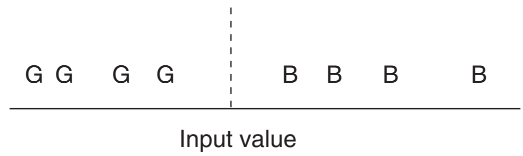


Figure 2.3 A dominant cause with continuous input and binary output.

Figure 2.4 shows another example of a dominant cause. The left-hand plot in the figure shows a continuous input and the right-hand plot a discrete input. In either instance, we see the full range of variation for both the input and output on their respective axes.

If both the input and output are binary (or discrete with a few possible values), we can depict a dominant cause using a table of percentages such as Table 2.1.

We hope to find a single dominant cause of variation such as shown in these examples. We may fail for several reasons. First there may be no single dominant cause; instead, we may find several causes, each with a relatively large effect. Second, we may find that the dominant cause involves two (or more) inputs.[3] Figure 2.5 shows two examples of a dominant

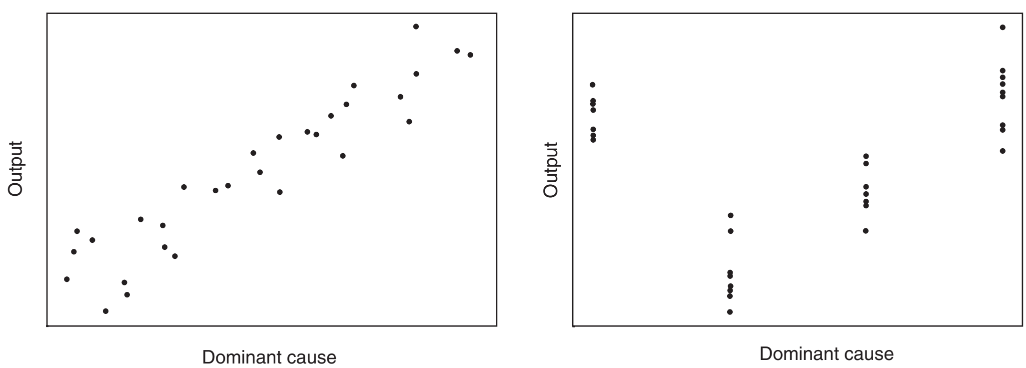


Figure 2.4 A dominant cause with continuous output.

Table 2.1 A dominant cause with binary input and output.

	Output pass	Output fail
Input low	80%	20%
Input high	97%	3%

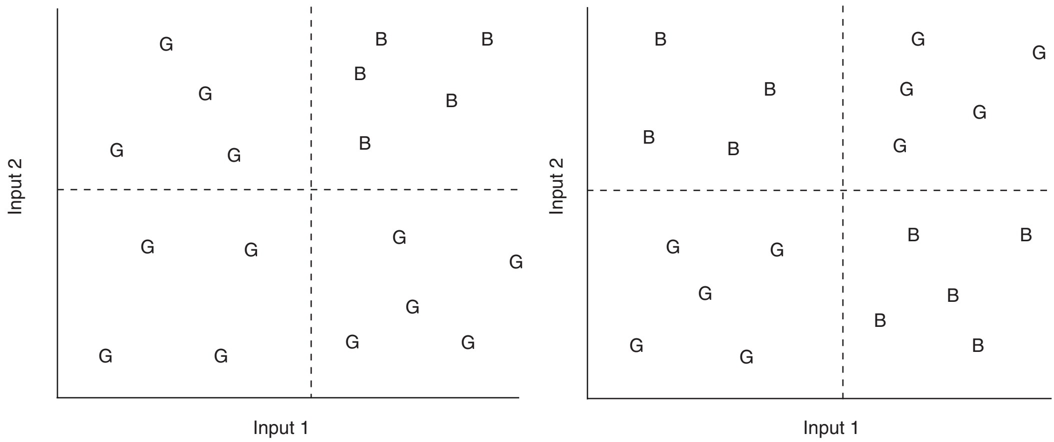


Figure 2.5 Two examples of a dominant cause involving two inputs with a binary output.

cause involving two continuous inputs, and a binary output where we denote good output as *G* and bad output as *B*. In the example presented in the left panel, both inputs must be simultaneously large in order to get bad output. In the right panel, good output results when the two input values are both large or both small. The latter case may arise if the output is a clearance between two assembled components whose critical dimensions are described by inputs 1 and 2.

Causes can be classified in many ways.[4] For us, the key issue is the size of the effects. We want to find dominant causes that contribute substantially to the variation in the output.

2.3 DISPLAYING AND QUANTIFYING PROCESS VARIATION

We will need to quantify process variation at several stages in the variation reduction algorithm. For example, we may set the goal to reduce the variation by 50%. We require a numerical measure of variation to define this goal explicitly.

We use the camshaft manufacturing process described in Chapter 1 to illustrate. The angle of the lobe axis is one important characteristic of the camshaft lobes. The specifications are ±400 thousandths of a degree measured as the deviation of the lobe axis from the nominal direction. A project team collected a sample of 108 camshafts, 9 per day over 12 days, and measured the angle error for each of the 12 lobes. The 1296 measurements of angle error and several other outputs are available in the file *camshaft lobe runout baseline*.

For a continuous output with two-sided specifications such as the angle error, we use the average and standard deviation (*stdev*) to summarize process behavior. From MINITAB, we get the following output.

```
Variable        N      Mean     Median     TrMean      StDev    SE Mean
angle        1296    -21.30     -18.00     -20.04      71.50       1.99

Variable       Minimum    Maximum         Q1         Q3
angle          -241.00     155.00     -67.00      30.00
```

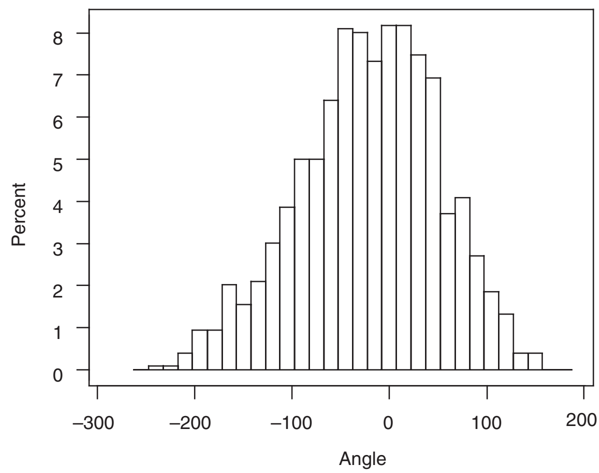


Figure 2.6 Histogram of camshaft lobe angle errors.

The average is –21.3 and the standard deviation is 71.5, as given in the columns `Mean` and `StDev`, respectively. See Appendix B for an interpretation of the other summary measures.

For the camshaft example, we can interpret the average and standard deviation using the histogram of angle error given in Figure 2.6.

The average is the point on the horizontal axis where the histogram would balance if we could cut it out of the page. In Figure 2.6, since the histogram is roughly symmetric about zero, the average (the balance point) is close to 0°. The width of the histogram is approximately six standard deviations when the histogram has a bell shape. Here, the width of the histogram is 400 thousandths of a degree and 6 *stdev* = 429. For a bell-shaped histogram, almost all of the characteristic values will fall within the range

$$average \pm 3\ stdev$$

From this argument, we see that the standard deviation is a measure of the unit-to-unit variation. The average is a measure of the process center. The distance from the average to the target is a measure of how well the process is targeted and hence a measure of the off-target variation. The average and standard deviation are sometimes combined in a capability ratio,[5] which can describe both kinds of variation simultaneously.

In the example, the target value for angle error is 0°. If the collected data represent the long-term behavior of the process, we cannot reduce variation significantly by better centering the process, that is, by adjusting the average to the target. Here, we must reduce the standard deviation to get a substantive reduction in the variation.[6]

Why is the standard deviation so large? Since the standard deviation measures the variation in angle error from lobe to lobe, there must be changes in varying inputs from lobe to lobe in the sample that explain the angle error variation. To reduce the standard deviation, we may first try to identify dominant causes; that is, varying inputs making major contributions to the standard deviation. Then we can try to eliminate the effects of these causes. We make this idea clear in the next section.

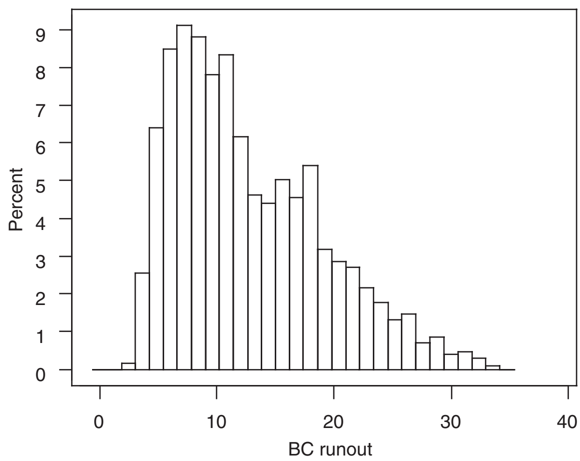


Figure 2.7 Histogram of camshaft lobe BC runout.

In Chapter 1, we described a problem in which the goal was to reduce variation in base circle (BC) runout, another characteristic of a camshaft. In that case, the target value was zero and the histogram of runout values is shown in Figure 2.7.

The average and standard deviations (given as follows by MINITAB) are roughly 12.6 and 6.4 microns, respectively.

```
Variable        N       Mean     Median      TrMean       StDev    SE Mean
BC runout    1296     12.643     11.100      12.271       6.389      0.177

Variable       Minimum    Maximum         Q1          Q3
BC runout        2.600     33.900      7.425      17.000
```

The histogram is not bell-shaped. The average ± 3 standard deviations is (–6.6, 31.8). We see that some values are above 31.8 and none are close to –6.6. We no longer have the interpretation that almost all of the values fall within three standard deviations of the average. However, we can still interpret the standard deviation as about 1/6 of the range since (33.9 – 2.6) ÷ 6 = 5.2. Here, to reduce variation around the target, we need to shift the average to the left and reduce the standard deviation. If we identify the dominant cause, we may find a low-cost way to reduce the variation.

Somewhat surprisingly, adjustment of the process center to the target can play a role in reducing unit-to-unit variation. To see how this happens, consider again the angle error data. Recall that there are 12 lobes on each camshaft with the positions numbered 1 to 12. We can use a box plot (see Appendix C), as given in Figure 2.8, to compare the performance of angle error from lobe position to position.

The lobe averages range from roughly 45 (lobe 2) to –90 (lobe 12). If we adjusted the process on each lobe separately so that the angle error average was on target for each lobe, then the overall standard deviation would be reduced to 61.6 from 71.5. You can see this result qualitatively by imagining all of the boxes in Figure 2.8 being shifted vertically to

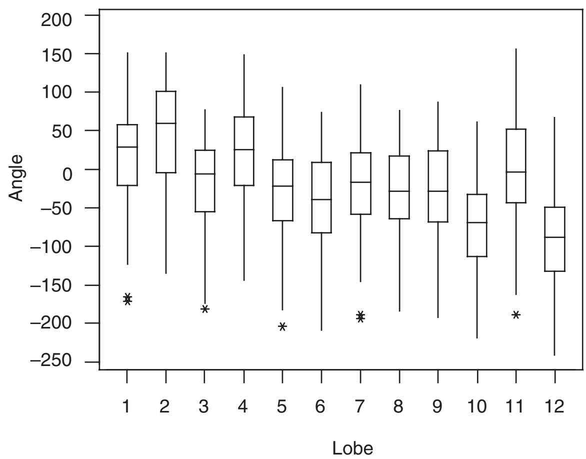


Figure 2.8 Box plots of camshaft lobe angle errors by lobe position.

have a centerline on zero. Then the range of values shown by the whiskers on the box plots and the overall standard deviation would be smaller.[7]

Data summaries such as the average, standard deviation, histogram, and box plots do not show how the process output varies over time. To show this behavior, we use a run chart, a simple plot of the output values against the order or time of collection. The run chart for the angle-error data from lobe 12 is shown in Figure 2.9.

The run chart shows how the process output varies over time. We may see cycles and smooth patterns on such a plot. The time structure of the output variation is important when we define the problem baseline (Chapter 6) and when we consider feedback controllers (Chapter 18). In Figure 2.9, we see most of the full range of variation from one camshaft to the next. There is no obvious longer-term pattern. In the production of V6 pistons, the diameter was recorded for one process stream at Operation 270 every minute for 200 minutes.

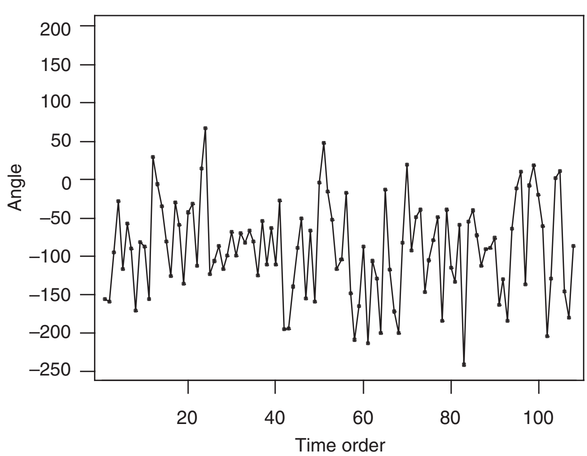


Figure 2.9 Run chart for angle error at lobe 12.

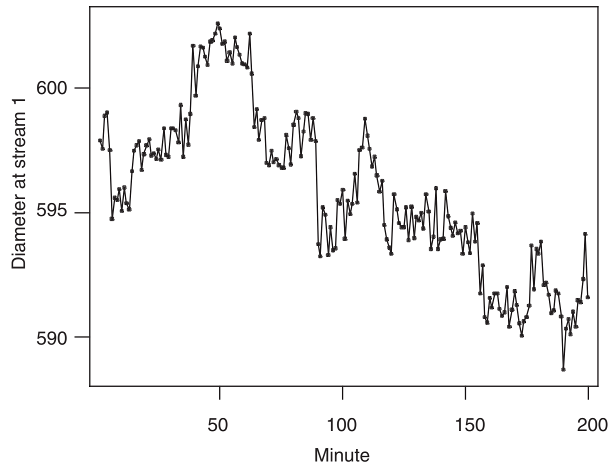


Figure 2.10 Stream one V6 piston diameter at operation 270 by hour.

The run chart is given in Figure 2.10. Here, the diameter shifts occasionally but relative to the long term, the output variation is small over the short term.

We can apply these numerical summaries and plots to data for a continuous output. If the output is binary (typically defective or not) or categorical with only a few values, we use proportions to quantify process variation. For example, in the process that produces manifolds described in Section 2.1, there are two manifolds produced in each mold. During an investigation, a team collected a sample of 40 castings, 20 from each cavity labeled A and B. The team classified each casting as defective (1) or not (0). We summarize the variation using the proportion defective. From MINITAB we get

```
Tabulated Statistics: output, cavity

 Rows: output     Columns: cavity
         A        B      All
 0      20       18       38
 1       0        2        2
 All    20       20       40
```

In this case, $2 \div 40 = 5\%$ of the castings in the sample are defective. The corresponding percentages are 0% and 10% for cavities A and B, respectively. We do not find histograms or bar charts helpful for these data. We can sometimes use a run chart to display clusters of defectives over time.

We have presented a number of charts and statistics for quantifying and describing process variation. We construct these summaries from a sample of units collected over time from the process. We must be careful that we collect data that gives us an accurate picture of the long-term process performance. For example, we would not have seen the structure of the variation over time in Figure 2.10 as clearly if we had sampled 200 pistons over a few hours rather than over 200 hours. We deal with this critical idea technically in Chapter 5 and practically throughout the book whenever we describe a process investigation. The bottom line is that we will not make progress without careful thought on how we collect process data.

2.4 MODELS FOR VARIATION AND THE EFFECTS OF CAUSES

Throughout the book, we use a simple model to connect the causes and the output. The model will help us identify dominant (and unimportant) causes of variation and to quantify their contributions to the variation in the output. As well, we can use the model to understand how to combine contributions from various causes and to understand how variation is transmitted through a process. The functional model, *output* = *f* (*input1, input2, ...*), which we introduced in Section 2.2, is useful to explain the idea of a cause and its effect but is too complex to apply in practice.

The basis of the model is a mathematical representation of a histogram that specifies the relative frequency with which different output values occur. To avoid confusion with the measured values, we denote the output in the model by an uppercase letter, typically Y. We associate a smooth curve, an idealized histogram, with Y as shown in Figure 2.11. In this case, the model is a Gaussian curve[8] (also called a bell or normal curve) that can be specified by two *parameters,* the mean (or center of the symmetric curve) and the standard deviation, a measure of the spread of the values. The mean and standard deviation (*sd*) associated with the model are directly analogous to the average and *stdev* of the histogram. The deviation between the mean and the target represents the off-target variation and the standard deviation describes the variation from unit to unit.

When we apply the model, we estimate the mean and standard deviation using the values of the output characteristic in a sample of units from the process.

The model behaves as expected if we rescale the output. Suppose the output value Y is changed by first multiplying by one constant (b) and adding another (a) so that

$$Y_{new} = a + bY_{old}$$

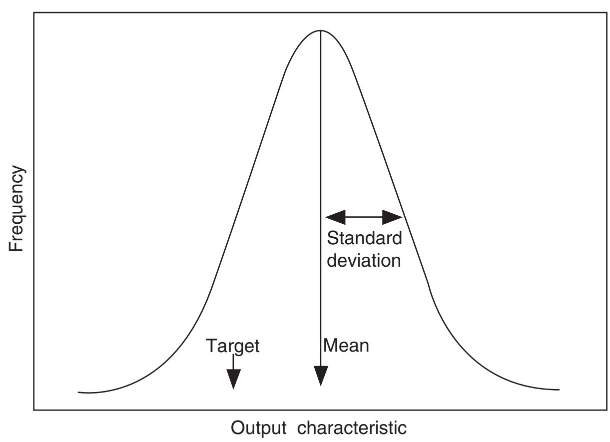


Figure 2.11 A Gaussian output model.

Then, in the model, the mean and standard deviation are changed to

$$\text{mean}(Y_{\text{new}}) = a + b\ \text{mean}(Y_{\text{old}}),\quad sd(Y_{\text{new}}) = |b|\ sd(Y_{\text{old}})$$

For example, suppose the model for a temperature output has mean 1234°F and standard deviation 56.7°F. If we change from Fahrenheit to Celsius (°C = 0.556°F – 17.778), then the new model has mean 668°C and standard deviation 31.5°C. Note that if we multiply Y by negative one, the standard deviation in the model does not change.

We want to decompose the model for the output into pieces corresponding to the effects of causes. Consider the following example. Suppose that two components A and B are stacked and the height of the assembly y is the output of interest. This is illustrated in Figure 2.12.

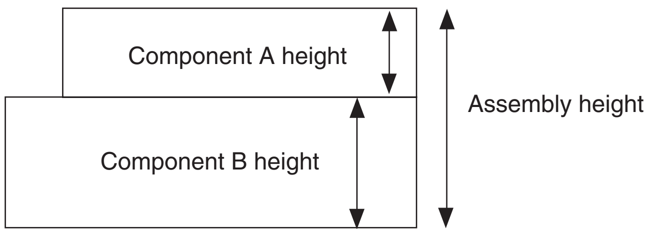


Figure 2.12 Stack height.

We have a model for the height of each component as described in Table 2.2. The means and standard deviations are based on data collected from the process that produces each component.

Here X_A and X_B represent the effect of component A and component B on the stack height. As such, we can construct a model for stack height from the two models of component heights, namely

$$Y = X_A + X_B$$

Adding the models does not mean that we add the idealized histograms. Fortunately, there are simple rules to describe how to combine the means and standard deviations. For the means, we get

$$mean(Y) = mean(X_A) + mean(X_B)$$

Table 2.2 Model for each component.

Component	Model	Mean	Standard deviation
A	X_A	12.212	0.045
B	X_B	17.567	0.033

In words, the mean of the sum is the sum of the means. This formula has important consequences when we adjust a process center to target by changing a fixed input. For example, we can use the model to predict the effect on the output if we shift the mean of component A.

For the standard deviation, we have

$$sd(Y) = \sqrt{sd(X_A)^2 + sd(X_B)^2} \tag{2.1}$$

In words, the standard deviation of the sum is the square root of the sum of the squares of the standard deviations of the terms in the sum. There is a demonstration in the exercises to convince you that this key formula is true. Equation (2.1) applies when the two component heights vary independently.

Using the model, the mean and standard deviation of the assembly height are

$$mean(Y) = 12.212 + 17.567 = 29.779, \;\; sd(Y) = \sqrt{(0.045)^2 + (0.033)^2} = 0.056$$

Equation (2.1) has many important consequences. The standard deviation of the sum is much less than the sum of the standard deviations. This is good news when you are building up assemblies, because the overall variation will be less than the sum of the component variation. However, it is bad news when it comes to reducing variation. To see why, suppose that (for a price) we contemplate reducing the standard deviation of the height of component B by 50% from 0.033 to 0.016. We can use the model and Equation (2.1) to predict the impact on the variation of the assembly height. The effect on the standard deviation of the assembly height is surprisingly small; the standard deviation becomes 0.048, a 15% reduction.

More generally, suppose we consider only two sources of variation, one attributed to a particular cause and the second to all other causes. To be specific, suppose that the specific cause (actually a group of causes) is the measurement system, so that all other causes are responsible for the variation in the true value of the outputs. If the effects are additive and independent, at least approximately, the overall standard deviation is

$$sd(\text{total}) = \sqrt{sd(\text{due to measurement})^2 + sd(\text{due to rest})^2}$$

Now suppose that the standard deviation due to the measurement system is 30% of the overall standard deviation, that is, the ratio of *sd*(due to measurement) to *sd*(total) is 0.3. What is the percentage reduction in overall variation if the variation due to the measurement system is eliminated?

We constructed Figure 2.13 to give the percent reduction in overall standard deviation if we could eliminate completely the contribution of an identified cause. From the plot, we see that when the ratio of the *sd*(due to cause) to the *sd*(total) is 0.3, the potential gain is about 5%. In other words, if the ratio of measurement system variation to total variation is 0.3, you can reduce the overall standard deviation by at most 5% if you replace the current measurement system with one that is perfect.

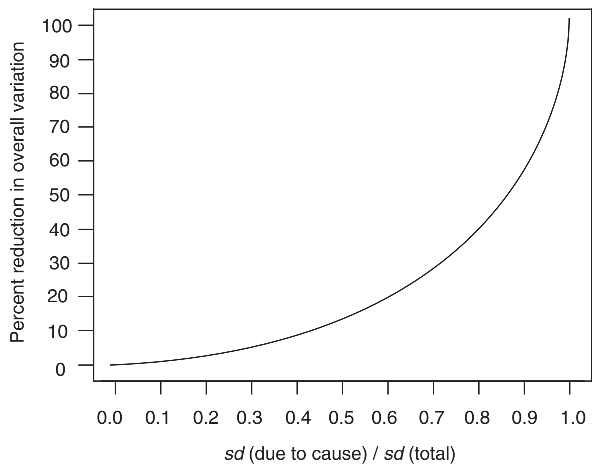


Figure 2.13 Reduction in variation if we remove a cause contributing a given proportion of the total variation.

In the stack height example, the ratio of $sd(X_B)$ to $sd(Y)$ is 0.59, and thus, by eliminating the variation in component B completely we can reduce the overall standard deviation by about 20%. Similarly, the ratio of $sd(X_A)$ to $sd(Y)$ is 0.80, and by eliminating the variation in component A, we reduce the overall standard deviation by about 41%.

The message from Figure 2.13 is that we need to address dominant causes of variation in order to make a significant reduction in variation. If we rank the potential gains (in terms of percent reduction in overall standard deviation) by eliminating the contributions of various causes, we expect to see a Pareto chart like Figure 2.14. There is little opportunity for gain in identifying causes with small contributions to the overall standard deviation. For this reason, the proposed variation reduction algorithm (see Chapter 4) focuses on finding and dealing with dominant causes.

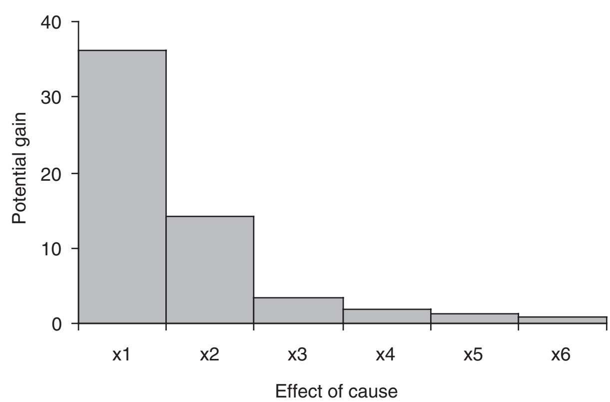


Figure 2.14 Pareto chart of potential gain.

We can also use a model to describe how variation is transmitted through the process. Suppose the scatter plot of the output versus some specific input is given by the left-hand side plot in Figure 2.4. We adopt the model

$$Y = a + bX + noise,$$

where Y represents the output, $a + bX$ the effect of the identified cause, X, and *noise* the rest of the causes, presumably unidentified. The parameters a and b describe the linear relationship between Y and X. From the model we can see that

$$sd(Y) = \sqrt{b^2 sd(X)^2 + sd(noise)^2}$$

In this example, we denote *sd* (due to X) by $b*sd(X)$. Using the model, we can predict how much of the standard deviation in Y is due to changes in X. For example if we could hold X fixed (that is, make its standard deviation 0), then the standard deviation of Y will be reduced to standard deviation of the noise. We can estimate this residual standard deviation by fitting a line to the scatter plot in Figure 2.4 (see chapters 10 to 12). We can also contemplate changing a fixed input to change the slope b; again, we can use the model to predict the effect of this change.

We can construct similar models for binary outputs. These are more complex and prove less useful so we omit them here.

Key Points

- Process descriptions should be kept as simple as possible while still capturing the needed information.
- Variation is an attribute of the process, not of single units produced by the process.
- We can quantify and describe process variation using summaries such as average, standard deviation, histograms, and run charts.
- Reducing variation in a cause has a substantial effect on the overall variation only if the cause is dominant.
- To determine the effect of various sources of variation on the overall variation, we have

$$sd(\text{overall}) = \sqrt{sd(\text{due to a specific cause})^2 + sd(\text{due to all other causes})^2}$$

End Notes (see the Chapter 2 Supplement on the CD-ROM)

1. The Pareto principle as applied to variation reduction problems is explored in more detail in the supplement.
2. We give a more formal derivation of the definition of a dominant cause. Some possible input-output relationships are also explored.
3. In the chapter supplement, we define what it means for a dominant cause to involve two (or more) inputs and explore the consequences.
4. We can classify causes of variation in many ways. For example, in Statistical Process Control (SPC), we speak of special and common causes. Taguchi (1986) calls causes controlled or noise and subdivides these classes even further. We look at these classification systems and their relationship to variation reduction.
5. There is an alphabet soup of capability rations—such as P_{pk}, C_{pk}, and so on—used to quantify process performance and its relationship to specifications. We give a brief discussion of these indices that are sometimes used to set the goal for a variation reduction project.
6. We have described two kinds of variation, deviation of the average from the target and variation from unit to unit. We provide a formula that relates the two kinds of variation to the "variation from the target," the single measure of variation that is likely related to cost.
7. In the camshaft lobe BC runout example, we showed how aligning the average angle error for each lobe could reduce the overall standard deviation. We give a key formula widely used in the analysis of variation to connect the overall standard deviation to the variation within groups and group to group.
8. The Gaussian distribution is widely applicable. We describe some of its key properties.

Exercises are included on the accompanying CD-ROM

3

Seven Approaches to Variation Reduction

A fool can learn from his own experiences; the wise learn from the experience of others.

—Democritus, 460–370 B.C.

There are many ways to change fixed inputs in the process to reduce variation in an output characteristic. In this chapter, we classify methods for variation reduction into seven generic approaches (MacKay and Steiner, 1997–98):

1. Fixing the obvious based on knowledge of a dominant cause of variation
2. Desensitizing the process to variation in a dominant cause
3. Feedforward control based on a dominant cause
4. Feedback control
5. Making the process robust to cause variation
6. 100% inspection
7. Moving the process center closer to the target

We must identify a dominant cause of the variation for the first three approaches but not the final four.

To implement any of the approaches, we need to change one or more fixed inputs. The possible changes include:

- Changing a set point (for example, machine settings, specifications for an input, supplier, and so on)
- Adding or changing a process step (for example, adding inspection, replacing a gage, retraining an operator, rewriting instructions, and so on)

- Changing the control plan (for example, adding a feedback or feedforward controller, changing the current controller, and so on)
- Changing the product design

These changes can be implemented anywhere in the process including at suppliers. We need to be careful since changing any fixed input may add significant operating costs or produce serious *side effects* defined in terms of other process outputs.

In this chapter, we introduce and discuss the seven variation reduction approaches. For each approach, we discuss how and when it works and the potential difficulties of implementation. We provide details and further examples in chapters 14 through 20.

The variation reduction approaches are an integral part of the proposed algorithm. The algorithm is the topic of Chapter 4. The ultimate aim of the algorithm is to lead to process improvement by implementing one (or more) of the seven variation reduction approaches. We provide further comparison and discussion of how to choose an approach in chapters 8 and 14.

3.1 FIXING THE OBVIOUS BASED ON KNOWLEDGE OF A DOMINANT CAUSE

After finding a dominant cause of variation, we can sometimes reduce variation in the output by implementing an obvious fix. For instance, we might identify temperature as a dominant cause and reduce temperature variation by more frequent adjustment. The effect of reducing variation in a (continuous) dominant cause is illustrated by the scatter plots in Figure 3.1. The vertical dashed lines specify the range in the input (the dominant cause here) and the horizontal dashed lines the corresponding range in the output. If we reduce the variation in the input, as shown in the right panel of Figure 3.1, the variation in the output will be substantially reduced.

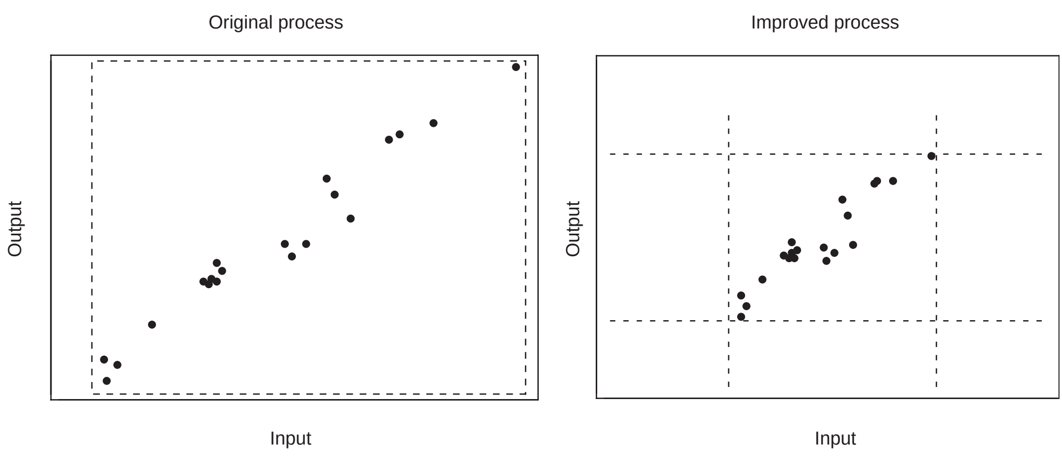


Figure 3.1 Reducing variation in a dominant cause.

We can sometimes eliminate variation in the dominant cause. In a drilling process, there was a frequent problem with holes drilled in the wrong location. Plates were drilled on a multispindle drill press. The defects were not found until assembly. Setting a plate upside-down or rotating it from the correct orientation before drilling was the dominant cause. The process engineer redesigned the fixture holding the plate so that it was impossible to mount the plate in the press incorrectly. The variation in the dominant cause was eliminated.

Figure 3.1 shows a continuously varying dominant cause. We can also reduce variation in a discrete cause, as shown in Figure 3.2. The dashed horizontal lines show the range of output. In the left panel of Figure 3.2, we see that machine number is a dominant cause of variation in the output. There are large differences in the average output level (process center) for different machines. In the right panel of Figure 3.2, we see that we can reduce variation by aligning the average output for each machine. There may be an obvious method for making such an adjustment.

For example, if a dominant cause is the difference in how two operators control the process, then we can reduce variation by retraining the operators and clarifying the control plan. In another example, a team identified the difference between two suppliers of valve lifters as the dominant cause of high oil consumption in truck engines. The team could eliminate most of the problem by switching to a single supplier or by establishing procedures to reduce differences between the suppliers.

In an engine assembly plant, the original problem was the high frequency of rejects for excessive noise at the valve-train test stands. The team discovered that there were large differences in average measured noise level for each of the three parallel test stands, even though the engines were haphazardly assigned to the stands for testing. After calibrating each stand, they found the average noise levels were roughly equal. The real challenge for the team was to keep the problem from recurring.

There are two conditions necessary for the fix-the-obvious approach to work. First, we must be able to identify a dominant cause. We waste valuable resources and make little or

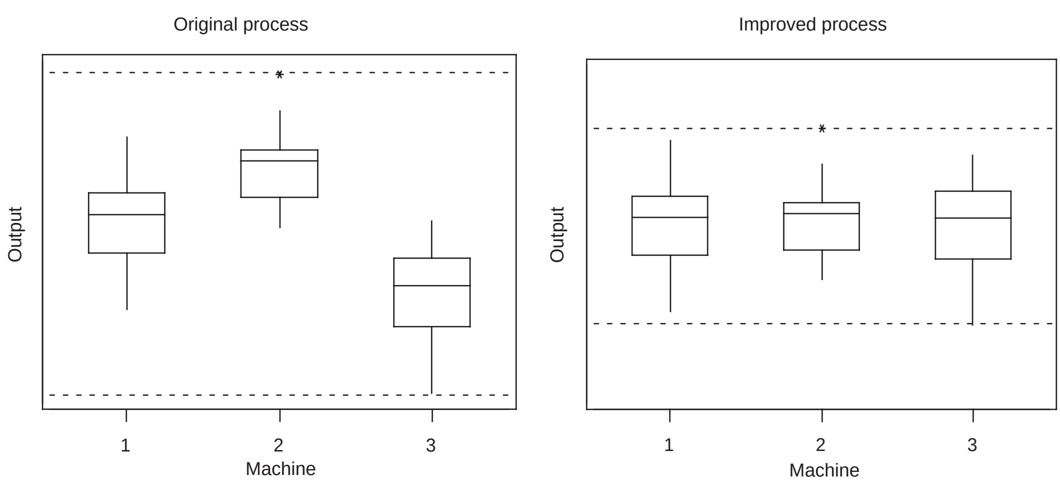


Figure 3.2 Aligning the process center stratified by a dominant cause.

no progress fixing problems based on a nondominant cause. Second, based on process knowledge, we must be able to address the cause with an obvious solution whose feasibility is without question.

3.2 DESENSITIZING THE PROCESS TO VARIATION IN A DOMINANT CAUSE

If we find a dominant cause as shown in the left panel of Figure 3.3, we can reduce variation in the output by flattening the relationship between the input (dominant cause) and the output as shown on the right panel of the figure. The dashed horizontal lines show the range of output values for the original and improved processes. We reduce the sensitivity of the output to variation in the input by changing some fixed inputs. We do not reduce variation in the dominant cause.

We find process desensitization useful when the variation in a dominant cause is hard or impossible to control. For example, we discussed a frost buildup problem in refrigerators in Chapter 1. The dominant causes of the variation in frost buildup were environmental and usage inputs, such as ambient temperature and the average number of times per hour that the door was opened. You can imagine that the instruction "Open this fridge at most once per hour" to reduce variation in a dominant cause would not be effective here. Instead, the manufacturer decided to investigate a number of design changes (that is, changes of fixed inputs) to reduce the sensitivity of the refrigerators to varying usage inputs. The changes investigated were directed by the knowledge of the causes of the problem.

For a second example, a team was charged with reducing scrap due to porosity found on a machined surface of an engine block. The porosity was due to gas bubbles trapped near the surface of the block when the molten iron solidified. The team decided to search for a dominant cause and discovered that low iron-pouring temperature resulted in a high frequency of porosity scrap. Looking for a more specific cause, they found that low-pouring

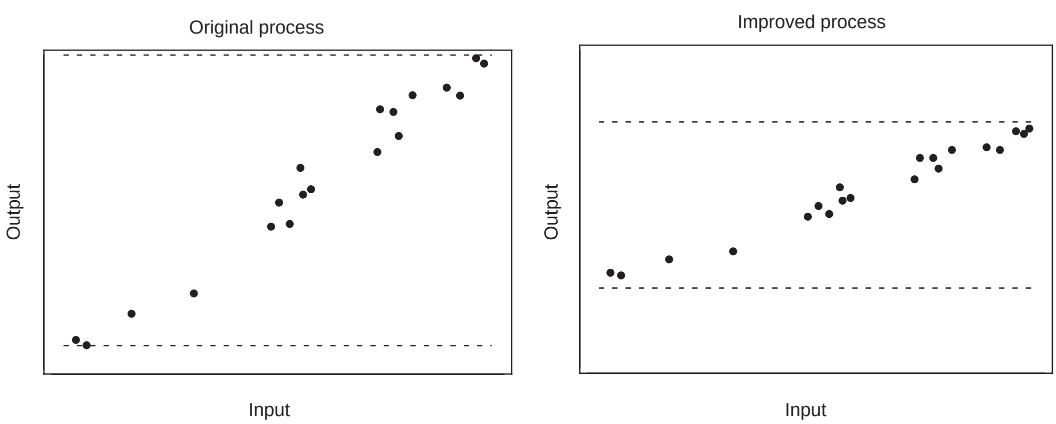


Figure 3.3 Desensitizing a process to variation in a dominant cause (dashed horizontal lines show the range of output values).

temperature was due to unscheduled and scheduled (for example, lunch) downtime at the pouring operation, where there was no method for maintaining iron temperature. The cost of adding controls was prohibitive. The team looked for ways to desensitize the porosity level to pouring-temperature variation. They discovered that changing the wash on the sand cores substantially reduced the sensitivity of the porosity to low pouring temperature. The foundry adopted this low-cost solution.

Process desensitization is a desirable approach since, once it is complete, no further action is required. First, we must find a dominant cause of variation. We can use knowledge of this cause to help select fixed inputs that we might change to desensitize the process. In the engine block porosity example, once the team understood that low pouring temperature was a dominant cause of porosity, they were led to consider changing the core wash and core sand composition. Without knowledge of a dominant cause, it is unlikely that they would have thought about changing these fixed inputs.

It is difficult to predict when desensitization will work. This is its great weakness. We require a great deal of process knowledge to make an output less sensitive to variation in a dominant cause. Which fixed inputs should we change? We require expensive designed experiments to determine these inputs (if any exist) and their appropriate settings. The experiments may fail to determine process settings that lead to improvement. Also, the new process settings may lead to increased operating costs or negative side effects.

3.3 FEEDFORWARD CONTROL BASED ON A DOMINANT CAUSE

With feedforward control, we adjust the process based on measurements made on a dominant cause of variation, anticipating the effect on the output. Suppose we have identified a dominant cause as in the left panel of Figure 3.4. If we have a high value for the input, we predict that we will get a high value for the output, so we adjust the process center downward to compensate.

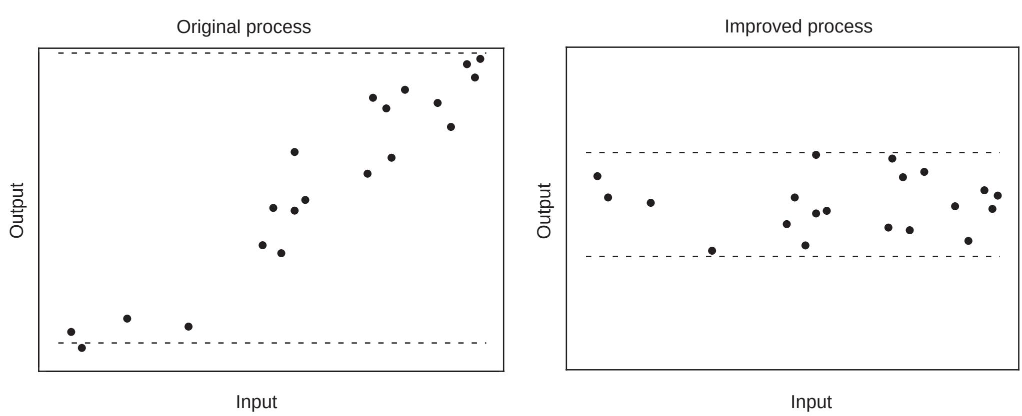


Figure 3.4 Reducing output variation using feedforward control.

Feedforward control is a form of process desensitization, as discussed earlier. However, with feedforward control, we achieve the desensitization by active adjustment as the process operates, rather than with a one-time change in the settings of some fixed inputs. We can see how feedforward control reduces the variation in the output in the right panel of Figure 3.4.

In the development of the truck alignment process discussed in Chapter 1, the geometry of the frame was a dominant cause of variation in the alignment characteristics. To reduce variation, the frame manufacturer measured each frame and recorded the geometric information with a bar code. The assembly plant read the information and manufactured a special component to adjust the alignment characteristics to remove the effect of frame geometry. This approach was relatively expensive but led to a dramatic reduction in the variation of the alignment characteristics.

Selective fitting is another example of feedforward control. Components are sorted and matched to produce good assemblies. This adds cost and complexity to the assembly process. In a problem to reduce steering wheel vibration, a team discovered that a dominant cause was an imbalance in the transmission. The problem was reformulated to reduce imbalance in the transmission. Then the team identified two mating components that were dominant causes of the transmission imbalance. They improved the balance of the transmission by measuring and vectoring the two transmission components, that is, assembling the components so that their individual contributions to imbalance tend to cancel. As a second step, they sorted transmissions on the basis of imbalance and used the better-balanced transmissions in the most sensitive vehicle models. Both changes are examples of feedforward control.

Feedforward control works under the following conditions. First, we must identify a dominant cause. Second, the relationship between the cause and the output must be well known and stable over time. Third, we must be able to measure the cause in a timely way. Finally, there must be a way to adjust the process to compensate for the anticipated effect of the cause. As a result, to implement feedforward control, we may need to find a way to adjust the process center.

If the dominant cause changes slowly and the conditions are met, we can use feedforward control with only occasional adjustments. For example, in a batch process, if a dominant cause is a raw material characteristic that is constant over a batch, we can use the value of the cause to set up the process to remove or reduce its effect. Here, we make an adjustment only once per batch.

There are substantial costs and risks associated with feedforward control. Costs arise because we need to identify a dominant cause, measure the cause, and repeatedly adjust the process. There is a danger of overadjustment if there is measurement error in determining the input (the dominant cause) value, if the relationship between the input and the output is not well understood, or if the adjustment scheme is imperfect. In addition, repeated process adjustment may be impractical or costly and may introduce undesirable side effects.

3.4 FEEDBACK CONTROL

The idea behind feedback control is to monitor the output characteristic and to predict future behavior from current and past observations. If the predicted output value is far from the target, we make an adjustment to the process. In Figure 3.5, the panel on the left

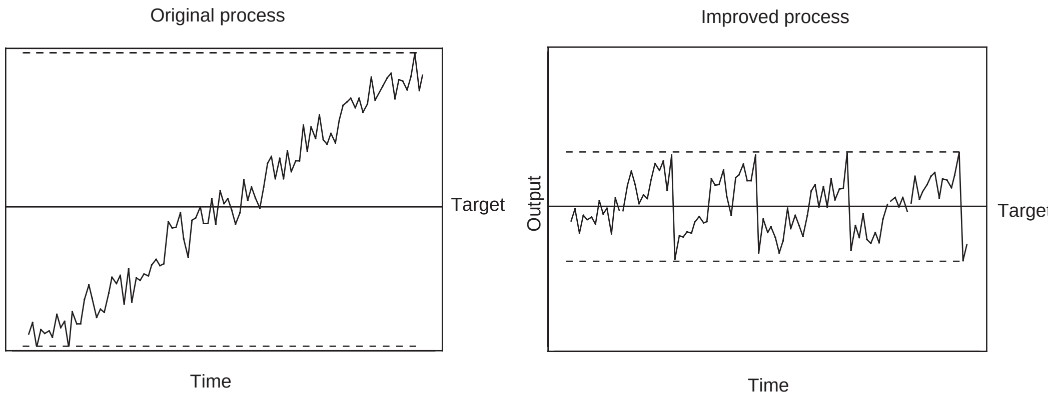


Figure 3.5 Feedback control.

shows the output of a process without feedback adjustment. The dashed horizontal lines indicate the range of output values. At any particular time, we can predict that the output will be larger in the future. If we predict that the output is far from the target, we adjust the process towards the target. The right panel of Figure 3.5 shows the dramatic reduction in variation of the output due to this approach. Here, using the feedback control procedure, we adjust the process whenever the output falls outside the adjustment limit (not shown, but near the top dashed horizontal line in the right panel of Figure 3.5). The size of adjustment is based on the deviation from the target of the last observed output value.

A team wanted to reduce variation in the film thickness in the painting of fascias. Too much paint resulted in defects such as sags and runs, and too little paint created appearance defects. The team discovered that a dominant cause of thickness variation was the flow rate of the paint. They reformulated the film thickness problem in terms of flow rate. The team then identified a dominant cause of variation in flow rate that was addressed by making equipment modifications. After these changes, the flow rate continued to show a pattern in the variation over time as shown in Figure 3.6.

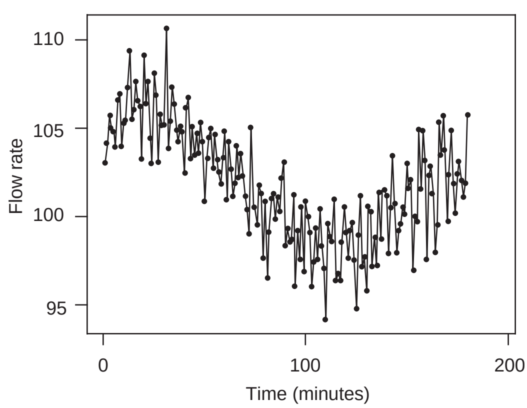


Figure 3.6 Variation in flow rate over time.

The team defined a feedback control scheme that involved considering an adjustment of flow rate every minute. If the flow rate fell outside prescribed limits, a valve was adjusted to move the process center to the target value of 103. The feedback scheme produced a further substantial reduction in film thickness variation.

As a second example, in a project to reduce scrap due to variation in the diameter of crankshaft journals, the team discovered that an automated 100% inspection gage drifted substantially over time. They could not determine the cause of the drift, so they decided to use a feedback controller to adjust the gage when appropriate. A reference part was measured several times just after the gage had been cleaned and calibrated. The team used these measurements to establish a centerline and adjustment limits for the gage. Every four hours the reference part was measured, and if the measured value fell outside the limits, the operators recalibrated the gage.

The use of setup procedures based on first-off measurements is another example of feedback control. For example, an operator used the measured output of the first few units produced after a tooling change to adjust the machine. Once a good setup was achieved, no further process measurements were taken or adjustments made until the next tool change.

We can apply feedback control successfully when three conditions are satisfied. First, the process must exhibit a strong time structure in the output variation as discussed in Chapter 2. Examples include drift due to tool wear and stratification due to batch-to-batch variation. Second, there must be an adjustment procedure to move the process center. Finally, the time to measure the output and adjust the process must be small relative to the rate of change of the process.

The major advantage of feedback control is that it requires no knowledge of the dominant cause of variation. We use only the measured output values. A drawback is the high cost of process measurements and adjustments. Finally, due to the feedback nature of the control, there is an inherent time delay. To identify when an adjustment is required, we must first observe some output values that are significantly different from the target value. Thus, feedback control is always reactive.

3.5 MAKING THE PROCESS ROBUST

To make the process robust, we first determine a performance measure, such as the standard deviation of the output values for 20 consecutive parts. Next we change one or more fixed inputs to see if the performance measure has improved. In Figure 3.7, we use a box plot to show performance, and the improvement due to changing the fixed inputs is clear. The horizontal dashed lines show the range of output values before the change in fixed inputs.

The challenge with this approach is to identify which fixed inputs to change. For example, to measure the concentration of silicon in cast iron, the operator poured coins from a sample of molten iron, prepared the coins by grinding the surface, and then used a spectrometer to determine silicon concentration. The repeatability, defined as the variation when the same batch of iron was measured by the same operator in a short time, was unacceptably high. To reduce this variation, the team decided to adopt the

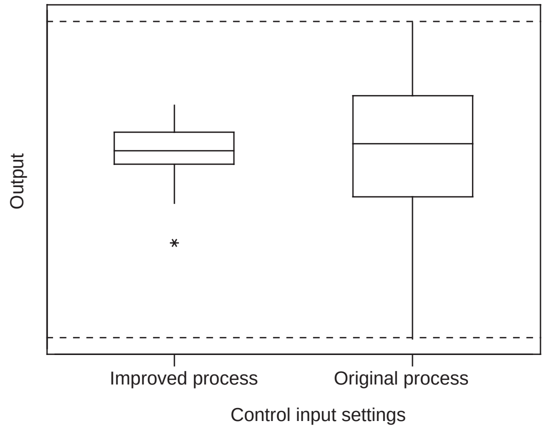


Figure 3.7 Making the process robust.

robustness approach. First they selected a run and a performance measure, the standard deviation of the measured silicon concentration in five coins poured from the same batch of iron. Second, they selected a number of fixed inputs, including mold shape, sample polishing method, and sample temperature. They organized a large experiment in which all of the selected inputs were changed in an organized fashion. For each combination of the inputs, they prepared and measured five coins and then calculated the performance measure. They found new settings for several inputs that reduced the repeatability variation.

As a second example, in a painting operation (discussed in Chapter 1), there was excessive variation in film build (paint thickness) from vehicle to vehicle at particular locations. As a consequence, to meet the minimum film build specification, the process operators kept the process center well above the lower specification. This resulted in high paint usage and occasional visual defects such as runs. To reduce the variation, the project team used the standard deviation of film build over five consecutive panels (it was too expensive to use cars for this investigation) to measure performance. Then they varied five fixed inputs, including some process parameters and paint properties, to explore settings to reduce the variation. The team discovered new settings of the fixed inputs that reduced the panel-to-panel variation by a factor of two. The average film build was then reduced with substantial savings in paint and rework costs.

Making the process robust will be successful if we can identify fixed inputs that can be changed to improve the performance measure. This approach is similar to desensitizing the process to variation in a dominant cause. In both, the goal is to make the process less sensitive to variation in the dominant cause. However, with the robustness approach, we select and change fixed inputs without first identifying the dominant cause. Without such knowledge, we find it more difficult to determine which fixed inputs to change and by how much. There is considerable risk with the robustness approach that significant resources will be used in a fruitless search for better process settings.

Note that we use the terms *robustness* and *desensitization* to label different approaches. Many authors use them interchangeably.

3.6 100% INSPECTION

100% inspection is the simplest variation reduction approach. We reduce output variation by identifying and then scrapping or reworking all items that have values of the output beyond selected inspection limits. We illustrate the effect of 100% inspection in Figure 3.8. We will scrap or rework (and reinspect) all units with output values outside the vertical dashed lines. Assuming no inspection or measurement error, the new full range of output is given by the dashed lines. The more the inspection limits are tightened, the greater the reduction in variation in the output characteristic.

In the production of cast-iron exhaust manifolds, a rare defect is a blocked port. The defect was not detected until the engine was assembled and tested. The team assigned to eliminate the defect had great trouble determining a dominant cause because the defect occurred so rarely. The team abandoned the search for the cause and, instead, designed and built a new device to automatically inspect all ports on every casting produced. Management agreed that the extra cost of the inspection was justified so that their customers could be assured that no manifolds with blocked ports would be shipped.

In the manufacture of a forced-air furnace, there was an emergency limit switch that shut down the furnace if the plenum got too hot. The heart of the switch was a plastic base in which a number of components were mounted to detect high temperatures and mechanically break the circuit. There was 100% inspection of the assembled switches. The scrap cost was high because the component and assembly costs could not be recovered from a scrapped switch. Using Pareto analysis, the team discovered that the most frequent reason for scrapping a switch was a broken tab in the plastic base. They also found that the tab was broken before any components were assembled into the base. The team decided to institute 100% inspection on the bases to scrap those with a broken tab before the components were installed. This reduced scrap costs by 50% and more than justified the inspection cost.

100% inspection is possible if the output characteristic can be determined for every unit in advance of shipping the product to a customer. Inspection has a number of negative features. The cost of reducing variation by tightening the inspection limits may be high due to increased rework and scrap costs and lost capacity. Also, the cost of inspection may be high if a new gage or additional labor is required. The reduction in variation will be less than

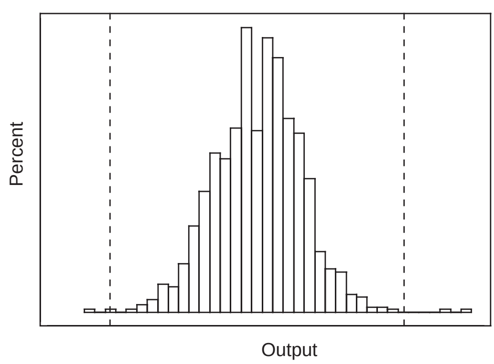


Figure 3.8 100% inspection.

anticipated if there are measurement or inspection errors. As a result, given the propensity of people to make inspection errors, most successful applications use automated inspection.

Inspection on the output has a negative reputation among quality professionals. However, we often apply 100% inspection on a dominant cause. This is called *source inspection,* one form of error proofing (Shingo, 1986).

3.7 MOVING THE PROCESS CENTER

This variation reduction approach is different from the others because we use it to change the process center to reduce off-target rather than unit-to-unit variation. We may need to move the process center closer to target as in Figure 3.9, or we may need to align subprocess centers as in Figure 3.2.

A team was charged with reducing the amount of shrinkage in a molded plastic casing. The casing was part of an assembly, and if there was excessive shrinkage, there was interference between the casing and the enclosed cable. Shrinkage was defined as a percentage change in casing length after a fixed curing time. The goal of the project was to reduce the average shrinkage. The team selected 15 fixed inputs, both raw material components and processing parameters. They planned an experiment where, for each combination of the input values, they measured the average shrinkage in a 500-meter length of casing. The results suggested two fixed inputs that could be changed to significantly reduce average shrinkage. The team did not determine the cause of shrinkage under the original process conditions.

In another example, a major source of scrap in the casting of aluminum differential carriers was a shrink defect resulting from the aluminum pulling away from the mold as it solidified. To reduce the rate of shrink defects, a team invented a five-point score that rated carriers from bad (score 1) to good (score 5). The goal was to increase the average score. Nine fixed inputs were varied in an experiment to determine levels that maximized the average score over 20 carriers. Again, the team made no attempt to determine a dominant cause of the shrink defect.

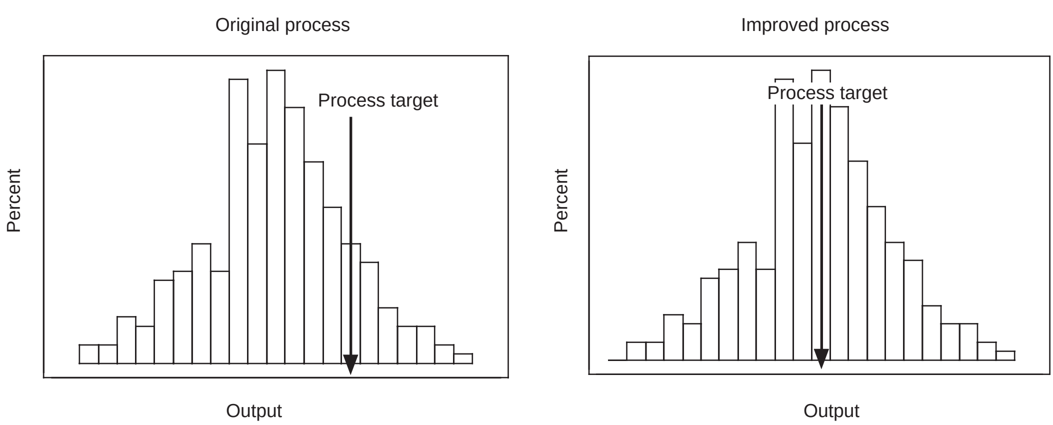


Figure 3.9 Moving the process center.

To apply the Move Process Center approach, we must discover a fixed input or inputs that we can change to shift the process center. Typically, we do this with an experiment in which we investigate several fixed inputs simultaneously.

There are risks associated with adjusting the process center. First, we may not find a fixed input that changes the process center and so waste the cost of the experiment. Second, changing the fixed input may add expense or produce negative side effects.

To apply feedback or feedforward control, we need a fixed input that we can change to adjust the output center a prescribed distance. We call such an input a *process center adjuster,* or just an *adjuster.* We use the same tools to find an adjuster or a fixed input to move the process center. With the Move the Process Center approach, we hope to make a single shift. With feedback or feedforward control, we plan to adjust the process often.

In many circumstances, we already know of a fixed input that will change the process center, but we may need to calibrate the size of the effect in order to target the process properly. We can use an experiment to determine the effect of the adjuster.

For many problems, moving the process center is identical to making the process robust. In the shrinkage example, since the output has a physical lower limit of zero, shifting the average and reducing the variation about zero are the same problem and the two approaches are the same. In the shrink defect example, the two approaches are again identical. In each case, we are led to look for fixed inputs that we can change to reduce the defect rate without finding a dominant cause.

Key Points

- To reduce variation, we must change fixed inputs. Without changing one or more fixed inputs, the process performance will not improve.
- We consider seven variation reduction approaches:
  1. Fixing the obvious using knowledge of a dominant cause of variation
  2. Desensitizing the process to variation in a dominant cause
  3. Feedforward control based on a dominant cause
  4. Feedback control
  5. Making the process robust
  6. 100% inspection
  7. Moving the process center
- Finding a dominant cause of variation is an important step in the first three approaches.
- The appropriate choice of variation reduction approach depends on the problem definition, the current state of knowledge about the process, and costs.

Exercises are included on the accompanying CD-ROM

4

An Algorithm for Reducing Variation

Begin with the end in mind.

—Stephen Covey

In this chapter we provide an algorithm to address variation reduction problems such as those described in Chapter 1. The algorithm is structured around the seven variation reduction approaches introduced in Chapter 3.

We believe that variation reduction problems (or any problems) are best addressed using a step-by-step method—that is, an algorithm. There are many advantages to adopting an algorithm. Some people and teams are natural problem solvers that can follow their own instincts. Most of us can use some guidance. We find the algorithm useful because it:

- Is easy to teach (systematic and structured).
- Can be managed.
- Helps to avoid silly mistakes or oversights.
- Provides documentation of progress and success (or failure).
- Helps to ensure that all possible solution approaches have been considered.
- Makes most people, especially when working in teams, better problem solvers.

On the enclosed CD-ROM, we describe in detail three case studies that are successful applications of the algorithm. Here we outline the algorithm and discuss a range of implementation issues. In later chapters, we describe each stage in detail.

4.1 THE STATISTICAL ENGINEERING VARIATION REDUCTION ALGORITHM

We propose the algorithm given in Figure 4.1 to reduce variation in high- to medium-volume manufacturing processes.[1]

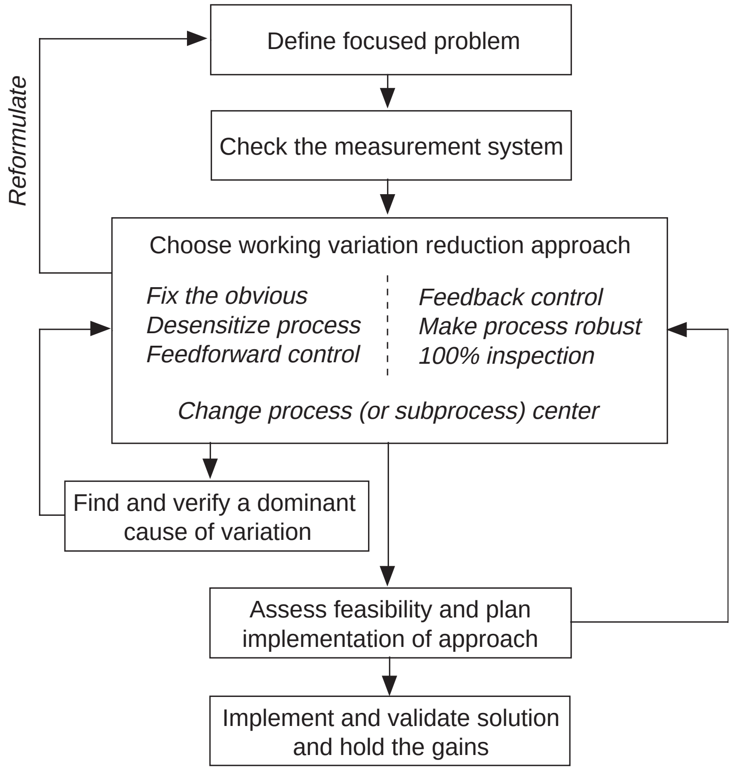


Figure 4.1 Statistical engineering variation reduction algorithm.

Process managers provide the information necessary to start the algorithm. This information should include a specified process, an improvement goal often expressed in monetary terms, and a schedule. The managers also enable people to carry out the work. For convenience, we refer to these people as the process improvement team or the team for short.

In the first stage of the algorithm, Define Focused Problem, the team narrows the process boundaries and selects the particular output characteristic(s) needed to specify the problem. For example, suppose the management statement of the problem is to reduce scrap costs for a particular process by 50% within a month. The focused statement may be to reduce a particular category of scrap, or scrap at a particular processing step. The team determines the nature of the excess variation (off-target, unit to unit) and establishes an appropriate baseline measure of the process performance. They then set a goal for the project in terms of the baseline performance that is consistent with the management goal.

One special feature of Statistical Engineering is the explicit and frequent use of the information provided by the baseline investigation. We establish a baseline to help:

- Set the goal in terms of a particular output.
- Search for the dominant cause.

- Assess the selected variation reduction approach.
- Validate the solution.

We cover the Define Focused Problem stage in detail in Chapter 6.

In the second stage, Check the Measurement System, the team ensures that the measurement system used for the selected output is not home to the dominant cause of variation. They also need to determine if the measurement system is adequate for use in later stages of the algorithm. We assess the measurement system by planning and executing an investigation to determine how much of the baseline variation can be attributed to the measurement system.

If we find that the measurement system is the home of the dominant cause of variation or is not adequate to support future investigations, we reformulate the problem into a new problem to address the variation in the measurement system. We can address this new problem with the algorithm. The process of interest is now the measurement system. We provide several examples of solving measurement system problems throughout the book.

We have seen many failures because a team tried to solve a problem without an adequate measurement system. Less often, we have seen teams become bogged down trying to improve a measurement system that is not the home of a dominant cause and is perfectly adequate for use in further investigations.We cover the stage Check the Measurement System in detail in Chapter 7.

At the next stage of the algorithm, Choose Working Variation Reduction Approach, the team looks ahead to decide how to reduce the variation. We believe that making this step explicit distinguishes this algorithm from the many others that have a similar intent. We have specified seven approaches to reducing variation, described in Chapter 3. At this stage of the algorithm, the team should assess each approach using:

- The nature of the problem and process
- The knowledge they have
- The knowledge required to implement the approach
- The likelihood and cost of obtaining the required process knowledge
- The likelihood of successful implementation
- The probable cost of implementation

The goal is to select a working approach and then direct efforts to determine if this approach is feasible. If it is, the team proceeds to the implementation stage; if not, they reconsider the approaches.

We have divided the approaches into two groups based on whether their implementation requires the identification of a dominant cause of variation. The approaches requiring the identification of a dominant cause are:

- Fixing the obvious (Chapter 14): Use knowledge of a dominant cause to implement an obvious solution.
- Desensitizing the process (Chapter 16): Change one or more fixed inputs to reduce the sensitivity of the output to changes in a dominant cause.

- Implementing feedforward control (Chapter 17): Use an adjustment scheme based on measured values of a dominant cause to anticipate and reduce its impact on the output.

The approaches not requiring the identification of a dominant cause are:

- Implementing feedback control (Chapter 18): Use an adjustment scheme based on measured values of the output to anticipate and reduce the impact of future changes.
- Making the process robust to cause variation (Chapter 19): Change one or more fixed inputs to reduce the baseline measure of variation.
- Using 100% inspection (Chapter 20): Use an inspection scheme to select units with less variation in the output.
- Moving the process center (Chapter 15): Change one or more fixed inputs to shift the process center.

The team must decide at this stage if they will search for a dominant cause. If the answer is yes, then they proceed with the search. If not, they should look at each of the other approaches that are not cause-based and select one.

In most applications of the algorithm, the team will decide to search for a dominant cause. We use the method of elimination to identify this cause. We partition all the causes of variation into families and then use process investigations or available data to rule out all families but one as the home of the dominant cause (see Chapter 9). We use elimination recursively to quickly narrow down the potential dominant causes to one or a few varying inputs. We describe a large number of methods to help search for the dominant cause in chapters 9 to 13.

When the team identifies and verifies a particular input as the dominant cause, they can consider the feasibility of one of the three cause-based approaches. If they rule out these approaches, they have three options:

- Reformulate the problem in terms of the dominant cause
- Reconsider the four non-caused-based approaches
- Search for a more specific dominant cause

If we decide to reformulate the problem in terms of the cause, we start the algorithm over with the goal of reducing variation in the identified input that is the dominant cause. Reformulation corresponds to "moving the problem upstream" or "searching for the root cause." We sometimes reformulate a problem several times. However, eventually we must select one of the variation reduction approaches.

Sometimes a team may reconsider the approaches with only partial knowledge of the dominant cause. They may have eliminated many possibilities but not be able to find a specific cause. With partial knowledge of a dominant cause, they may find one of the non-caused-based approaches more feasible. We describe the Choose Working Variation Reduction stage in detail in chapters 8 and 14.

At the next stage of the algorithm, Assess Feasibility and Plan Implementation of Approach, the team looks at the feasibility of the selected approach. They:

- Examine the process to see if it meets the conditions for the approach to be effective.
- Determine what further knowledge is required.
- Plan and conduct investigations to acquire the knowledge.
- Determine the solution, that is, what and how fixed inputs will be changed.
- Estimate the benefits and costs of the proposed changes.
- Look for possible negative side effects.

We describe assessing the feasibility of each approach in chapters 14 to 20.

If the selected approach is feasible, the team proceeds to validate and implement the solution. Otherwise, they must reconsider other variation reduction approaches.

We arrive finally at the Implement and Validate Solution and Hold Gains stage. Here the team assesses the baseline performance of the changed process to ensure that the project goal has been met. They must also examine other process outputs to make sure they have not created a new problem in order to solve the original one. Finally the team implements and locks the change into the process or its control plan. We recommend the team monitor the process output and audit the process change until they are certain the solution is effective and permanent. As well, the team should document what they have learned and identify future opportunities to reduce variation further.We discuss the Implement and Validate Solutions and Hold Gains stage of the algorithm in Chapter 21.

4.2 HOW TO USE THE ALGORITHM EFFECTIVELY

We have seen the proposed algorithm work well on a large number of projects and fail on others. We believe the key drivers to success can be divided into two groups related to process and management issues.

Process Issues

The process needs to be under reasonable control before starting a project using the formal algorithm. We once were asked to assess the likelihood of success of a project to reduce the amount of rework due to dirt defects on painted bumpers. In a quick walk-through of the painting process with a painting expert, we saw fiberglass bats fall off the ceiling onto the painting line, paint drips everywhere, operators sweeping in areas with unbaked painted parts, and so on. The process was a mess, riddled with poor practices, even to our untrained eyes. We suggested there was little value in using the algorithm with this uncontrolled process. Instead, we gave the simple message:

Fix the obvious!

We think that this is an important message both at the start and during any project to reduce variation. Many times, we have learned after starting a project that scheduled equipment maintenance has been abandoned, sometimes for years, or that operators are ignoring the control plan for the process, often because no one told them that such a document existed. A solution that results from any initial work needed to get a process under control may be thought of as fixing the obvious. In general, we expect to apply the algorithm to processes where:

- A control plan is being followed.
- Equipment is maintained.
- Gages are calibrated.
- Personnel are adequately trained.
- Housekeeping is addressed.
- Industry standard operating principles (see, for example, Todd et al., 1994) are followed.

Quality standards, such as ISO 9000 (Hoyle, 2001) and QS-9000 (AIAG, 1998) provide a mechanism to establish reasonable control. Bhote and Bhote (2000) call this activity *process certification.* We do not require that the process be under statistical control, that is, stable as defined in terms of a control chart, to implement the algorithm. See the discussion about classifying causes in the supplement to Chapter 2 for more explanation.

The algorithm is best suited to address chronic problems (long-standing adverse situations; Juran and Gryna 1980, p. 99) rather than problems that are sporadic. Sporadic problems, where the status quo is suddenly adversely affected, are difficult because they require immediate attention and quick solution. Sporadic problems often lead to containment of product and a corresponding large cost due to logistics, delay, and lost inventory. In this context, applying the proposed stage-by-stage algorithm, with its contemplative nature, is likely not an option. The algorithm has no mechanism for containment. However, many of the specific tools and methods discussed in this book as part of the algorithm are useful to find the cause of a sporadic problem and a solution.

The appearance of a sporadic problem is sometimes used to initiate a project whose goal is to look at both the new sporadic problem and the related long-term chronic problem. For example, a sudden large increase in brake rotor balance rejects from 25% to 50% prompted a process improvement project whose goal was to reduce balance rejects to less than the chronic rate. It is also important to realize that a recurring sporadic problem is best thought of as a chronic problem. Firefighting to address sporadic problems is not effective in the long run. Using the discipline of the proposed algorithm provides greater assurance of finding a permanent solution.

To apply the algorithm, we need to measure process inputs and outputs in a timely fashion. If we are unable to make such measurements, we may not gain the process knowledge required to move through the stages of the algorithm. In a project to reduce warranty costs, the team defined the focused problem in terms of the failure rate of wheel bearings within the warranty period of three years. Their goal was to reduce this failure rate from 3% to less than 0.3%. The key output is the time to failure of the bearing. Other than historical data stored in the warranty database, the team recognized that it would be difficult to apply the algorithm because it takes such a long time to measure the output for any new vehicle. To proceed, they decided to replace the time to failure in the field by a surrogate measure, the bearing failure time measured under extreme conditions in a laboratory.

In another situation, a shipping company set out to reduce the frequency of short and wrong stock shipments to its only customer. There was a lag of up to six months in the customer's reporting of shipping errors and little confidence in the accuracy of the reports. The long lag meant that these measurements were difficult to use for variation reduction. Before the algorithm could be applied, the customer-based measurements were replaced by local measurements based on the results of a daily preshipment audit. It was assumed that changes to improve the process expressed in terms of the audit results would also improve performance for the customer.

To apply the algorithm, we need a high- to medium-volume process that produces frequent parts or units. We assume that we can conduct process investigations relatively quickly and at low cost. If we cannot do so, then the algorithm is likely to fail or we will be very limited in our choice of approaches. In the bearing failure example, each measurement of failure time in the lab was very expensive, so that it was not feasible to carry out many investigations to find the cause of the variation in failure time. The team decided to make the process robust, the only feasible variation reduction approach.

The algorithm is designed to identify low-cost changes to the process or product that will meet the goal of the project. We consider the application of the algorithm a failure if the team cannot identify a low-cost solution. In some instances, the team has a solution in mind (for example, expensive new equipment) and does not consider all of the possible approaches at the Choose the Variation Reduction Approach stage. They do not adequately consider lower-cost changes such as an improvement to the control plan or a change of process setting to desensitize the process to variation in a dominant cause. On the other hand, there may be no cost-effective solution.

We urge you to remember that in most variation reduction problems, the process is currently producing a high proportion of excellent output. Because the process does an excellent job, most of the time, there likely will be a low-cost way to make it even better.

Management Issues

Management makes critical contributions to the successful application of the algorithm. Some specific management tasks are to:

- Choose projects and set reasonable goals.
- Select the process improvement team.
- Provide a supportive infrastructure (training, time, project reviews, and so on).
- Conduct or review cost/benefit analyses for any suggested process/product changes.
- Avoid local optimization in the choice of projects and proposed solution.
- Ensure that organizational learning occurs.

It is difficult for managers to make good decisions in terms of these points if they do not understand the algorithm and how it functions.

The algorithm is project-based, so we want to focus on projects with the potential to have a large impact on customer satisfaction, cost, or both. We know that there are costs associated with the algorithm and that the outcome is uncertain. We can use customer surveys, market research, and warranty data to highlight important customer concerns or desires. We need to link customer concerns to process outputs that can be measured at the manufacturing site; we can use a tool such as quality function deployment (Revelle et al., 1998). We can identify good cost reduction or productivity improvement projects by examining scrap and rework rates and using Pareto analysis (Juran et al., 1979) to rank possible projects. In many cases, other factors will also influence the decision. For instance, management may be planning to update or remove a production process. If such wholesale change is imminent, we should not waste resources addressing identified problems for that process.

When choosing a project (and also later when choosing a potential solution), managers must consider the issue of global versus local optimization. Local optimization may occur when we have a narrow focus and forget that the process of interest is likely part of a much larger system. The output of one process is an input to the next. Reducing variation in an output that has little effect on the downstream process is an example of local optimization. Goldratt (1992) provides an interesting discussion of the local optimization issue (for example, bottlenecks) in the context of cycle times and machine scheduling.

An example of local optimization occurred in the engine block leak example introduced in Chapter 1. The team solved the problem of center leaks by adding several chaplets (small steel inserts) to better support the core. This solution reversed a previous process change. The chaplets had been removed some months earlier to reduce cost. At that time, it was not realized that the cost savings from removing the chaplets would be overwhelmed by increased scrap costs. The truck pull problem described in Chapter 1 provides another example. The problem was initiated because variation in truck pull is noticeable to the customer. However, pull is a derived output characteristic that depends on other outputs such as camber and caster. See Chapter 6 for further discussion. Local optimization would have been a concern had management presented the problem as one to reduce variation in camber and caster. It turns out that camber variation has little influence on pull (relative to the effect of caster). The team could have spent considerable time and effort reducing variation in camber, but the (local) improvement would not have been noticeable to the customer.

The managers must specify the project goal, usually in monetary terms, and provide a schedule. Without a goal, the team will not be able to decide if an approach is feasible, or when the project is finished. The goal must be reasonable; the use of the algorithm will not (often) produce miracles. Resources must be allocated to fit the schedule. A good reference on project management is Lewis (2002).

Teams should include participants with expert knowledge of the selected process. As well, the team must include at least one individual familiar with the algorithm and the associated statistical methods. The core team should be relatively small; we recommend one to three people. Many others will be consulted as necessary to help with specific tasks.

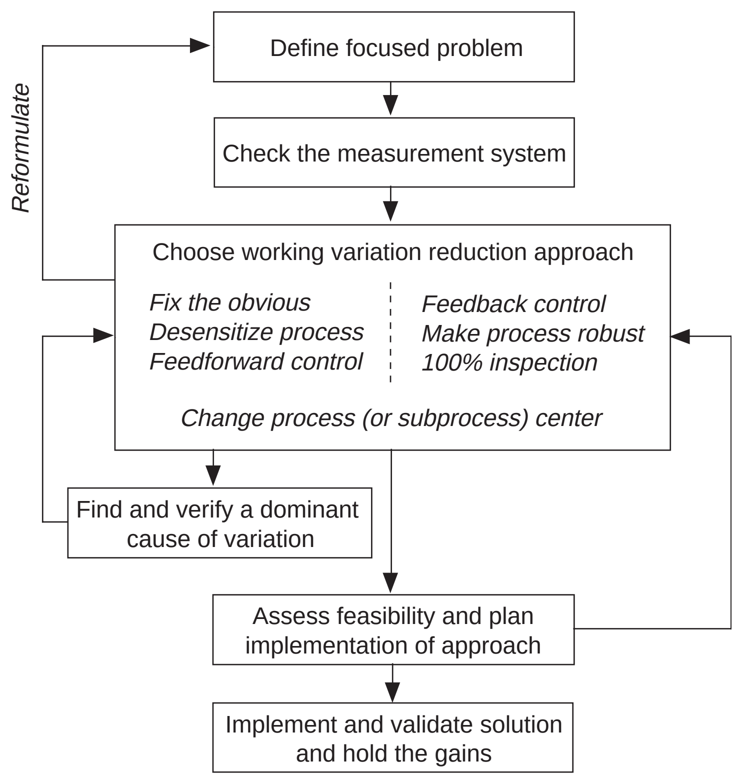
Successful teams must have strong management support during the conduct of the project. Management must provide team members time away from other duties and help in obtaining necessary resources such as testing time. Management can provide training in the use of the algorithm if necessary, can facilitate access to experts, and can provide contacts with customers and suppliers as needed. Management should conduct periodic reviews of the project to keep the effort focused and to provide a mechanism to terminate the project if the likelihood of success appears too small or if costs are too high.

Management also has a strong role to play in weighing the likely costs and benefits of proposed solutions. The team can provide a business case to justify any suggested process/product change. Management approval is necessary to authorize and pay for such changes.

At the conclusion of a project, management must ensure that important lessons learned are disseminated throughout the organization. This may involve updating the corporate memory by changing design guidelines, and so on. A useful source for information on learning organizations is Senge (1990).

Key Points

- We present a variation reduction algorithm specifically designed to produce low-cost improvements to high- to medium-volume processes.



(continued)

- The Choose Working Variation Reduction Approach stage is designed to force the team to consider a wide range of possible solutions in terms of cost, likelihood of success, and prerequisite process knowledge.
- Key drivers to the successful application of the algorithm include the proper selection and management of projects, the provision of necessary resources including people's time, and training.

Endnote (see the Chapter 4 Supplement on the CD-ROM)

1. There are many other algorithms for reducing variation and improving process performance. The Six Sigma algorithm DMAIC (define, measure, analyze, improve, control) is widely used. We discuss a selection of these alternative algorithms and compare them to the algorithm proposed here.

5

Obtaining Process Knowledge Empirically

There is no substitute for knowledge.

—W. Edwards Deming, 1900–1996

New process knowledge is an essential ingredient to the variation reduction algorithm introduced in Chapter 4. Remember, the basis for the algorithm is the idea that we can find a cost-effective solution by better understanding how and why the process behaves as it does.

We use an empirical approach to acquire new process knowledge. Empirical means derived from experiment and observation. That is, we learn about process behavior by carrying out investigations (see Box, 1999). We follow a five-step process (we call this process a *framework* to avoid confusion with the process we are trying to improve) to help plan, execute, and draw conclusions from such investigations. We refer to the framework by QPDAC, an acronym that stands for Question, Plan, Data, Analysis, and Conclusion (Oldford and MacKay, 2001). The purpose of each step is:

Question:	Develop a clear statement of what we are trying to learn.
Plan:	Determine how we will carry out the investigation.
Data:	Collect the data according to the plan.
Analysis:	Analyze the data to answer the questions posed.
Conclusion:	Draw conclusions about what has been learned.

The QPDAC framework highlights the important issues and forces us to think critically about the inevitable trade-offs necessary in designing and conducting an empirical investigation. At the end of the chapter, we give a checklist of issues that you should address in every such investigation. By so doing, you can be assured of solid conclusions at reasonable cost.

We expect to apply QPDAC several times in any application of the variation reduction algorithm. We need to use empirical investigations to increase process knowledge in most of the stages of the algorithm. Examples include:

- Quantifying the problem baseline (Chapter 6)
- Assessing the effectiveness of the measurement system (Chapter 7)
- Finding a dominant cause of variation (chapters 9–12)
- Verifying the dominant cause of variation (Chapter 13)
- Determining the feasibility of the variation reduction approaches (chapters 14–20)
- Validating that the proposed solution has addressed the original problem (Chapter 21)

5.1 QUESTION, PLAN, DATA, ANALYSIS, AND CONCLUSION (QPDAC) FRAMEWORK

In this section, we describe the QPDAC framework in more detail. To make the ideas concrete, we look at an investigation that was part of a project to reduce the variation in the diameter of machined aluminum V6 pistons. The manufacturer selected this project because the cost of internal scrap due to diameter variation was high. At the Define Focused Problem stage of the variation reduction algorithm, the team planned and carried out an investigation to establish a baseline measure of process performance. Maximum diameter (as the piston was rotated) was measured at a specified height above the skirt of the piston.

Question

In the Question step, we specify *what* we are trying to learn about the process. The goal is to pose one or more clear questions that we can address in the subsequent steps of QPDAC. Without clear questions, it is impossible to determine a good plan and draw appropriate conclusions.

In the Question step, we need to answer the following:

- To what group of units do we want to apply the conclusions?
- What input and output characteristics are needed to specify the question?
- What attributes of the process specify the question?
- What is the question of interest?

To introduce some terminology, we call a *unit* an individual realization (or product) of the process under investigation and the *target population* the group of units about which we want to draw conclusions. We sometimes specify the target population in terms of the *target process* that produces the units.

In the example, to establish a baseline performance measure, the team wanted to learn how the piston-making process would behave if it was left to operate normally. A unit was a piston and the target population was all pistons to be produced in the future under the current operating conditions (the target process).

We use the process language from Chapter 2 that defines input and output characteristics. In the example, we are only interested in the diameter of the piston in order to pose the question. In more complex investigations, we will specify a number of input and output characteristics to help define the questions.

To state clearly what we are trying to learn, we specify *attributes* of the target population. An attribute is a function of the characteristics over *all* the units in the target population. In the piston example, the team needed an attribute to quantify the output variation in the target population. They decided to use the standard deviation of the diameters in this set of pistons. Note that none of these pistons had yet been produced.

Attributes can be numbers such as averages, proportions and standard deviations, or pictures such as histograms or scatter plots. We define attributes in terms of one or more input and output characteristics.[1] In many applications of QPDAC, we formulate several questions, so we define several attributes.

We use the selected attributes to specify the question. In the piston example, to establish a baseline measure of process performance, the team asked the specific question:

What is the standard deviation of the diameters of pistons to be produced in the future if we leave the process to operate as it is currently?

After applying QPDAC, the team was able to provide an answer to this question. They then proceeded to look for new ways to operate the process that would reduce the variation. This search involved several applications of QPDAC. When they found a promising new method of operation, they asked and answered a question about another attribute of the target population:

What is the standard deviation of the diameters of pistons to be produced in the future if we operate the process under the new method?

By comparing the two attributes, the team gained valuable information helpful in making the decision about whether or not to change the method of operating the process. The cost of the change and potential side effects also entered into the decision. With knowledge of both attributes, the team had the process knowledge to make a decision to permanently change the operating method.

In the example, the target population contained a large unknown number of pistons, all to be produced in the future. This is the common situation where it is impossible to examine each unit in the target population. As a consequence, we will never know the target population attributes exactly. Our goal in the final four steps of QPDAC is to learn enough about the attributes of interest, subject to the constraints of time and cost, to make good decisions, in spite of the inherent uncertainty.

The outcome of the Question step is one or more clear questions about well-defined attributes of the relevant target population or process.

Plan

In the Plan step, we specify *how* we will answer the questions generated in the Question step. The result of this step is a plan to gather a *sample* of units, to measure a prescribed set of characteristics on these units, and to store the information collected. To get a detailed plan, we need to determine:

- What are the units and population available for the investigation?
- How will we select units to be included in the sample?
- What characteristics of the sampled units will we measure, deliberately change, or ignore?
- For those characteristics that we plan to measure, do we have confidence in the measurement systems?
- For those input characteristics that we will deliberately change, how will we make such changes?
- How will we deal with the logistical issues?

The *study population* is the collection of units from which we can choose the sample. In the V6 piston diameter example, the team chose the study population to be all pistons produced by the process in the next week. That is, they planned to collect the sample over the next week of production.

What were the consequences of this choice? First, the team would take at least a week to complete the investigation. Second, if the standard deviation of the diameters of pistons produced in the given week was different from the corresponding standard deviation in the long-term future (the target population), there will be an *error* in the conclusions. The team could have reduced the likelihood of this *study error* by extending the time of sampling to a month or even longer, but then they would pay a price in terms of time and cost.

We can only suspect that study error might be present. We cannot quantify the error without complete knowledge of all units in both the study and target populations. If we have this knowledge, then the investigation is pointless because we already have the answer to our question. You should always think about the relationship between the target population and a proposed study population in terms of a trade-off between possible study error and cost. Remember that the study population is the set of units from which we will get our sample. Even though we cannot quantify the study error, it is clear that some choices of study population are much better than others. In the V6 piston diameter example, the team would have been unwise to define the study population as the next 100 pistons to be produced, since it is likely that a dominant cause may take longer to vary over its normal range.

In most applications of QPDAC, it is not feasible to examine every unit in the study population because of cost and time constraints. Rather, we collect a sample of units using a *sampling protocol.* The sampling protocol specifies how we select the sample and how many units we choose. The goal of the sampling protocol is to produce a sample of units with attributes that match those of interest in the study population. We define *sample error* as the difference between the sample and the study population attribute. We cannot determine the sample error because we do not know the attribute in the study population.

We have numerous choices for a sampling protocol, including random sampling, haphazard sampling, systematic sampling (for example, sample every 10th unit), and convenience sampling (take what we can get).[2] When specifying a sampling protocol, we need to balance cost and convenience against possible sample error. For a given sampling method, larger sample sizes are more likely to yield a smaller sample error. However, it is more expensive to gather and deal with larger samples.

In the V6 piston manufacturing process, about 10,000 pistons were produced per day for five days in the week. The team decided to commit resources to collect a sample of 500 pistons. The issue of sample size is a difficult one, and the choice is usually driven by resource considerations. For convenience, the team decided to use a systematic sampling protocol in which they would pick every 100th piston. They expected such a sample to give a good representation of the week's production. That is, they thought the standard deviation of the diameters of the sampled pistons would be close to that of all the pistons produced over the week. In other words, they expected little sample error.

Next we decide what characteristics of the sampled units to measure, to ignore, or to change. We must measure or record any input or output characteristic used in the definition of the attribute of interest. However, it is often advantageous to measure additional characteristics. In the V6 piston diameter baseline investigation, for very small cost, the team recorded the time at which each piston in the sample was measured. They hoped that the pattern of variation of diameter over time would give valuable clues to be used later in the problem solving.

We typically ignore most characteristics of the sampled units; this is a conscious choice. In the V6 piston diameter example, the team decided to record the diameter, the time, and the day of measurement and to ignore all other input and output characteristics. They made this choice because of the question of interest. However, as a general rule, if there is little cost involved, we should record other inputs and outputs, especially if we have automated measurement and data collection systems available. We may be able to use these data later in the variation reduction algorithm.

In the Plan step, we specify how we will measure the selected output and input characteristics on the sampled units. The gages, operators, methods, materials, and environment all make up the *measurement system.* The difference between the measured and true value is called *measurement error*. Due to measurement errors, the attribute calculated using the measured values of the sampled units might differ from that using the true values. We need to worry about the contribution of errors from the measurement system to the overall error.

In some investigations, we deliberately change an input characteristic on one or more units in the sample to understand the effect on the output characteristic. If no input characteristics are deliberately changed, we call the plan *observational.* The plan to determine the baseline performance of the piston process was observational since the team did not deliberately change any inputs as the sample pistons were collected. Of course, many inputs changed from piston to piston in the sample. The key point is that the team let these changes occur naturally and did not deliberately manipulate any input.

If we deliberately change one or more input characteristics on the sampled units, we call the plan *experimental.* For example, as part of a project to reduce the proportion of steel stampings that rusted during shipping, a team specially oiled a number of stampings and shipped these parts in the same crates as stampings that were not oiled. The team hoped to show that the oiling significantly reduced the frequency of rusting. This was an experimental plan. The oiling procedure would not have happened without deliberate intervention. In Chapter 13, we look at the details of experimental plans, which are often called *designed experiments.*

We require an experimental plan for many questions, such as the previous example regarding the effect of oiling on rust. To investigate the effects of fixed inputs that do not normally vary, we must intervene and change them. Such interventions can cause difficulty in the production process. For other questions, we may have a choice between an experimental and observational plan. Where possible, we use observational plans because they have the major advantage of not disrupting the current operation of the process. Also, observational plans are usually cheaper to conduct than experimental plans.

A final task in the Plan step is to organize the logistics of the investigation. This can be nontrivial, especially for complex plans. We must consider who, how, where, and when as related to the investigation. For example, we must decide who will collect the sample of units, make the appropriate measurements, and record the data. We must take care to ensure that everyone involved knows what they are supposed to do and that everyone not directly involved knows what they are not supposed to do.

In the V6 piston diameter example, imagine the confusion generated if someone had decided to make an unexpected change (that is, change a fixed input) to the process during the week in which the team was collecting the sample. Remember the question:

> *What is the standard deviation of the diameters of pistons to be produced in the future if we leave the process to operate as it is currently?*

If the fixed input had been changed in the middle of the sampling period, the team would have little confidence in the conclusions from their investigation. Worse, they may not have known that such a change was made.

We have also seen the opposite problem. A team took great care to explain that they were collecting data to establish a baseline and that no process changes should be made. The operators took the team at their word and ignored the adjustment procedure in the control plan until a supervisor noticed that a large number of out-of-specification parts were produced. The team had failed to explain that the current control plan was part of the process and was to be executed as usual.

If the plan is experimental (that is, involves deliberate changes to the process), we must make all interested parties aware of what is happening since we may put customers at risk. The changes may have unexpected consequences, and we must take special care if we plan to ship the product produced during the experiment.

Part of the logistics is to plan for data storage and processing. For small investigations, we can write down the measurements as they are made. Usually we want to store the data electronically for processing. We recommend a row/column spreadsheet format, where each row represents a different unit and each column represents a different characteristic, as described in Appendix A.

The output of the Plan step is, not surprisingly, a detailed plan for carrying out the investigation. In most applications of QPDAC, we consider all the substeps in the plan but not necessarily in the order that they are presented here. As well, we often iterate among the substeps. Sometimes, in the middle of the planning, we are forced back to the Problem step to clarify the question being asked.

In Figure 5.1, we summarize the connections among the target population and the data and show where errors can occur. Our task is to create a plan that is not too complex or expensive and yet controls the potential errors.[3]

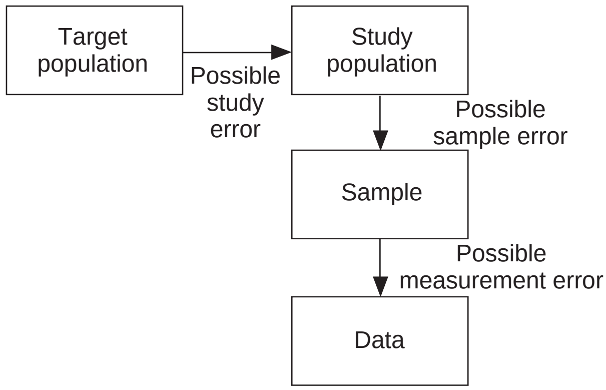


Figure 5.1 Target population to data.

Data

In the Data step, we execute the plan and collect the data. This is often the most time-consuming and costly part of the investigation. This is especially true for those investigations where we paid little attention to the plan, since there is a great opportunity for things to go wrong. In the Data step, we also record any deviations from the plan.

In the example, the team measured 469 pistons and recorded the diameter, day, and hour of measurement. The data were recorded in production order. We give only the last three digits of the diameters in microns. The diameter is measured in millimeters and the deviation from 101 millimeters is multiplied by 1000 before being recorded. The process did not operate on the last part of the shift on day four. The data from the piston baseline investigation are stored in the file *V6 piston diameter baseline*.

Analysis

The goal of the Analysis step is to use the data to answer the question(s) posed in the Question step.

For most of the investigations discussed in this book, we use simple numerical and graphical summaries. However, we also consider some more advanced statistical analysis techniques provided by MINITAB. We introduce these analysis methods and tools as needed in subsequent chapters.

The standard deviation of the sample data in the V6 piston diameter example is 3.32 microns. In Figure 5.2, we show a histogram and run chart of the sample piston diameters.

By plotting the data in several ways, we can detect unusual values (outliers). There is a concern, because many of the numerical attributes such as the average and standard deviation are sensitive to outliers. As a result, any outliers that greatly affect the calculated attributes should be identified.[4] When outliers are present, great care must be taken that the conclusions drawn from the investigation are truly representative of the target population.

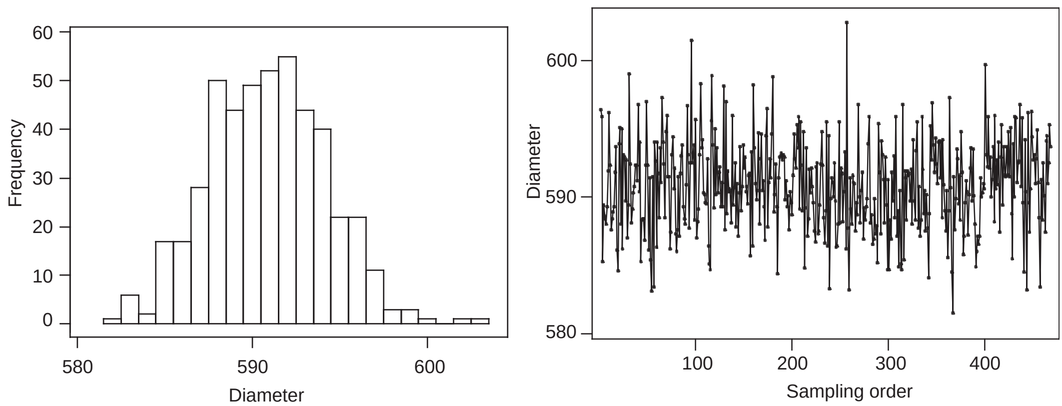


Figure 5.2 Histogram and run chart of V6 piston diameter.

Conclusion

In the Conclusion step, we answer the questions posed about the target population attributes of interest. We also discuss limitations of the answer due to possible errors, both as envisaged in the Plan step and due to deviations from the plan in the Data step. In thinking about errors, remember the three types: study error, sample error, and measurement error.

We use the results from the conclusion step to help us decide what to do next. We need to interpret the conclusion and the associated risk of error in the context of the variation reduction algorithm and the problem itself. To do so, we use basic understanding of the operation of the process, appropriate theory, and knowledge gained in earlier empirical investigations.

In the V6 piston diameter baseline investigation, the estimated process standard deviation was 3.32 microns, which served as a baseline measure of process performance against which the team eventually assessed the effectiveness of deliberate process changes. The team was confident that the estimated standard deviation was a good representation of the long-term variation in the current process, because their experience with the process suggested that in one week, most of the (important) varying inputs that could have changed would have changed. In addition, because of the well-designed plan, they expected that sample error would have little impact.

They specified the numerical goal for the project in term of the estimated standard deviation and decided to proceed to the next stage of the variation reduction algorithm.

5.2 EXAMPLES

Here are two more examples to help reinforce the ideas and language of QPDAC.

V6 Piston Diameter: Comparing Two Measurement Systems

In the process to make V6 pistons, there were gages for measuring the diameter at Operation 270, the last grinding operation, and at Operation 310, final inspection (see Figure 5.3).

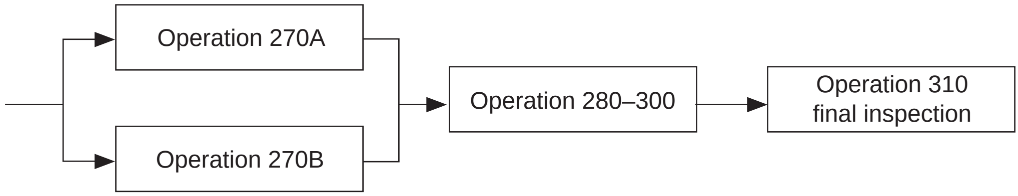


Figure 5.3 Part of a V6 piston machining process.

Operators used the measurement system at Operation 270 to control the two grinders that operate in parallel. The gage at final inspection operated in a controlled temperature environment. Every piston was inspected to ensure that the diameter and several other characteristics were within specification.

During an application of the variation reduction algorithm, the team decided to investigate the relationship between the two measurement systems to help understand the causes of diameter variation at Operation 310.

The team defined a unit to be the act of taking a measurement on a piston. The target population was the set of all such acts that would occur in the future under current conditions. The output characteristic was the measured diameter. One key input was the measurement system used.

To define an attribute useful for comparing the two measurement systems, think of measuring the same piston twice, once with each measurement system. Repeat this over all possible pistons and times (here we think of doing this only conceptually since the number of possible pistons and times is very large). Finally, plot the measured diameters from the Operation 270 gage against the corresponding measured diameter from the final gage. The scatter plot (the attribute of interest) might look like Figure 5.4. Note that the closer the plotted points fall to the 45° line shown on the plot, the better the agreement between the two systems. A point on the line corresponds to equal measured diameters by both systems. Of course, the scatter plot might be very different from the one that is shown.

The specific question was to estimate the scatter plot and determine some of its properties.

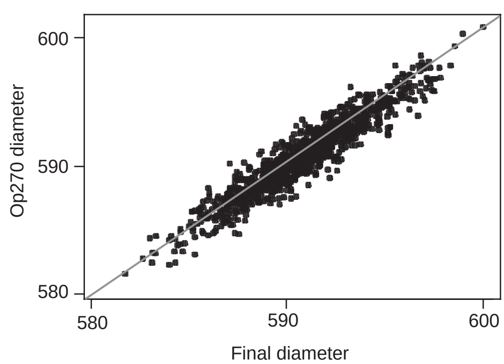


Figure 5.4 Hypothetical target population scatter plot comparing two measurement systems.

The team decided to carry out the investigation the following day. Every hour, the operator at Operation 270 measured four pistons, two from each of the parallel grinders. A designated team member marked these pistons so that they were uniquely identified, recorded the measured diameter, and set the four pistons aside. After six hours, he removed the pistons to the final gage room and let them to come to ambient temperature. During the next shift, the final gage operator measured the 24 pistons and recorded the diameters.

We use the language introduced in the previous section to examine this plan in detail. What is the study population? Measurements could be made on one day. As well, only the 24 selected pistons could be measured with both gages. There is a possibility of study error if the relationship between the two systems changed over time (day to day) or was dependent on the nature of the pistons being measured. To help control this source of study error, the team made a good decision to use pistons from both Operations 270A and 270B.

The sampling protocol specifies which units (acts of making a measurement) will be selected from the study population. Here, the team decided to use 24 pistons and measure each twice at specified times. We can criticize the step in which all 24 pistons were measured by the final gage over a short time. There may be sample error if the 24 pairs of diameters did not represent the relationship between the two gages on the day of the investigation. In terms of sample error, the team could have improved the plan by making the measurements with the final gage throughout the day.

For each piston, the team recorded the diameter from both gages, the time of measurement and the process stream (A or B) at Operation 270. They ignored all other process characteristics such as other dimensional and physical properties of the pistons, the ambient temperature at Operation 270, and so on. The team did not change any inputs, so this was an observational plan.

Since this was an investigation of the measurement systems, there was no need to consider the issue of measurement systems separately.

The logistics of who does what and when were well organized, since one person was assigned responsibility for executing the plan and storing the data in row/column format in a spreadsheet, convenient for subsequent analysis.

The team executed the plan without difficulty. They collected and stored the data in the file *V6 piston diameter gage comparison*.

The analysis consisted of simple numerical and graphical summaries of the data. We show a scatter plot of the pairs of measurements from the sampled units in Figure 5.5 and give the average diameter by gage over all 24 pistons in Table 5.1.

Table 5.1 Piston gage comparison results.

Gage	Average diameter
Operation 270	589.63
Final	591.14

The team concluded there was a strong relationship between diameters measured on the two gages. However, the final gage gave a measured diameter that was systematically higher by about 1.5 microns on average. The difference was small but unexpected. The

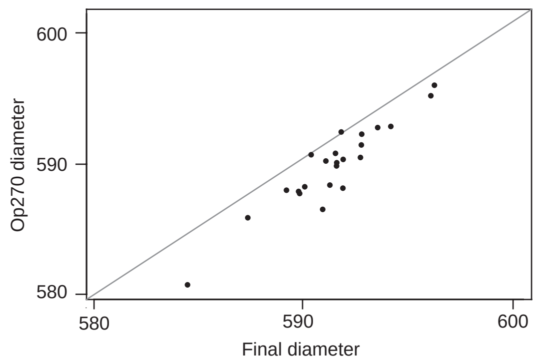


Figure 5.5 Scatter plot comparing sample diameters at the final and Operation 270 gages.

team later explained the difference when they discovered that steel, not aluminum, masters were used to remove the effect of temperature changes from the Operation 270 gage.

The team's only reservation was that the investigation had been limited to one day and there was some concern that the systematic difference between the two gages might become larger on a hotter day.

The team used the knowledge gained to recommend a new calibration procedure for the Operation 270 gage.

Painted Fascia Ghosting

In the painting of two color plastic fascias (the cover over a car bumper), the black prime coat was applied to the whole part. Next, some of the black area was masked with paper taped to the fascia, and then the color coat was added. When the masking was removed, there was occasionally a residual pattern, called ghosting, on the matte black surface under the tape. See the process map given in Figure 5.6.

Customers could detect ghosting since the black surface of the fascia was prominent. The plant could rework the surface to remove the ghosting with added cost and lost production. Management assigned the process engineer the task to eliminate the ghosting problem that occurred on about 3% of the fascias.

The engineer followed the Statistical Engineering algorithm. The process operators visually judged ghosting on a scale of 1 to 5 (1 = no ghosting, 5 = heavy ghosting). The plant reworked fascias with a ghosting score greater than 2 and reluctantly shipped those with a score of 2. The engineer carried out a small investigation to verify that the measurement system was repeatable; that is, the operators were consistent in their assessment of ghosting. Next, after several QPDAC applications, she was convinced that the cause of the problem

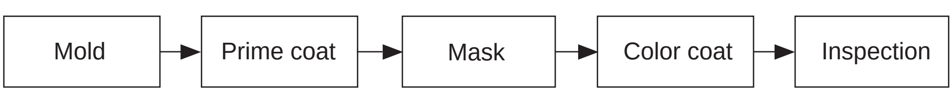


Figure 5.6 Fascia molding and painting process.

was some environmental factor such as ambient temperature or humidity that changed from day to day and could not be easily controlled. More ghosting problems seemed to occur on hot, humid days. She also knew that the ghosting appeared under the tape during the baking of the color coat and that the problem seemed more frequent with certain colors. Two other tape suppliers claimed that their products would not produce ghosting. She decided to investigate whether the other sources of tape were robust to the effects of the environmental factors and color. In the jargon of the variation reduction algorithm, she chose the Desensitize the Process approach. The engineer decided to carry out an experimental plan to compare the performance of the current tape, here denoted *C,* against the two other possibilities, *A* and *B*.

A unit was a fascia and the target population was all fascias to be made in the future. The output characteristic was the degree of ghosting and the key input characteristic was the type of tape used to mask. The three attributes of interest were the proportions of fascias with ghosting level 1 if the tape type was *A, B,* or *C*. The three questions of interest were:

> *What is the proportion of fascias in the future production with score 1 if tape A [or B or C] is used?*

The engineer knew that she required an experimental plan since the tape type was a fixed input that could not change without intervention. Because of the high cost of scrap, she decided to use fascias that had been scrapped upstream of the masking operation for her study population. She also decided to use only those fascias produced on a hot, humid day and to paint these fascias with the color having the most frequent ghosting problem. Furthermore, she planned to mask three small areas, labeled I, II, and III on the primed surface on both sides of each fascia with the three different tapes. See Figure 5.7.

The study unit was one of the small areas on the scrap fascia. The study population was all such units available on the selected day. The study units and conditions under which they were processed were very different from the target population. There was likely to be study error because the conditions were selected to be extreme, so the degree of ghosting seen in the investigation was likely to be high. As well, the engineer knew that she was unlikely to get a large number of scrap fascias for her investigation, so she expected a relatively small sample size.

Accordingly, she decided to revisit the Question, which was changed to the following:

> *Under extreme conditions, what is the average ghosting score for tapes A, B, and C?*

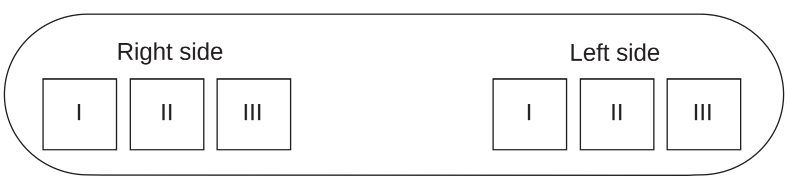


Figure 5.7 Schematic of the study units.

Note that these new questions involved averages, not proportions. The engineer thought that if one tape performed better on average in the extreme conditions, then this tape was likely to produce a higher proportion of fascias with score 1 in the future.

Going back to the Plan step, she decided to use all available scrap fascias (with a maximum of 15) produced on one day. She would chose the day based on a weather forecast. She planned to record the fascia number, the side, the position, the tape type, and the ghosting score for each unit in the sample. She had already established a reliable measurement system for ghosting.

Since this was an experimental plan, she had to decide how she was going to select which tape was applied to each unit. Her plan was to use all three tapes on both sides of each fascia available. Within a side, she would assign the tapes to the three positions at random.

The logistics were critical to the success of the plan. The engineer involved an operator to decide how to mask the small areas. She carefully labeled the tapes to avoid confusion. She prepared 15 schematics like Figure 5.7 that showed which tape went on each position. She planned to give these schematics to the operator one at a time to help him use the correct tape to mask each area in the sample. She would identify each fascia with a number and write it on the schematic. When the operator assessed ghosting, he could write the score on the sheet above each box.

She notified the process owners of the plan to ensure that the experimental fascias would be set aside after painting and to avoid interference with normal production procedures. A time to run the experimental parts could be set up once the day was selected.

The plan was executed without any hitches. Twelve fascias were processed and a total of 72 ghosting scores were recorded. The data were transferred from the schematics to the file *fascia ghosting robustness*. Table 5.2 is a summary of the data collected.

Table 5.2 Summary of ghosting scores.

	Ghosting score					
Tape	**1**	**2**	**3**	**4**	**5**	**Average**
A	16	7	1	0	0	1.375
B	5	9	7	2	1	2.375
C	7	9	6	2	0	2.125

Under the extreme conditions of the plan, tape *A* was clearly superior with a much lower average score. All but one area had a score of 1 or 2. The major limitation was study error; it was possible that tape *A* would not perform better than tapes *B* or *C* under normal operating conditions.

The engineer validated the performance of tape *A* under normal operating conditions with another application of QPDAC. Since the costs of the three brands of tape were similar, management accepted the engineer's recommendation to switch to tape *A*. In the longer term, the proportion of fascias with any detectable ghosting (score >1) fell to less than 0.3%.

5.3 SUMMARY

We summarize the terminology used within QPDAC as follows:

QPDAC Terminology

Term	Meaning
Attribute	Numerical or graphical summary of the characteristics over a collection of units
Experimental plan	Plan where some inputs are deliberately changed on the sample units
Measurement error	Difference between the measured and true value of a characteristic
Measurement system	Gages, people, methods, material, and environment involved in measuring a characteristic
Observational plan	Plan where all inputs vary naturally on the sample units
Sample	Units selected from the study population and measured in the Data step
Sample error	Difference between the attribute in the study population and sample
Sampling protocol	Procedure by which the sample is selected from the study population
Study error	Difference between the attribute in the study and target populations
Study population	Collection of units available for investigation
Target population	Collection of units produced by the target process that we want to draw conclusions about, usually all units produced now and in the future
Unit	A part or the act of making a measurement

None of the individual steps or substeps in QPDAC is difficult. However, pitfalls abound.

In the V6 piston diameter example, the team first established the baseline measure of process variation by measuring the diameters of 50 pistons produced consecutively. The estimated standard deviation was 2.27, suggesting a highly capable process. There seemed to be little reason to carry on with the project. However, everyone involved knew that the process needed improvement. The team had given little thought to the notions of the target and study population or the sampling protocol. Local management recognized the problem and the baseline investigation was redesigned as described in Section 5.1.

Here we provide a checklist for the major steps and substeps for any empirical investigation. Make sure you consider each substep as you plan and execute a process investigation.

A Checklist of Questions for Each Step of QPDAC

Question	What are the units? To what population of units will the conclusions apply? What characteristics are involved in the question? What attributes of the target population make up the question? What question(s) need to be answered? Be specific!
Plan	What units might be included in the investigation? How well does this study population match the target population? How will the sample of units be collected? How well is the sample likely to match the study population? What characteristics will be measured/ignored? Are all the necessary measurement systems adequate? What input characteristics (if any) will be deliberately changed? How will changes be made? Who will carry out the plan? Who needs to know that the investigation is being conducted? How will the data be stored?
Data	Was the plan executed as expected? What deviations occurred? Are the data stored as expected?
Analysis	What are the sample attributes? Draw a picture if possible! Are there any unusual patterns (e.g., over time) of concern in the data? Are there any unusual values (outliers)?
Conclusion	Has the posed question(s) been answered? Are any of study, sample, or measurement errors likely to be large enough to affect future decisions?

Key Points

- We use the QPDAC (Question, Plan, Data, Analysis, and Conclusion) framework to carry out empirical investigations to increase our knowledge of the process.
- With careful use of the QPDAC framework, we can avoid or control study, sample, and measurement errors and generate the process knowledge required for the variation reduction algorithm.
- We expect to apply the QPDAC framework several times in any application of the variation reduction algorithm.

Endnotes (see the Chapter 5 Supplement on the CD-ROM)

1. We can use simple attributes such as averages, rates, standard deviations, histograms, and run charts to establish baselines in the Define Focused Problem stage of the variation reduction algorithm. In the supplement to this chapter we describe some other more complex attributes.
2. There are many different sampling protocols. The most famous, random sampling, plays a small role in the variation reduction algorithm given here. In the supplement, we look at the ramifications of choosing among different sampling protocols.
3. We show the nature of the possible errors in empirical investigations in Figure 5.1. Remember that we want to answer questions about attributes of the target population/process using measured values of characteristics in the sample. In the supplement, we describe study, sample, and measurement errors in more detail.
4. Outliers can have a large influence on the conclusion drawn from an investigation. In the supplement we explore the effect of outliers and describe methods for their identification.

Exercises are included on the accompanying CD-ROM

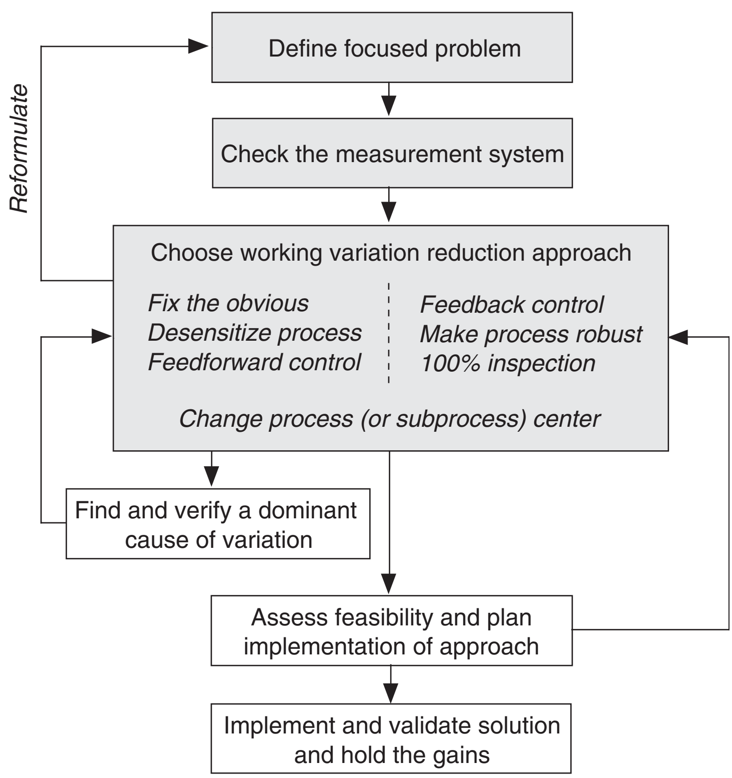
PART II

Getting Started

If you know a thing only qualitatively, you know it no more than vaguely. If you know it quantitatively—grasping some numerical measure that distinguishes it from an infinite number of other possibilities—you are beginning to know it deeply.

—Carl Sagan, 1932–1996

In this second part of the book, we explore the first three stages of the variation reduction algorithm: Define Focused Problem, Check the Measurement System, and Choose Working Variation Reduction Approach. We require a quantitative goal for the problem and the full extent of variation of the output to help design and analyze investigations conducted in the search for a dominant cause or a solution, and to check that a proposed solution meets the problem goal. Next, we look at how to assess and improve measurement systems. A measurement system is necessary to support all process investigations and may itself be a dominant cause of variation. Finally, we discuss how to choose a working approach among the seven variation reduction approaches. In any application, this choice is important because it determines what we do next.



Statistical Engineering variation reduction algorithm.

6

Defining a Focused Problem

Our plans miscarry because they have no aim. When you don't know what harbor you're aiming for, no wind is the right wind.

—Lucius Annaeus Seneca the Younger, 4 B.C.–A.D. 65

There are two major tasks in the Define Focused Problem stage of the Statistical Engineering algorithm. First, the team must translate the project and its goal into a problem statement about excessive variation in a measurable output characteristic. The problem should be specific so that it is likely that there will be only one or two dominant causes. Second, the team must quantify the magnitude of the problem. We require this baseline measure to:

- Help set the goal for the problem.
- Help design investigations and interpret the results as we proceed through the algorithm.
- Allow validation that the problem has been solved.

The baseline is important in all subsequent stages of the algorithm.

6.1 FROM A PROJECT TO PROBLEMS

We start many applications of Statistical Engineering with a project defined in vague terms, and a goal specified in terms of dollars or customer concerns. In Chapter 1, we discussed the block leakers project, where the goal was to reduce the scrap rate of engine blocks that leaked after machining. The management of the foundry assigned a team to reduce cost and improve the relationship with the engine plant, their direct customer. The goal of the project was to reduce the scrap rate from more than 4% to less than 1%.

The team started by turning the project into three problems. The engine plant inspected every block for leaks using a pass/fail test. The team instituted a new test for failed blocks that determined the location of the leak. Using a sample of 100 leaking blocks, they found that 92% of the leakers fell into three classes based on the location of the leak, as shown in Table 6.1.

Table 6.1 Leak classification for 100 leaking blocks.

Class of leak	Percent
Center	32
Cylinder bore	26
Rear intake wall	34
Other	8

The team suspected that each class of leak would have its own dominant cause. Hence they defined three problems based on the three classes. In each case, they set the ambitious goal of eliminating the entire leak class. If they could achieve these goals completely, they would far surpass the project goal.

Some projects are defined in terms of processes that are replicated at different sites or in terms of a class of products. We may choose to focus the problem by concentrating on a single product or manufacturing line. In the truck alignment problem described in Chapter 1, several assembly plants built the same truck using the same components and assembly process. To concentrate resources, upper management assigned the project to a team in one plant. The idea was to apply the same solution to all plants.

In specifying a problem, we may need to define an output characteristic that can be measured locally. A team at an engine assembly plant was charged with reducing warranty claims due to excessive oil consumption. There were only a few claims, but the plant management initiated the project because of the potential damage to the long-term reputation of the engine. To define a specific problem, the team first spent considerable time and effort developing a dynamometer test that could reproduce in the lab the failure mode seen in the field. This effort was necessary because the field failures were so rare. The team specified the problem in terms of the dynamometer measurement. They were confident that if they could reduce oil consumption in the dynamometer test that they could eliminate the field failures.

We try to specify the problem in terms of a continuous output to improve the efficiency of the problem solving. If we use a discrete output, such as pass/fail, we require larger sample sizes for all of the subsequent investigations. For example, in the engine block porosity project described in Chapter 3, the project goal was to reduce the scrap rate from about 4% to below 1%. The team invented a new output, a porosity score based on the size, location, and number of holes on the surface. A block that was scrapped had a high score, but more importantly for the problem solving, every block could be assigned a porosity score that reflected the severity of the problem on a continuous scale. The team defined the problem goal in terms of reducing the variation in the porosity score.

We need to be careful defining the output. For instance, in a problem with seat cover appearance, the team measured shirring on a six-point scale using boundary samples. Scores of 1 to 4 were acceptable, and high scores came from *either* too much or too little shirring. With shirring score defined in this way, the team found it difficult to find a dominant cause of shirring variation, because both high and low values of the cause led to high scores. The choice of output forced the team into the Make Process Robust approach, which failed. See the Chapter 19 exercises for further details.

We need to be able to measure the output characteristic quickly. We sometimes aggravate or accelerate the usage conditions so that the problem occurs sooner. In the oil consumption example, the team created a dynamometer test with aggravated conditions in an attempt to quickly simulate use of the engine under extreme field conditions. There can be considerable difficulties linking the original management project goal to a goal for the output measured using an aggravated or accelerated test. We want to avoid study error where the cause or the solution of the problem under the aggravated conditions is not the same as under the normal conditions. To avoid study error, we need to check that the field failure mode can be replicated using the aggravated test.

We can sometimes use aggravation to deal with a problem with a binary output. For example, in a painting process, there were line defects on the roof of about 10% of the vehicles painted. The team used panels with an increased clear coat film build to make the defect occur more often. Using the panels under the aggravated conditions, they felt they could find clues about the dominant cause more quickly since the defect occurred more often. This advantage had to be balanced against the risk that the dominant cause of line defects on panels with increased clear coat film build (that is, under the aggravated conditions) was not the same as the dominant cause of lines on vehicle roofs with the normal clear coat film build.

To illustrate the transition from a project to problem(s), consider a project to reduce scrap in a process that produced piston rods. Based on past performance, the monthly scrap rate averaged 3.2%. Management set a goal of reducing the rate to less than 1.6%. We show a finished rod in Figure 6.1.

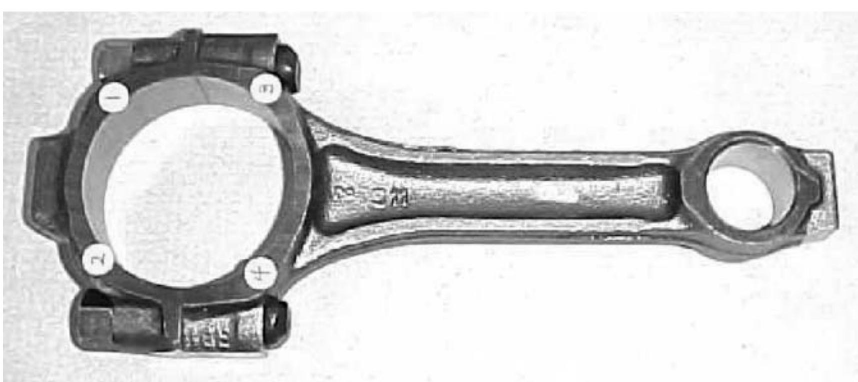


Figure 6.1 A connecting rod.

Looking at process records, the team found that scrap occurred at several processing steps. To focus the problem, they used Pareto analysis on the records from one month, as shown in Figure 6.2. They found that 62% of the scrap was identified at a grinding operation. Further investigation revealed that about 90% of the grinding scrap was due to undersized rods. After grinding, the thickness was measured on every rod at the four locations, marked (with small white circles and faint numbers) in Figure 6.1. A rod was scrapped if the thickness at any location was below the lower specification limit. Rods with thickness above the upper specification limit were reworked. The team focused their attention on reducing variation in rod thickness, a continuous output.

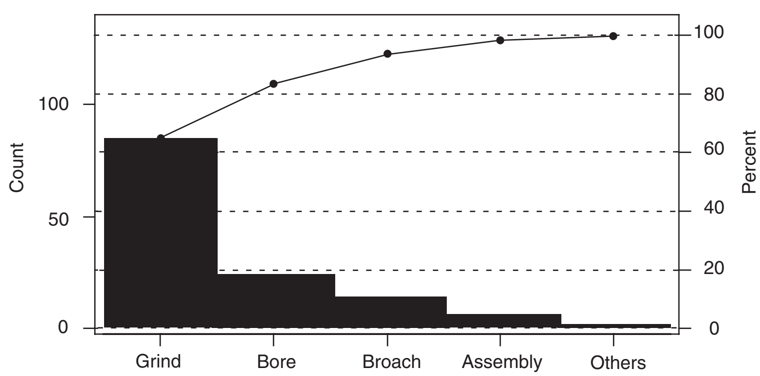
The team set the problem goal to produce all rods within the thickness specification. Achieving this goal would eliminate 90% of the scrap at the grinder, or 56% (90% of 62%) of the total scrap, and hence meet the project goal.

In summary, the key elements in focusing a project to one or more problems are:

- Identify and address the most important failure modes.
- Replace a binary or discrete output characteristic by a continuous one, if possible.
- Define the problem in terms of an output that can be measured locally and quickly.
- Choose the problem goal to meet the management project goal.

If multiple problems arise from a single project, we recommend that the team address the problems one at a time. We need to be careful that, in improving the process with respect to one output, we do not make it worse with respect to another, unless the gain outweighs the loss.

We find it helpful to think about what, where, when, and to what extent the problem exists. We can use a Pareto analysis to help focus the project to a problem that is both narrowly defined and a major contributor to the concern that generated the original project.



Defect	Grind	Bore	Broach	Assembly	Others
Count	85	24	14	6	2
Percent	64.9	18.3	10.7	4.6	1.5
Cum.%	64.9	83.2	93.9	98.5	100.0

Figure 6.2 Pareto chart for rod scrap by operation.

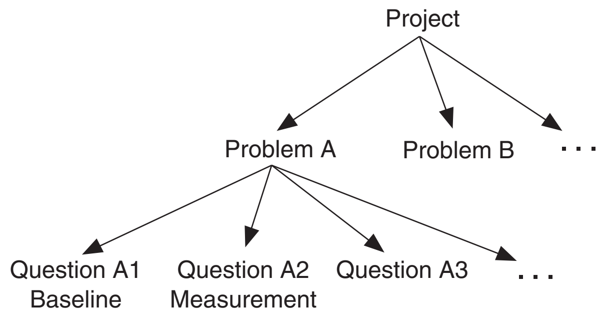


Figure 6.3 Hierarchy of project, problems, and questions.

Figure 6.3 shows the hierarchy we propose to define projects, problems, and questions. Company management initiates projects using concerns regarding quality and cost. The improvement team focuses the project into one or more problems. The team then attacks each problem using the Statistical Engineering variation reduction algorithm given in Chapter 4. In each application of the algorithm, the team will ask many questions about the process behavior. They address these questions using the QPDAC framework described in Chapter 5.

6.2 THE PROBLEM BASELINE

The *problem baseline,* also called the *baseline,* is a numerical or graphical summary of the current process performance. In the language of QPDAC, the baseline is an attribute of the process operating under current conditions. We select a baseline and estimate its value or appearance in the Define Focused Problem stage.

We specify the problem goal by stating how the baseline should be changed. At the Validate Solution and Hold the Gains stage of the variation reduction algorithm, we recalculate the baseline with the changes to the fixed inputs to demonstrate the improvement and to determine whether or not we have met the problem goal. If we have focused the problem, we should be able to translate the change in baseline to progress toward the overall project (management) goal.

We can choose among many attributes for a baseline. If the output that defines the problem is continuous, we may use the average, the standard deviation, a capability ratio such as P_{pk} (see Chapter 2 supplement), a histogram, or a run chart, among many others. For a binary output, we use a rate or a series of rates taken over time, often plotted on a run chart. We summarize the choices in Table 6.2.

We prefer not to use a capability ratio such as P_{pk} since it combines the average and the standard deviation. We use different approaches to change the process center and to reduce unit-to-unit variation, so we prefer to address two problems when we need to change the process center and reduce variation in the output.

Table 6.2 Problem types and corresponding baseline measures.

Problem type	Possible baseline measures
Too much scrap or rework	Proportion of scrap or rework, run chart of proportion scrapped
Process center off target	Average
Unit-to-unit variation too large with two-sided specifications	Standard deviation, histogram, P_{pk}
Unit-to-unit variation too large with one-sided specifications	Average, standard deviation, histogram, one-sided P_{pk}
Measurement	Measurement variation and bias

In the rod thickness example, the team selected a histogram with specifications limits as the baseline. See Figure 6.4. We discuss the investigation to produce this histogram in the next section. The team set the problem goal to reduce variation in thickness so that the histogram would fall entirely between the specification limits 10 to 60.

The data are found in the file *rod thickness baseline*. The height was recorded in thousandths of an inch, measured as the deviation from 0.900 inches. The summary is:

```
Variable         N      Mean     Median     TrMean      StDev    SE Mean
thickness      800    34.575     36.000     34.840     11.023      0.390

Variable       Minimum    Maximum         Q1         Q3
thickness        2.000     59.000     28.000     43.000
```

The standard deviation in the baseline investigation was 11.0. Since the process is roughly centered and the histogram is bell-shaped, the team needed to reduce the standard deviation by about 25% so that six times the standard deviation matches the specification range.

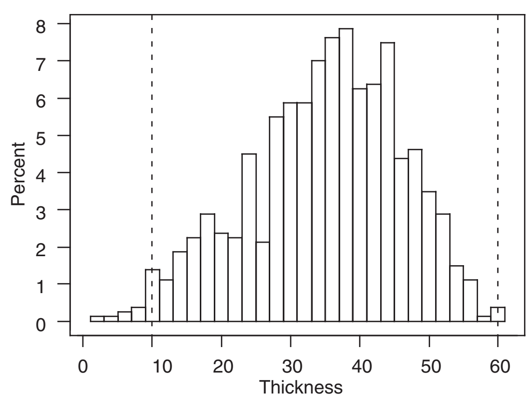


Figure 6.4 Baseline histogram of rod thickness.

There are many choices for a baseline measure. We once asked a process engineer how he could tell if a planned process change would lead to an improvement. After some thought, he told us that he would receive fewer phone calls per week from his customers. In a problem with a system to measure silicon concentration in cast iron, the baseline was the measurement system R&R (repeatability and reproducibility). The problem goal was to reduce R&R to less than 20% of the process variation. See Chapter 7 and its supplement for more information on assessing measurement systems.

As part of establishing the baseline for a continuous output, we determine the *full extent of variation* in the output. We use the full extent of variation in planning and interpreting the results in subsequent process interpretations. When looking at a histogram of the output, we define the full extent of variation as the range within which the vast majority of values lie. The range (minimum to maximum) defines the full extent of variation when the sample size is large (that is, the sample size is in the hundreds) and there are no outliers. For the rod thickness example, the full extent of variation is 2 to 59 thousandths of an inch.

More generally, for a histogram with a bell shape (as given in Figure 2.11), the full extent of variation corresponds to the range given by the average plus or minus three times the standard deviation. Defined in this way, the full extent of variation covers most of the output values.

Sometimes, we use the baseline investigation to generate clues about the dominant cause of the variation. For the rod line example, 40 rods were measured at the four positions each day for five days. We summarize the day-to-day performance using box plots in Figure 6.5.

We see most of the full extent of variation within each day. We can draw two important conclusions from this observation:

- The dominant cause of variation changes within days and we can rule out slowly varying causes that change from day to day or over a longer time frame.
- In subsequent investigations to search for a dominant cause, we can set the study population to be rods produced on a single day.

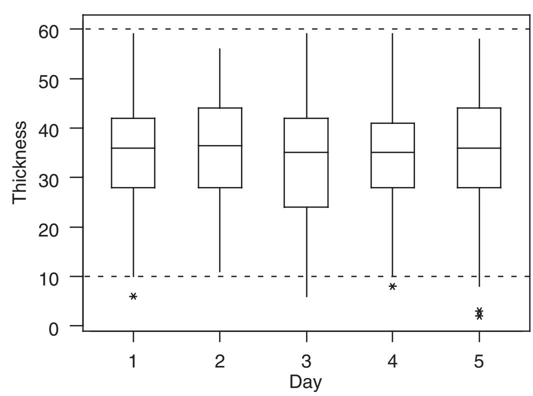


Figure 6.5 Rod thickness by day (dashed lines show specification limits).

6.3 PLANNING AND CONDUCTING THE BASELINE INVESTIGATION

To estimate the baseline, we need to carry out a process investigation. The purposes of the baseline investigation, in order of importance, are to:

- Estimate the baseline for the current process.
- Determine the full extent of variation in the output.
- Generate clues about possible dominant causes.

We need to keep the investigation simple, so we pay little attention to the third purpose.

We discuss common issues in baseline investigations using the QPDAC framework with the rod line scrap reduction problem as the example.

Question

The target population is all units produced by the process now and in the future (assuming no changes are made). When we focus the problem, we specify the output characteristic. In the rod line example, the output is the rod thickness measured at any of the four locations. The attribute of interest (we call this the baseline) is a measure of the process performance related to the goal. In the rod example, the team chose the histogram of the thickness in the target population as the baseline.

Plan

We need to specify a study population from which we will collect a sample of units. The big issue is the period of time over which the study population extends. In the example, the team decided that the study population was all rods ground within a week. They felt that a single week was sufficient to see the full extent of variation in the target population. There is always a trade-off between avoiding study error with a longer period and the cost and time it takes to complete the baseline investigation.

Remember, the goal of the baseline investigation is to understand the process performance as the process currently operates. As a result, process adjustments should be made according to the current control plan.

We choose a sampling protocol that spreads the sample units across the study population; in other words, we sample across the time period selected. In most cases, a combination of a systematic and haphazard method is used. If the output is measured on all units and the data are stored automatically, we can use the whole study population. In the rod example, 40 rods were selected haphazardly throughout each day for five days.

We specify the sample size to be large enough to avoid substantial sample error. As a rule of thumb, we suggest sample sizes of hundreds of units if the output characteristic is continuous and thousands if the output is binary. These sample sizes may seem large, but tables 6.3 and 6.4 demonstrate the difficulty of estimating a small proportion and the advantage of a continuous output.

In Table 6.3, we show roughly how well we can estimate (with 95% confidence) a standard deviation as a function of the sample size.[1]

Table 6.3 Relative precision for estimating a standard deviation as a function of sample size.

Sample size	Relative precision
50	±20%
100	±14%
200	±9%
500	±6%
1000	±5%

In the rod thickness example, we can estimate (with 95% confidence) the baseline standard deviation in the study population to within about ±5% with a sample of size 800.

For a binary output, we cannot give such a simple table since the precision of the estimate depends on both the sample size and the unknown defect rate. In Table 6.4, we give the relative precision (95% confidence) for estimating a proportion defective with a given sample of size.

Table 6.4 Relative precision (%) for estimating a proportion *p* with 95% confidence.

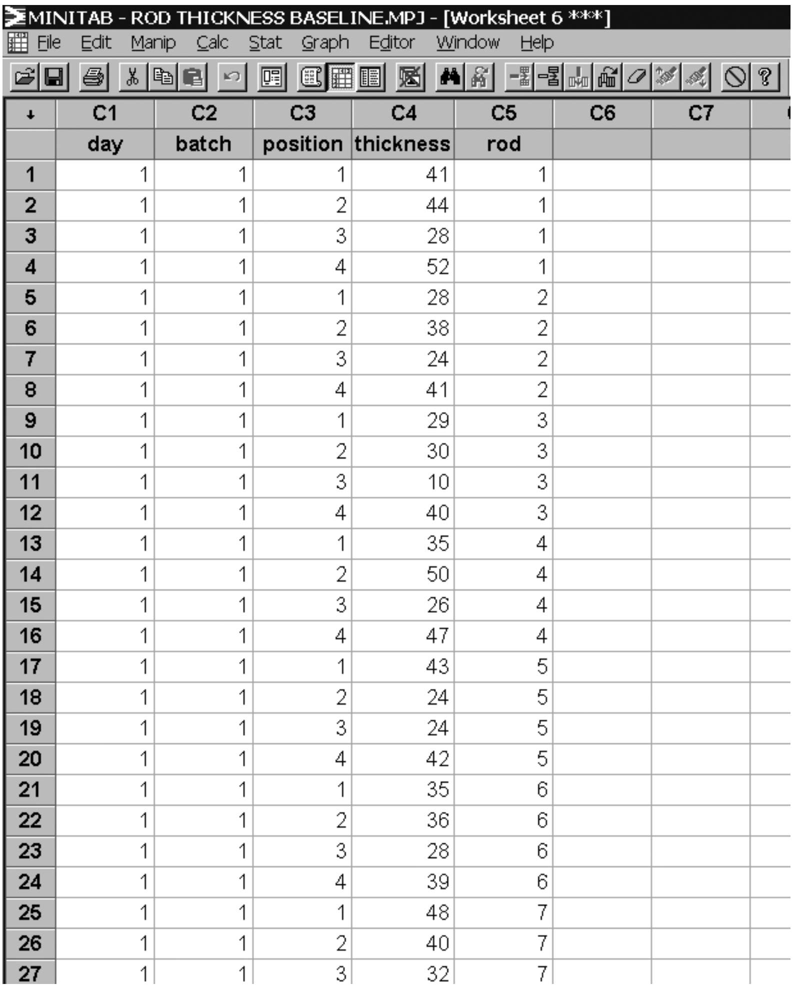
	Sample size			
Proportion defective	**1000**	**2000**	**5000**	**10000**
.05	±27%	±19%	±12%	±8%
.04	±30%	±22%	±14%	±10%
.03	±35%	±25%	±16%	±11%
.02	±43%	±31%	±19%	±14%
.01	±62%	±44%	±28%	±20%

In the example, the known baseline scrap rate is about 2%. Since the team examined only 200 rods in the investigation, the sample proportion of defectives is almost useless as a baseline measure since it is so imprecise. If instead they had used a baseline sample of 2000 rods they would have estimated the proportion to within about ±0.6% (2% ¥ 0.31).

In the baseline investigation, we must measure the output and input characteristics that define the problem. We can decide to measure other characteristics, such as the time the unit is produced, to generate clues about a dominant cause. We do not need to assess the measurement system at this point because we want to include variation due to the measurement system in the baseline.

Data

The plan is implemented in the Data step. In the rod line example, the data were collected according to the plan without incident. The data are given in the file *rod thickness baseline*. We store the data in MINITAB using the row/column format convenient for statistical analysis as shown in Figure 6.6. Each row corresponds to a single measurement, and the columns give the corresponding day, batch, position, thickness, and rod number. See Appendix A for more on row/column data storage.


MINITAB - ROD THICKNESS BASELINE.MPJ - [Worksheet 6 ***]

File Edit Manip Calc Stat Graph Editor Window Help

↓	C1	C2	C3	C4	C5	C6	C7
	day	batch	position	thickness	rod		
1	1	1	1	41	1		
2	1	1	2	44	1		
3	1	1	3	28	1		
4	1	1	4	52	1		
5	1	1	1	28	2		
6	1	1	2	38	2		
7	1	1	3	24	2		
8	1	1	4	41	2		
9	1	1	1	29	3		
10	1	1	2	30	3		
11	1	1	3	10	3		
12	1	1	4	40	3		
13	1	1	1	35	4		
14	1	1	2	50	4		
15	1	1	3	26	4		
16	1	1	4	47	4		
17	1	1	1	43	5		
18	1	1	2	24	5		
19	1	1	3	24	5		
20	1	1	4	42	5		
21	1	1	1	35	6		
22	1	1	2	36	6		
23	1	1	3	28	6		
24	1	1	4	39	6		
25	1	1	1	48	7		
26	1	1	2	40	7		
27	1	1	3	32	7		

Figure 6.6 Row/column format storage of rod thickness baseline data.

Analysis

To quantify performance, we use the sample attribute corresponding to the population attribute selected in the Problem step. In the rod line example, we gave the sample histogram (Figure 6.4) and estimated the standard deviation as 11.0. We also look at other simple numerical and graphical summaries such as box plots and run charts to look for unusual values and patterns in the data.

For example, suppose we collect data as in the rod line investigation and we see a run chart as in Figure 6.7. Since there is an obvious trend, we would worry that the one-week study population was not long enough to capture the full extent of variation in the target population.

Similarly, if there were a large day effect, for example as illustrated by Figure 6.8, we would worry about study error. In this case, five days is not long enough to obtain a good

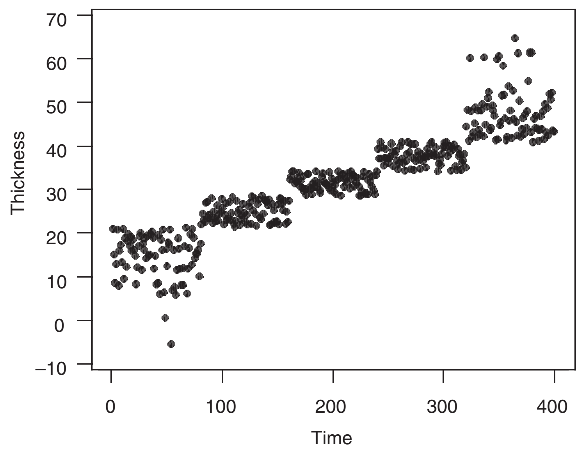


Figure 6.7 Artificial example showing a thickness trend.

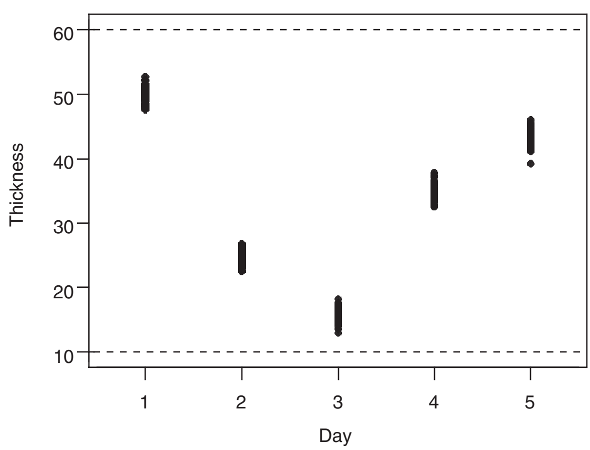


Figure 6.8 Artificial example showing a large thickness day effect.

estimate of the long-term process variation (target population). Here, since the variation within each day is so small, the effective sample size for estimating the overall variation is closer to five (the number of days) than 800 (the number of thickness measurements). With a large day effect, the additional measurements taken each day give information about the within day variation but do not provide much more information about the overall variation.

Conclusion

In drawing conclusions from the baseline investigation, we report the estimates of the process attributes of interest such as the standard deviation, histogram, and run chart. In addition, we note any material limitations due to possible study or sample error. As illustrated by figures 6.7 and 6.8, we may discover these limitations from the sample data. We may also have limitations due to concerns about the plan. If the sample size is small, we worry about sample error. In the rod example, the baseline is given by Figure 6.4. The full extent of variation is 2 to 59 thousandths of an inch.

Baseline Investigation Summary

Question

The purpose of the investigation is to:

- Estimate the baseline, an appropriately chosen attribute of the current process
- Determine the full extent of variation of the output characteristic

The team must select an appropriate baseline—for example, a histogram, a standard deviation, or a proportion.

Plan

- Choose a study population covering a period long enough to see the full extent of variation in the output.
- Determine what outputs and inputs to measure. The inputs should include the time of production.
- Select a sample well spread across the study population with respect to time and other (possibly) important inputs such as machine, position, and so on. The sample size should be hundreds of parts for continuous outputs and thousands of parts for binary outputs.

Data

Record the input and output values with one row for each output value measured (row/column format).

Analysis

- Summarize the data using the appropriate sample performance measure(s). For:
  –a continuous output, use the average, standard deviation, histogram, and run chart
  –a binary output, use the proportion defect and a run chart
- Check for patterns in the output over time (and possibly other inputs).
- Check for outliers.
- Estimate the full extent of variation in the output.

Conclusion

- State the problem and goal in terms of the estimated performance measure(s).
- Determine the minimum time required to see the full extent of variation.
- Consider possible study and sample error.

Comments

We can use existing data to establish a baseline. With the QPDAC framework we can examine the data and how they were collected to ensure that we have confidence that the calculated baseline reflects the true long-term process performance. For example, suppose the output is monitored using a control chart (Montgomery, 1996; Ryan, 1989). We can select an appropriate time period (that is, a study population) from the recent past and use the control chart data over that period. We do not require the process to be stable, as defined by the control chart, to estimate the problem baseline.[2]

Some projects will generate several problems, defined in terms of different outputs. We may be able to plan and carry out a single investigation that simultaneously establishes the baseline and full extent of variation for each problem. We need to be careful to avoid study and sample error for each of the outputs.

6.4 EXAMPLES

We give two further examples of focusing the problem and estimating the baseline.

Truck Pull

We described a project in Chapter 1 where management identified truck alignment as a key issue based on a Pareto analysis of warranty costs and customer quality surveys. They set a project goal to match the performance of a competitor based on these surveys. The same truck was built at several plants. To concentrate resources, the management established a team at one plant to work on the project. They assumed that any remedy found at that plant could be applied to the others.

From mathematical modeling of the truck geometry, the team knew that customers could detect pull, a torque on the steering wheel. Pull is a function of the front wheel alignment as well as many other characteristics such as tire design and pressure, the road camber, and condition. Within the assembly plant, pull was measured as a function of caster and camber only. We have

$$\text{pull} = 0.23*(\text{left caster} - \text{right caster}) + 0.13*(\text{left camber} - \text{right camber}) \qquad (6.1)$$

The specification limits for pull are 0.23 ± 0.35 Newton-meters. The plant measures and records caster and camber on every truck. Any truck with pull outside the specification limits is repaired before release.

To focus the problem, the team started by relating pull as measured in the plant to warranty costs due to alignment problems. They divided the pull specification range into seven classes, each of width 0.10. Using historical data on the alignment characteristics and a warranty database, the team grouped trucks into the seven classes, and for each class calculated the average warranty cost due to alignment issues. Figure 6.9 shows that warranty costs are higher for the extreme classes, corresponding to pull values near the specification limits. This suggests that warranty costs can be reduced by reducing variation in pull. The team assumed reducing pull variation would improve customer satisfaction as measured by the quality surveys.

To establish a baseline, the team selected data from the previous two months. The data are given in *truck pull baseline.* They felt that this time period was long enough so that they would see the full extent of the variation in the process. The histogram for these data is shown in Figure 6.10, with the specification limits given by the dashed lines.

The team also constructed box plots of pull by day as shown in Figure 6.11. We see clear evidence of some day-to-day variation, with drifts in the pull center over time.

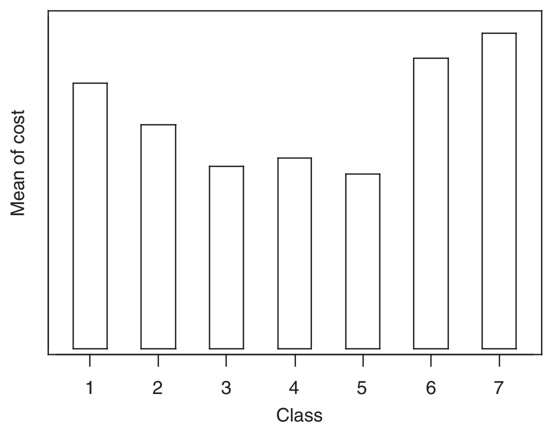


Figure 6.9 Average warranty cost versus truck pull class.

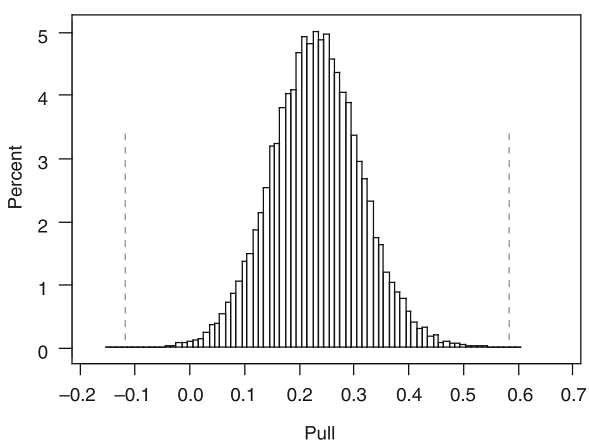


Figure 6.10 Truck pull baseline histogram.

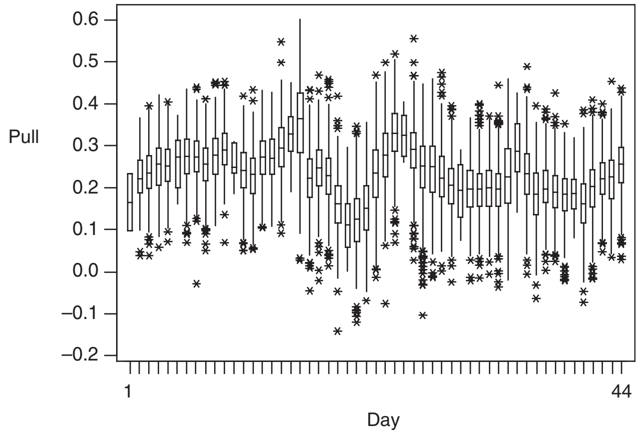


Figure 6.11 Box plots of truck pull by day.

The numerical summary of the pull data is:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
pull	28258	0.23093	0.23084	0.23070	0.08234	0.00049

Variable	Minimum	Maximum	Q1	Q3
pull	-0.14060	0.60120	0.17618	0.28502

The team set the problem goal to reduce the standard deviation of pull from 0.082 to 0.050. If they could achieve this goal, almost all trucks would leave the plant with pull in classes 3, 4, and 5, and warranty costs would be reduced.

To further focus the problem, the team used Equation (6.1). Pull is a simple function of cross caster (difference between left and right caster) and cross camber. Looking at summaries of the cross caster and camber in the baseline data, we have

Variable	N	Mean	Median	TrMean	StDev	SE Mean
cross caster	28258	-0.99982	-0.99600	-0.99795	0.36724	0.00218
cross camber	28258	0.00748	0.01500	0.00991	0.20124	0.00120

Variable	Minimum	Maximum	Q1	Q3
cross caster	-2.73600	0.49800	-1.23900	-0.75700
cross camber	-1.01500	1.08400	-0.11800	0.14100

Using Equation (6.1) and ignoring the small correlations (the correlations between the right and left sides of both camber and caster are around –0.2, and the correlation between cross camber and cross caster is around 0.2), we have

$$\text{stdev}\left(pull\right) = \sqrt{\left(0.23\right)^2 stdev\left(cross\ caster\right)^2 + \left(0.13\right)^2 stdev\left(cross\ camber\right)^2} \qquad (6.2)$$

Since 0.23*stdev(cross caster)* is much greater than 0.13*stdev(cross camber)*, cross caster is a dominant cause of pull variation. Based on this knowledge, the team refocused the problem based on caster. There is little benefit to reducing variation in camber. Using Equation (6.2), if they cut the variation in cross caster in half from 0.367 to 0.185, they could achieve the goal of reducing the pull standard deviation to 0.05. Numerical summaries for right and left caster are:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
l-caster	28258	3.5190	3.5230	3.5204	0.2241	0.0013
r-caster	28258	4.5188	4.5210	4.5192	0.2427	0.0014

Variable	Minimum	Maximum	Q1	Q3
l-caster	2.4490	4.5480	3.3700	3.6720
r-caster	3.0440	5.9380	4.3600	4.6780

The team could accomplish the goal if they reduced the variation in both right and left caster by half. They decided to address right caster first, hoping that any process improvements could also be made to the left side.

We give a histogram of right caster in Figure 6.12. There is a large number of trucks in the data set, so we need to be careful when defining the full extent of variation. The team decided not to use the observed range since there were several outliers, as shown in the box plot in Figure 6.12. Rather, looking at the histogram and the numerical output, they defined the full extent of variation as 3.80° to 5.25° (4.52 ± 3*0.24).

In summary, the team focused the project as defined by the management to reducing variation in right caster from the baseline standard deviation of 0.24° to 0.12°. They connected the management and problem goal through an investigation of the warranty database. They could not connect the problem directly to the customer satisfaction surveys. They established the full extent of right caster variation. They were confident that there was little

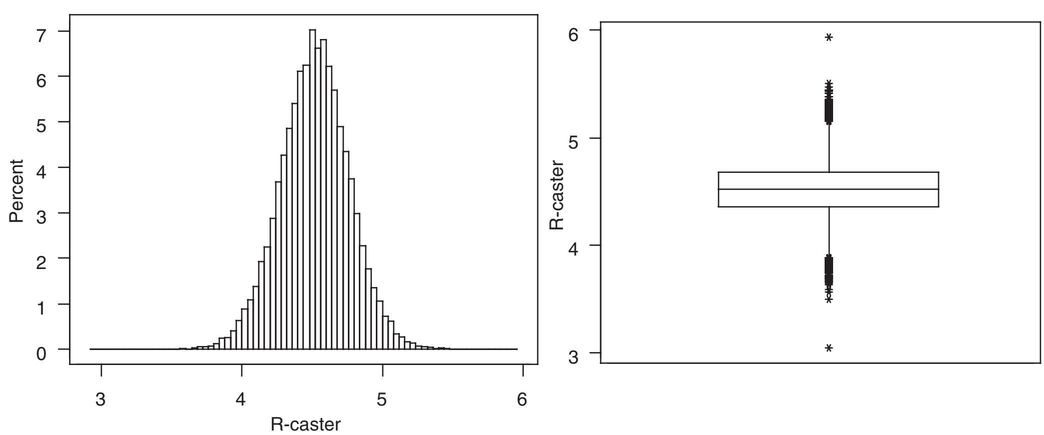


Figure 6.12 Histogram and box plot of right caster from baseline.

study error in their conclusions since they believed two months of data described the performance of the alignment process in the long term.

Pump Noise

A finished vehicle audit at an assembly plant repeatedly detected unacceptably noisy windshield washer pumps. The plant charged for the repairs and put great pressure on the pump manufacturer to solve the problem. The pump manufacturer established a project to eliminate complaints due to noisy pumps.

A team began by developing a noise measurement system that could be used in the manufacturing facility before the pumps were assembled into vehicles. Each member of the three-person team subjectively assessed 24 pumps in vehicles for noise using a five-point scale. The vehicles selected were both acceptable and unacceptable to the vehicle assembly plant. The average score for each pump was recorded. The 24 pumps were then removed and the noise was measured using the new in-plant system. After some adjustments, the team was able to achieve a strong correlation between the subjective human and the in-plant measurement systems (see Figure 6.13). The data from the final measurement investigation are given in *pump noise measurement*. Using the subjective measure, the team judged that a score of 4 or greater was unacceptable to the customer. Accordingly, they set a limit of 8 as the upper noise specification as determined by the in-plant system.

The team was confident that they could detect noisy motors in their facility using the new measurement system. Over a two-day period, they selected 100 motors haphazardly from the current production and measured the noise level. The data are given in *pump noise baseline*. The baseline histogram is given in Figure 6.14.

In the sample, 18% of the motors had a measured value exceeding the new specification limit. The problem goal was to reduce this percentage to 0. The full extent of variation in the baseline using the in-plant measurement system is 0 to 15.

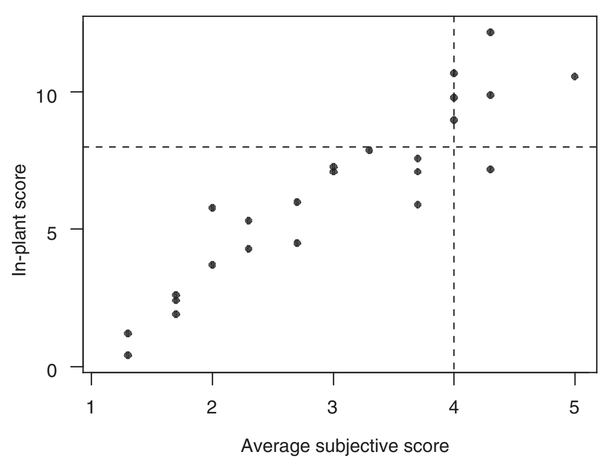


Figure 6.13 Correlation between in-plant and subjective pump noise measurement systems.

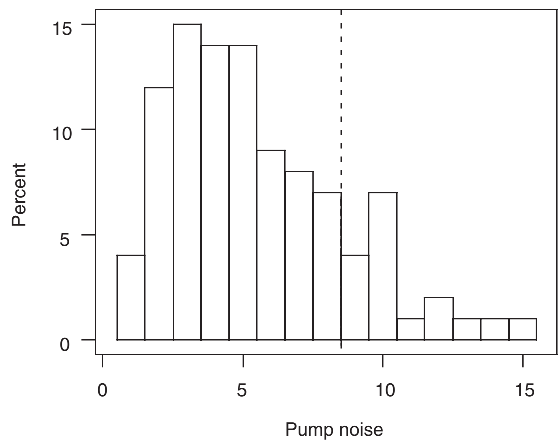


Figure 6.14 Baseline histogram for noisy pumps.

The team did not consider possible study or sample error. They assumed that the sample of 100 motors from a two-day period would accurately describe the long-term performance of the process. They were driven to proceed by the urgent nature of the problem.

6.5 COMPLETING THE DEFINE FOCUSED PROBLEM STAGE

Define a Focused Problem is the first stage of the Statistical Engineering algorithm. The purpose of the stage is to understand and quantify the nature of the problem in a way that makes sense in the production environment. When we start this stage, we have a project description and goal from management, and some prior experience with process. The key tasks necessary to complete this stage are:

- Define a focused problem or problems with an established link to the project goal.
- Estimate the problem baseline, that is, give a graphical or numerical summary of process performance.
- Set a problem goal, stated in terms of the baseline, that matches the project goal.
- Estimate the full extent of output variation in the current process.

The first task may be the most difficult. We must translate a vague management description, such as too much scrap or too many complaints, into specific problems expressed in terms of an output characteristic and the baseline. Since we assume that there is a single dominant cause of the problem, we need to be as specific as possible. We may have to revisit this stage later.

We try to describe the problem in terms of a continuous output. This may not be possible. In many applications, the team invents a new continuous output and a way to measure it to specify the problem. The pump noise problem is a good example.

We need to specify a baseline that we use to define the goal of the problem. We estimate this measure of current process performance using a baseline investigation or existing data. In either case, we must be careful that the time frame for the study population is long enough to give an accurate picture of the current process performance.

We use the full extent of variation to help plan and analyze future process investigations. In cases where there are outliers (unusual values) in the baseline investigation, we ignore them in determining the full extent of variation, unless the outliers define the problem.

The full extent of variation for a binary output is given by the two possible values of the output. The notation of the full extent of variation is not that helpful in planning investigations in this case, other than to suggest we need to examine both defective and nondefective units.

Key Points

- A project should be translated into one or more specific problems, each of which:
  –Has a single dominant cause
  –Can be quantified in terms of a measured output characteristic
- The goal of each problem should be directly linked to the project goal.
- We estimate a problem baseline to quantify the goal, to assess a proposed solution, and to help in the search for a dominant cause and solution.
- For problems with a continuous output, we determine the full extent of variation to help plan and analyze subsequent investigations.

Endnotes (see the Chapter 6 Supplement on the CD-ROM)

1. See the supplement to this chapter for a brief introduction to confidence intervals in the context of estimating the process standard deviation and other attributes. We explain how to use confidence intervals and how to determine them using MINITAB.
2. Does a process need to be stable in the sense of Statistical Process Control before we can establish a baseline performance measure? We consider this issue and firmly answer no!

Exercises are included on the accompanying CD-ROM

7

Checking the Measurement System

When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind. It may be the beginning of knowledge but you have scarcely, in your thoughts, advanced to the state of science, whatever the matter may be.

—William Thompson (Lord Kelvin), 1824–1907

Check the Measurement System is the second stage of the Statistical Engineering algorithm. The purpose is to ensure that we have an effective measurement system for the output characteristic that defines the problem. We assess the measurement system for two reasons:

- The measurement system may be home to the dominant cause of variation.
- We use the measurement system to produce data in subsequent investigations in the variation reduction algorithm.

The measurement system provides a window to view the process. If the window is foggy, we are not able to see clearly what is going on. We must improve the measurement system if we judge it to be inadequate. Sometimes, we can meet the problem goal just by improving the measurement system.

There are many types of output characteristics and measurement systems. For example, the output of the system may be as complex as a force versus time curve, or as simple as a score on the scale from 1 to 5. In this chapter, we describe how to assess a nondestructive measurement system for a continuous characteristic. In the supplement, we discuss variants on the plans to assess measurement systems for binary characteristics[1] and for destructive measurement systems.[2]

7.1 THE MEASUREMENT SYSTEM AND ITS ATTRIBUTES

A measurement system is not just a device or gage. For any characteristic, it includes the:

- Gages, masters, and other materials
- People using the gages
- Procedures for use, including calibration and mastering
- Environment in which the gage is used

In a project to reduce camshaft scrap and rework, the team focused on journal diameter. At the final 100% inspection, the diameter of each of the four journals was measured at two locations. We show the front and rear locations for journal 1 in Figure 7.1.

The inspection gage had eight heads, one for each location. The operator placed the camshaft into the gage. The diameter was defined as the maximum value as the part was rotated through 360°. The specifications were ±12.5 microns, measured from the target value. Camshafts that did not meet specifications were sent to a rework area where operators used a similar gage to aid in repair or to confirm that a part should be scrapped. The two gages were mastered daily using a certified steel cylinder that could be placed in the two heads corresponding to each journal. There was also a yearly calibration procedure.

In this measurement system, there are two gages (or 16 if you count each head separately), a single master, several operators, a prescribed method, and a changing environment in which temperature, for example, was uncontrolled.

We do not get the true value of the characteristic when we measure a part. We call the difference between the true and measured values the *measurement error.* We cannot determine the measurement error since the true value of the characteristic is unknown unless we measure a part with certified known value.

We think of the act of measuring as a process. There is a direct analogy between a measurement process and a production process. Measurement systems produce measurements (the units) rather than parts. In this light, assessing the measurement system is similar to establishing the problem baseline (see Chapter 6).

We need to specify a target population. We plan to use the measurement system to determine process behavior and, perhaps, to monitor and control the process in the future. The target population is all the acts of measuring that we plan to make. For each such act,

Figure 7.1 Front and rear camshaft journal locations.

the output characteristic is the measurement error. We define two important attributes of the measurement system for this target population:

- *Measurement variation* (also called *precision,* or better *imprecision*): the standard deviation of the measurement errors
- *Measurement bias* (also called *accuracy,* or better *inaccuracy*): the average measurement error

We can picture measurement variation and bias as in Figure 7.2. The idealized histogram shows the measurement errors for all measurements in the target population. This is the set of measurement errors if we repeatedly measured many parts over a long time period.

There are many other attributes of the measurement system that may be of interest. For example, do measurement bias and variation depend on the true value of the part being measured? Or do measurement bias and variation change over time? We can define and estimate other attributes to look at these issues related to linearity and stability of the system (Automotive Industry Action Group [AIAG], *Measurement System Manual,* 1995).[3]

For many gages, the supplier provides a standard deviation to indicate the capability of the gage. However, these performance measures are determined under narrow conditions and almost always underestimate the measurement variation as we have defined it over a broad target population with varying parts, operators, environment, and so on.

For many problems, we do not assess measurement bias, especially when the problem baseline is quantified in terms of a standard deviation. If we remove bias from the measurement system and change nothing else, we shift the center of the output characteristic but do not change the variation. If there are two measurement systems within the same process, then we may look at the *relative measurement bias* of the two systems. See Section 7.3.

We start by discussing how to assess measurement variation.

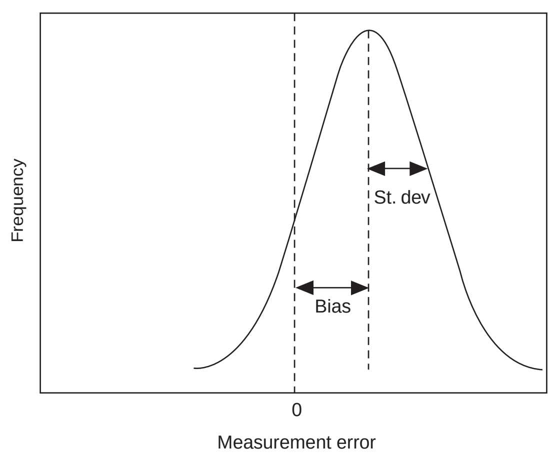


Figure 7.2 Representation of measurement variation (standard deviation) and bias.

7.2 ESTIMATING MEASUREMENT VARIATION

Measurement variation is the standard deviation of the measurement errors over a broadly defined target population. Surprisingly, we can estimate the measurement variation without knowing the true value of the characteristic for any part, that is, without knowing the measurement errors. Suppose that we measure a characteristic of the *same* part several times. For example, suppose we measure the diameter of the same camshaft journal at the same location five times with results

$$1.4, 1.1, 1.3, 0.2, 0.1$$

Since we are repeatedly measuring the same part, the variation in the measured values is due solely to the variation in the measurement errors. We are assuming that the measuring process does not change the true diameter, that is, that the measurement process is not destructive. If the five measurements were taken over conditions matching the target population, we can use the standard deviation of the five values to estimate the measurement variation.

We now describe an investigation to estimate the variation of the measurement system. We use the camshaft journal measurement system as the example restricting our attention to the final inspection gage and the front location on journal 1. In reality, the team looked at all eight locations on both gages.

Question

The purposes of the investigation were to:

- Estimate the standard deviation of the measurement errors over a wide range of parts, a variety of operators, changing environmental conditions, and a long period of time
- Compare the measurement variation to the variation due to the rest of the process

Plan

The first step in the Plan is to define a study population. In other words, what measurements can we possibly take in the investigation? We need to decide:

- Which parts to measure
- Which operators to include
- What time frame to use

We recommend using three parts with values spread across the full extent of output variation seen in the baseline investigation. Here, the team used three camshafts with initially measured diameter values –12.2, 0.9, and 12.8, as shown in Figure 7.3.

If several operators use the measurement system, then we recommend including at least two operators in the study population. If the gage is automated so that it is known that there is no operator impact, then we can use a single operator. In the camshaft example, there were three operators, one from each shift.

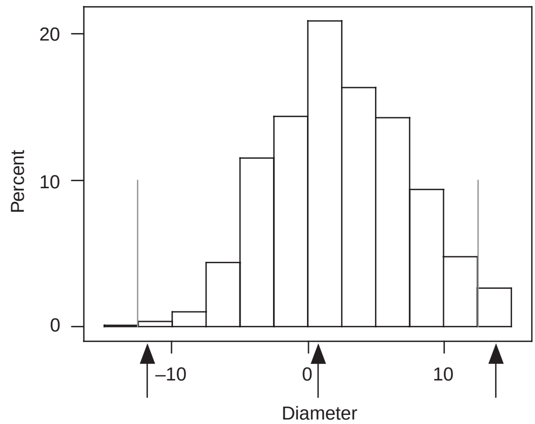


Figure 7.3 Baseline performance for journal 1 front diameter. Arrows show diameter values chosen for measurement investigation.

We can get some guidance from the baseline investigation to help select an appropriate time frame for the study population. Ideally, we assess the measurement system over a period long enough to see the full extent of variation given in the baseline. This ensures that the dominant cause (if it is in the measurement family) has time to act over the course of the investigation. In the example, the team selected a one-week period for the study population. In our experience, most teams make the mistake of selecting a time frame that is far too short because they want to finish the measurement assessment stage quickly.

We need to specify a sampling protocol to determine exactly what measurements will be taken. We plan to measure each part a number of times. We recommend selecting at least two time points within the chosen study period and having each operator measure each part at least twice at each time point.

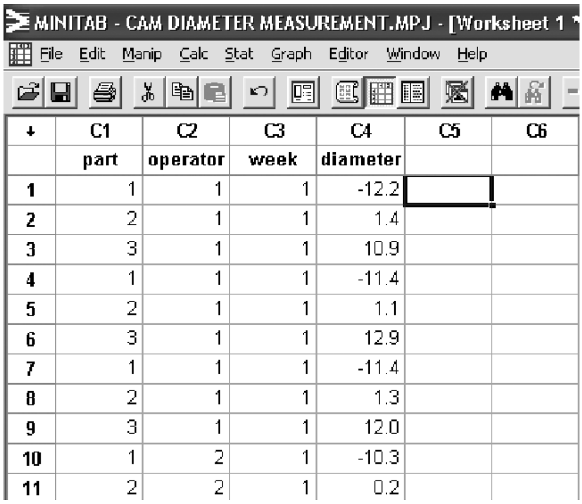
In the camshaft example, two days were chosen, one week apart. On both days, the team planned to have each operator measure each camshaft three times. They would give the parts to the operators in random order for each determination. A team member would record the results. With this plan, each of the three parts is measured 18 times (2 days by 3 operators by 3 determinations). The sample has a total of 54 measurements.

Other than day, operator, part number, and time, the team decided to record no other characteristics. If they had suspected that the attributes of the measurement system would change due to some environmental factor such as temperature, they could record the temperature at each determination. If the measurement system is the home of the dominant cause of the overall variation, the team can use these data to look for clues about why the measurement variation is so large.

The plan should be executed under normal conditions. For example, the operators taking the measurements should use their usual method, and gages and parts should not be specially cleaned.

It is a good idea to randomize the order of measurement within each time period so that the operators cannot remember an earlier result for a particular part characteristic. This point is especially important if there is a subjective element to the measurement system.

There are many other possible plans.[4]


MINITAB - CAM DIAMETER MEASUREMENT.MPJ - [Worksheet 1

File Edit Manip Calc Stat Graph Editor Window Help

↓	C1	C2	C3	C4	C5	C6
	part	operator	week	diameter		
1	1	1	1	-12.2		
2	2	1	1	1.4		
3	3	1	1	10.9		
4	1	1	1	-11.4		
5	2	1	1	1.1		
6	3	1	1	12.9		
7	1	1	1	-11.4		
8	2	1	1	1.3		
9	3	1	1	12.0		
10	1	2	1	-10.3		
11	2	2	1	0.2		

Figure 7.4 Camshaft diameter measurement investigation data in row/column format.

Data

In the camshaft journal diameter example, the team executed the plan. The 54 measured values and the corresponding time, operator, and part number were stored in the file *camshaft journal diameter measurement*. Figure 7.4 shows some of the data stored in the suggested row/column format (see Appendix A) in a MINITAB spreadsheet.

Analysis

We estimate the measurement variation by calculating the standard deviation of the measurements made on each part. The variation in these measurements is due solely to variation in the measurement errors. For the camshaft example the average and standard deviation for each part are:

```
Descriptive Statistics: diameter by part

Variable   part     N     Mean     StDev
diameter   1       18   -10.90     0.638
           2       18     1.20     0.796
           3       18    12.68     0.820
```

We combine the three standard deviations with the formula

$$sd(measurement) = \sqrt{\frac{0.638^2 + 0.796^2 + 0.820^2}{3}} = 0.756$$

to produce the estimate of the measurement variation. We can get this value directly from a one-way analysis of variance (known far and wide as ANOVA) using MINITAB as described in Appendix D.

We can also look at how the measurement variation changes over time, over parts, and over operators. We find this analysis useful to look for outliers (see the supplement to Chapter 5) and to generate clues about a dominant cause if the measurement variation is too large.

In Figure 7.5, we plot the measured values versus the part number (left panel) and, more revealingly, the deviation from the part average versus part number (right panel). We see that the variation due to the measurement system is roughly the same for the three parts and that the measurement system can distinguish among these parts. There are no outliers.

In Figure 7.6, we show the results for the two weeks used in the example investigation. We see no obvious changes in the measurement variation over time. If there were clear differences over time, we would be concerned about study error and would recommend repeating the plan over a longer time frame.

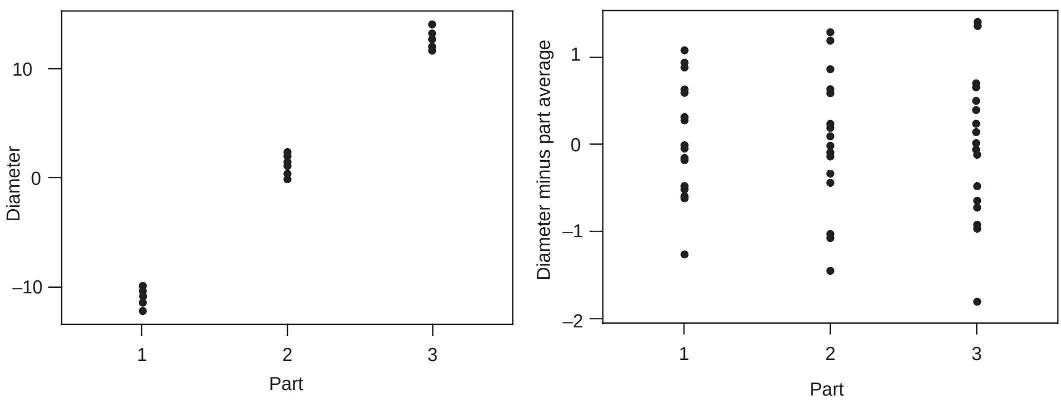


Figure 7.5 Diameter and diameter minus part average by part number.

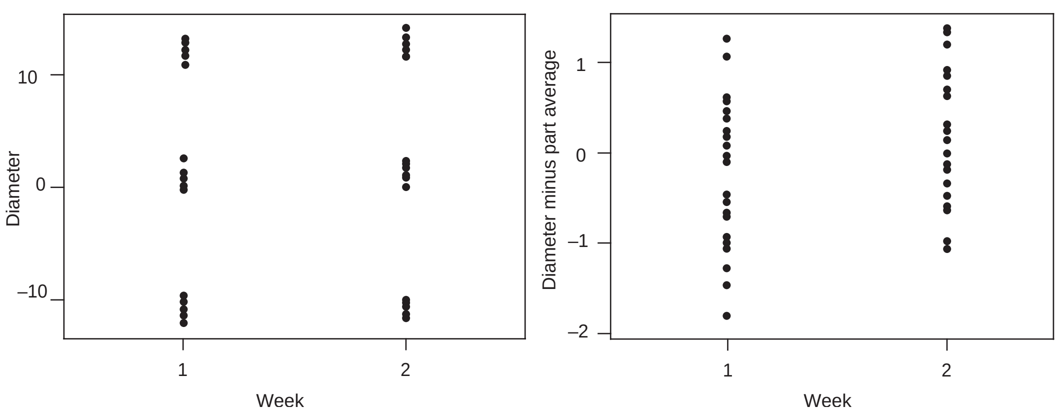


Figure 7.6 Diameter and diameter minus part average by week.

Conclusion

We estimate the measurement variation to be 0.756 microns. We cannot interpret this standard deviation in terms of how close a measured value is likely to be to the true value because we do not know if the measurement system has bias. Perhaps the easiest interpretation is to say that if we measure the same part at two different times with different operators, then the difference in the two measurements is likely to fall in the range[5]

$$\pm 2 \times \sqrt{2} \times \left(\text{estimated measurement variation}\right).$$

In the camshaft example, this range is ±2.14 microns. Since we looked at only 54 measurements, there is possible sample error in the estimate of the order ±20% (see Table 6.2).

To quantify the effect of measurement variation, recall from the model described in Chapter 2 that we can partition the overall variation into two pieces—one due to the variation in the true values of the characteristic and the other due to the variation from the measurement system. That is, we have

$$sd(total) = \sqrt{sd(\text{due to } process)^2 + sd(\text{due to } measurement)^2} \tag{7.1}$$

From the baseline investigation in the Define Focused Problem stage (discussed in Chapter 6), we have an estimate of *sd(total)* that we used as a baseline to define the problem. We labeled this estimate *stdev(total)*. From the measurement investigation, we estimate *sd(measurement)*. Then we can use Equation (7.1) to judge if the measurement system is a major contributor to the overall variation.

In the camshaft example, the estimated total standard deviation is 6.055 (determined from a baseline investigation) and the estimated standard deviation from the measurement system is 0.756. Using Equation (7.1) with the estimates, we can solve for the contribution from the rest of the process, that is,

$$stdev\left(\text{due to } process\right) = \sqrt{stdev(total)^2 - stdev(\text{due to } measurement)^2} \tag{7.2}$$

In the example, *stdev*(due to *process*) is $\sqrt{6.055^2 - 0.756^2} = 6.008$. We see that the measurement system has very little impact on the overall variation.

The effectiveness of the measurement system depends on the relative sizes of the variation due to the process and measurement. We summarize the measurement effectiveness using the *discrimination ratio* D given by Equation (7.3). Larger values of this ratio are better since we are better able to distinguish among parts using the measurement system.[6]

$$D = \frac{stdev\left(\text{due to } process\right)}{stdev(\text{due to } measurement)} \tag{7.3}$$

If D is less than about 2, the measurement system is home of a dominant cause of variation. We should reformulate the problem in terms of the measurement system. In this case, improving the measurement system may solve the original problem.

If D exceeds 3, then we know that the measurement system is not the home of a dominant cause and we can proceed with the next stage of the algorithm and choose a working variation reduction approach.

If D falls between 2 and 3, the measurement system is not a dominant cause, but the measurement system should be improved (see Section 7.5). This recommendation may be ignored depending on the nature of the problem and the difficulty and cost of improving the measurement system. If the discrimination ratio is between 2 and 3, we may have difficulty interpreting the data in future process investigations.

Measurement Variation Assessment Investigation Summary

Question

For all acts of measurement with the measurement system in the future:

- What is the measurement variation, i.e., the standard deviation of measurement errors?
- Is the measurement variation sufficiently small?

Plan

- Ensure that the measurements are made under normal operating conditions.
- Select:
  - –Three parts that cover the full extent of variation in the output
  - –Two or three time periods that cover a period over which you expect to see the full extent of the variation in the output, if feasible
  - –Two or three operators, if multiple operators are normally used
  - –Two or three gages, if multiple gages are normally used
- Make three measurements under each combination of part, operator, time period, and gage.

Data

Record the measured output and corresponding operator, gage, time, and part number with one measurement per row.

Analysis

- Calculate the average and standard deviation of the measurements by part number.
- Combine the within-part standard deviations to estimate the measurement variation.

$$\mathit{stdev}(\text{measurement}) = \sqrt{\frac{\text{stdev(part 1)}^2 + \text{stdev(part 2)}^2 + \text{stdev(part 3)}^2}{3}}.$$

- Plot the measurements and the deviation from part average by the part number, time period, operator, and gage. Look for unusual patterns and outliers.

Conclusion

- Calculate the discrimination ratio,

$$D = \frac{\mathit{stdev}(\text{due to } \mathit{process})}{\mathit{stdev}(\text{due to } \mathit{measurement})}$$

- If:
  - *D* < 2, the measurement system is the home of a dominant cause. Reformulate the problem in terms of measurement variation.
  - 2 < *D* < 3, the measurement system is not the home of a dominant cause but should be improved.
  - *D* > 3, the measurement system is not the home of a dominant cause and is adequate.
- If the measurement system is not adequate, look for clues pointing to the dominant cause of measurement variation.

In the camshaft example, there were eight diameters measured on each part. We may not be able to find three parts with the full extent of variation on each characteristic. Finding and using different parts for each diameter was too complicated, so the team selected parts based on journal 1 only.

We may be able to use the results from recent gage repeatability and reproducibility (R&R) investigations to estimate the measurement variation.[7] However, we must assess the risk of study error since typical R&R investigations are conducted over a short time frame. In many cases, the estimate from R&R substantially underestimates the measurement variation.

7.3 ESTIMATING MEASUREMENT BIAS

To estimate measurement bias we must measure units with known values. We use *known* here in a relative sense. As a rule of thumb, the variation and bias of the measurement system used to determine the known values should be at least 10 times less than the variation and bias of the system under investigation.

The units with known values may be certified standards such as gage blocks or they may be actual parts that have been measured on another measurement system that has low variation and bias. For the camshaft journal diameter system and others that have specialized fixtures, we recommend the use of certified parts to avoid possible study error.

We follow the same plan as we used to estimate the measurement variation. In the camshaft journal diameter example, the original problem was excessive scrap and rework. The rework operators had noticed that some of the parts rejected by the final inspection gage were acceptable when measured by the rework gage. The team decided to assess the bias of both measurement systems. Here we consider only the final inspection system and, as earlier, the front location on journal 1.

There were no certified parts available. Instead, the team used the three camshafts that were described in Section 7.2. For each part, the journal diameter was measured five times on the in-house coordinate measuring machine (CMM). The data are given in Table 7.1.

Table 7.1 Journal diameter CMM data.

	Repeated measurements					Summaries	
Part	1	2	3	4	5	Average	Standard deviation
1	–8.62	–8.35	–8.61	–8.26	–8.68	–8.504	0.19
2	4.24	4.00	4.17	3.93	4.47	4.162	0.21
3	14.64	14.75	14.62	14.20	14.56	14.554	0.21

The team took the average CMM values to be the true values. Note that the variation of the CMM was about 0.20 microns. By averaging five readings, this variation was reduced to 0.09 ($0.2/\sqrt{5}$; see “Use Combined Measurements” in Section 7.5). The team assumed that the bias of the CMM was negligible.

To estimate the bias in the final gage, we calculate the measurement error for each of the 54 diameters measured in the initial investigation. Recall that

$$\text{measurement error} = \text{measured value} - \text{true value}$$

The data file *camshaft journal diameter measurement2* gives the measurement errors for this investigation. We use the average of the observed measurement errors as the estimate of bias.

From the numerical summary that follows, we estimate the measurement bias to be –2.44 microns. The final gage gives consistently smaller measured diameters than the CMM.

```
Variable        N      Mean     Median     TrMean      StDev    SE Mean
meas. error    54    -2.437     -2.434     -2.432      0.866      0.118

Variable       Minimum    Maximum          Q1          Q3
meas. error     -4.455     -0.511      -2.954      -1.780
```

We plot the measurement errors against the part number and week in Figure 7.7. The plot shows that the measured values at the final gage were consistently smaller for all parts and there was no change from one week to the next.

The team considered two possible limitations to this conclusion. First, was there something peculiar about the three parts used or the time during which the investigation was conducted? Since the bias persisted across all three parts (see Figure 7.7) and did not change much over the week, the team was confident that there was a systematic difference between the final gage and the CMM. Second, was the bias due to the CMM? The team, with all members from production, had more faith in their own system that was specially designed to measure journal diameters than the CMM that was an all-purpose gage. Recall that there were eight gage heads to measure the two positions on the four journals in the inspection system. We have given the results for one position on journal 1. The estimated bias for the other seven heads was very small, less than 0.5 microns. Since the same CMM program was used to certify all four journals, the team concluded that the observed bias on journal 1 was not due to the CMM.

Given that the diameter specifications were ±12.5 microns and the process performance used up more than the full tolerance (Figure 7.3), the team knew that the bias was large enough to cause inspection errors. They arranged maintenance on the gage, and a subsequent bias investigation showed that the problem had been corrected. They implemented a once-a-shift check of the system using a reference part (one of the parts certified by the in-house CMM) to indicate if the bias problem recurred.

Note that the plan to investigate measurement bias is identical to that for estimating measurement variation except that we use parts or standards with known values. We can simultaneously estimate both bias and variation from such a plan.

When we have two or more measurement systems for the same characteristic, measurement variation and bias can be the source of acrimony and confusion among the users. We have seen battles between customers and suppliers, each insisting that the value produced by their own measurement system is correct. In one instance, we saw a process in which a transmission part was inspected on four different systems, the last being at the customer. The customer

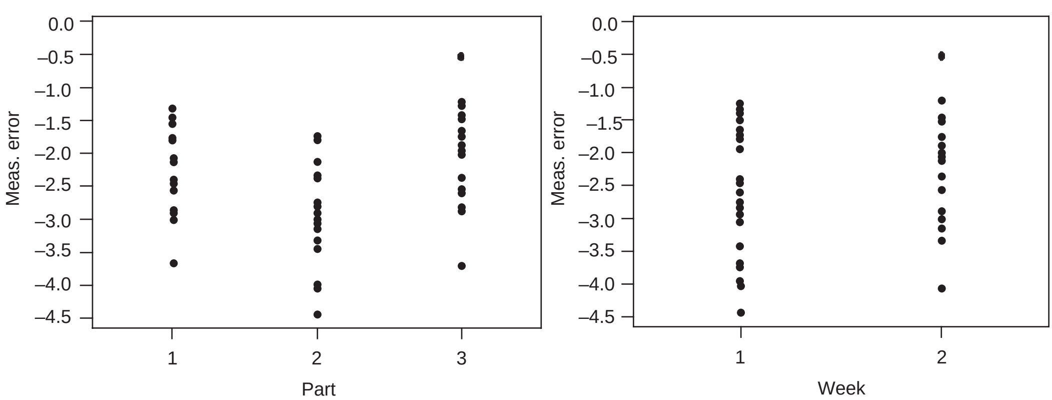


Figure 7.7 Measurement error stratified by part number and week.

occasionally found parts out of specification even though they had been through three upstream inspections. Needless to say there was a lot of finger pointing and high inspection costs.

The *relative bias* of two measurement systems is the difference in bias. We can estimate the relative bias without using parts with known values. We need to measure a number of parts on both systems and compare the results. There are many possible plans. We discussed an investigation to estimate the relative bias of two measurement systems for V6 piston diameter in Chapter 5.

We find it difficult to give a general rule to decide if bias is too large. We need to assess the bias in terms of the original problem, the overall variation, and the consequences.

In the example, given the baseline variation, the team knew that the negative bias in the final inspection gage was large enough that they were shipping oversized parts. The bias also helped to explain why the repair gage accepted parts that were undersized at final inspection.

7.4 IMPROVING A MEASUREMENT SYSTEM

If the variation or bias of the measurement system is too large, we must improve the system to solve the problem or to proceed to the next stage of the variation reduction algorithm. There are several possible strategies.

Fix the Obvious

The first rule, as always, is to fix the obvious. In the camshaft example, the team decided that maintenance of the gage was necessary because the bias was large enough to explain false rejects. They also knew how to implement and use the reference part procedure to detect recurrence of the bias.

In general, we can adjust for bias when it is consistent across parts and time by subtracting an offset (equal to the estimated measurement bias) from the measured value. We need to assume linearity (constant bias for all parts) for this offset strategy to be effective. If the properties of the measurement system change too quickly over time, we may calibrate the system more frequently. Or we may discover that maintenance of the system is long overdue.

Use Another Measurement System

In some cases we may be able to use a better measurement system for the duration of the problem solving. For example, we may be able to use an offline measurement system that has smaller variation but a cycle time that is too large for use in regular production. Another idea to reduce variation in a measurement system with a large subjective component is to use a single operator to make all of the measurements in subsequent investigations.

Alternately, we may invent a new measurement system. In a bottling operation, management raised a concern over crooked and misplaced labels. The label was to be placed on the neck of the bottle a set distance below the bead. In establishing the problem baseline, a height gage was used that measured the distance between the bottom of the bottle and the bottom of the label. As the bottle was rotated, the system measured the maximum and minimum distance and calculated the average height and crookedness (maximum height – minimum height). In the measurement system assessment, the discrimination ratio for average height was 3.1. Although the measurement system was not the dominant cause of the variation, the team believed they would have difficulty in subsequent investigations because of the measurement variation. They decided to use a feeler gage to measure the distance between the label and the bottle bead. The discrimination ratio of the feeler gage system was 6.5. The team felt comfortable proceeding to the next stage of the variation reduction algorithm using the feeler gage.

Use Combined Measurements

We can sometimes combine a number of measurements to reduce bias or variation. For example, if there is a significant bias in a weighing process, to determine the weight of an object we may first weigh the object together with a container and then weigh the container on its own. The weight of the object is the difference. This method eliminates the measurement bias as long as bias does not depend on the true weight. There is an increase in measurement variation.

We can reduce variation by a factor of $1/\sqrt{2} = 0.7$ if we average the values from two determinations on each part. More generally we reduce variation by a factor of $1/\sqrt{n}$ if we average n measurements on each part. Averaging has no effect on bias. For this method to be effective, we assume that the repeated measurement of a part by the same operator over the short term captures the long-term measurement variation in the measurement system. This method was used in the camshaft journal diameter example to certify parts using the in-house CMM. The diameter of each camshaft journal was measured five times. To help ensure that the five measurements reflect the long-term measurement variation, the camshaft was removed from the CMM and refixtured between measurements.

Reformulate the Problem and Use the Statistical Engineering Algorithm

We may find a way to reduce measurement variation using the Statistical Engineering algorithm. The results from the measurement investigation can serve as a problem baseline. As in other applications of the algorithm, to determine the appropriate (measurement) variation reduction approach, we need to examine how the measurement system behaves. We provide examples of applying the various approaches to measurement systems later in the book.

If we record the values of other inputs such as temperature, time, and so forth, in the measurement system investigation, we can use these data to search for clues to the dominant cause of the measurement variation.[8] Again, note the analogy to the baseline investigation. See further examples in chapters 9 through 12.

7.5 COMPLETING THE CHECK THE MEASUREMENT SYSTEM STAGE

The purpose of the Check the Measurement System stage is to ensure the measurement system is not a dominant cause of the variation and that the measurement variation and bias are sufficiently small to support further process investigations. To complete this stage, the key tasks are:

- Estimate the measurement variation and possibly bias.
- Calculate the discrimination ratio, D, as given by Equation (7.3).
- If the measurement variation or bias is too large, reformulate the problem in terms of the measurement system.
- Fix obvious problems in the measurement system.

If we substantially increase the discrimination ratio, is it necessary to repeat the baseline investigation using the altered measurement system?

If $D < 2$, then we reformulate the problem in terms of the measurement system and use the results from the measurement system investigation as a baseline for the reformulated problem. We will reassess the original baseline when we validate the solution in the last stage of the Statistical Engineering algorithm. If $D > 2$, the measurement system is not a dominant cause of variation. Thus, reassessing the baseline is unnecessary because any change to the measurement system will have little impact on the full extent of variation.

Key Points

- We must assess the measurement system for the output characteristic that defines the problem because it may be the home of the dominant cause of variation and because it will be used in future process investigations.
- The key attributes of the measurement system are bias and variation. Bias is the average and variation is the standard deviation of all measurement errors made in the future.
- To estimate the measurement variation, we define a study population over a wide range of conditions, including a variety of parts, operators, gages, and times.
- To estimate bias, we apply the same plan proposed for estimating measurement variation using parts with known values.
- To assess the adequacy of the measurement system, we look at the impact of the bias and compare the measurement variation to the standard deviation of the true values of the characteristic.
- If necessary, for the purpose of problem solving, we may use a different measurement system than that used to define the baseline performance.

Endnotes (see the Chapter 7 Supplement on the CD-ROM)

1. Binary measurement systems that classify units into one of two classes are relatively common. In the supplement we discuss how to assess such systems.
2. We have provided plans and analysis methods to estimate variation and bias for continuous characteristics that are not affected by taking a measurement. In the supplement, we look at similar methods for destructive measurement systems.
3. We consider the AIAG definitions of repeatability, reproducibility, linearity, and stability. We compare these attributes to measurement variation and bias and discuss their usefulness in the variation reduction algorithm.
4. We discuss gage R&R and Isoplots™ (measurement system scatter plots) as tools for estimating the measurement variation. We compare these assessment procedures to that proposed in Section 7.2.
5. It can be difficult to interpret the estimate of measurement variation. We provide an explanation in terms of how large a difference you might see between two measurements on the same unit.
6. We have suggested a criterion for assessing the variation of a measurement system and making decisions on how to proceed. We look more deeply at this criterion and compare it to other standards such as to gage R&R < 30%. For example, we look at the effect of measurement variation on power and sample size and the effect of measurement variation when comparing measured values to specification limits.
7. See note 4.
8. There are many other attributes of the measurement system apart from bias and variation. These become important when we wish to improve a system that is not adequate. We can stratify results by operators, gages, or time. We can use the data from the basic plan for estimating variation to generate clues if we want to reduce the variation in the measurement system.

Exercises are included on the accompanying CD-ROM

8

Choosing a Working Variation Reduction Approach

Change is not made without inconvenience, even from worse to better.

—Samuel Johnson, 1709–1784

In this stage of the Statistical Engineering algorithm, we select a working approach from the seven possibilities described in Chapter 3. You may be tempted to skim over this stage, since our directions are somewhat vague. However, without a working approach, it is not clear what to do next. If you do take further action, you are implicitly adopting a working approach. We believe that explicit consideration of the approach at this point will lead to a better choice and produce better results sooner.

At this stage of the algorithm, we may not know enough to choose the approach that will eventually be implemented. Based on what we do know, using both engineering knowledge and previous process investigations, we select the working approach and then gather the required information to assess its feasibility. So at this point, we are trying to pick the most feasible approach with incomplete knowledge. We may have to return to this stage of the algorithm a number of times as we obtain more process knowledge. We fix obvious faults in the process as we uncover them.

When a team arrives at the Choose Working Variation Reduction Approach stage, they will have the following:

- The problem expressed in terms of an output characteristic and a baseline measure of process performance
- A goal expressed in terms of the baseline measure tied to the project goal
- Some knowledge of the process and its behavior (for example, a process map, a control plan, results of a baseline investigation and other past investigations, applicable science and engineering knowledge, and so forth)
- Some knowledge of the constraints (for example, economic reality, span of control, political and cultural constraints)
- Confidence in the measurement system unless the problem is defined in terms of this system

We return to the question of choosing a working approach in Chapter 14, where we specifically discuss the consequences of having searched for, and hopefully found, a dominant cause. In this case, the team will have acquired considerable new knowledge of the process behavior.

We have the following options at this point in the algorithm:

- Search for a dominant cause of variation. Identifying a dominant cause is a prerequisite for reformulation or the cause-based approaches:

  –Fixing the obvious

  –Desensitizing the process to variation in a dominant cause

  –Feedforward control

- Assess the feasibility of one of the non-caused-based approaches:

  –Feedback control

  –Making the process robust

  –100% inspection

  –Moving the process center

We consider each approach based on the requirements for that approach and our current knowledge of the process. We also consider the likely implementation costs, classified into three broad categories:

- Costs related to determining if the approach can be effective
- Costs of the change to the process to implement the approach
- Ongoing costs to maintain the approach

We incur the first type of cost when we search for a dominant cause, when we try to find process settings that are robust or less sensitive to variation in a dominant cause, when we look for an adjuster to change the process center, and so on. We incur initial costs such as changes to the product or process design, changes to the control plan, training of personnel, and so forth when we first implement an approach. Finally, we incur ongoing costs such as the measurement and adjustment costs needed for feedback or feedforward control and inspection.

At this first iteration of the Choose Working Variation Reduction Approach stage, we are unlikely to have sufficient process knowledge to make a good assessment of the implementation costs. We consider implementation costs for each approach in detail in chapters 14 through 20.

The best option is always problem specific and we find it difficult to give precise directions. Instead, we pose a series of questions that the team should ask.

8.1 CAN WE FIND A DOMINANT CAUSE OF VARIATION?

We need to identify a dominant cause of the variation in order to implement one of the cause-based variation reduction approaches. We strongly recommend that the team search for a dominant cause and then decide which of the three to adopt as a working approach. We have seen many problem-solving efforts flounder because the team tried to change the process without first identifying the dominant cause of the variation. If we know a dominant cause, we have a much better chance of finding a cost-effective process change that will meet the problem goal.

Despite our strong preference for trying to find a dominant cause, in some problems the team may know enough to answer no to this question.

Manifold Blocked Port

In the casting process that makes exhaust manifolds, described in Chapter 2, management adopted a project to eliminate a defect called "blocked ports." The foundry customer detected this rare defect only a few times per year. At each occurrence, there was a loss of goodwill and an expensive containment effort that typically found no other defective manifolds.

The project team knew that observing the process to find the cause would likely be fruitless since there were very few blocked ports. Changing fixed inputs to see their effect was equally hopeless. They decided to look at the feasibility of automated 100% inspection to check every manifold for blocked ports.

Electric Box Durability

The management of a firm that built metal boxes for exterior electrical components decided to increase the durability of the boxes to satisfy its customers, large utility companies. The management had received numerous complaints about premature rusting on boxes in service.

The team considered looking for the cause of the variation in durability. They could trace field failures to only a few processing conditions, such as the lot of paint. They could not determine the values of most varying inputs for any particular box in the field. The team also suspected that the dominant cause of variation was environmental and outside of their control. Finally, they had little confidence in the measurement of the failure time as determined by the utility companies.

The team could measure durability in-house by painting standard panels and using an accelerated test in a salt spray chamber. They decided not to search for a dominant cause of durability variation and instead considered process changes that could increase average durability in the salt spray chamber. That is, they adopted Move Process Center as the working approach.

Comments

One other reason for not looking for a dominant cause is historical. If many teams have failed to find the cause in previous projects, it may be inefficient to keep searching.

Having recommended looking for the cause in almost all cases, we believe the team should consider the four non-caused-based approaches before reaching a decision to search for a dominant cause. Based on current knowledge and potential costs, the team may see that one of these approaches is feasible.

8.2 CAN WE MEET THE GOAL BY SHIFTING THE PROCESS CENTER WITHOUT REDUCING VARIATION?

In other words, should we adopt the Move Process Center approach? For example, if the baseline histogram matches one of those shown in Figure 8.1, and the goal is to reduce the proportion of parts out of specification (specification limits shown by the dashed lines), then we can consider this approach.

To make the approach effective, we need to find a change to some fixed input that will shift the center of the process without high cost or large negative side effects. We also must be able to maintain the change in the long term.

Sand Core Strength

In Chapter 1, we introduced a problem of excessive breaking of sand cores in a foundry operation. The baseline histogram of core strengths is reproduced in Figure 8.2.

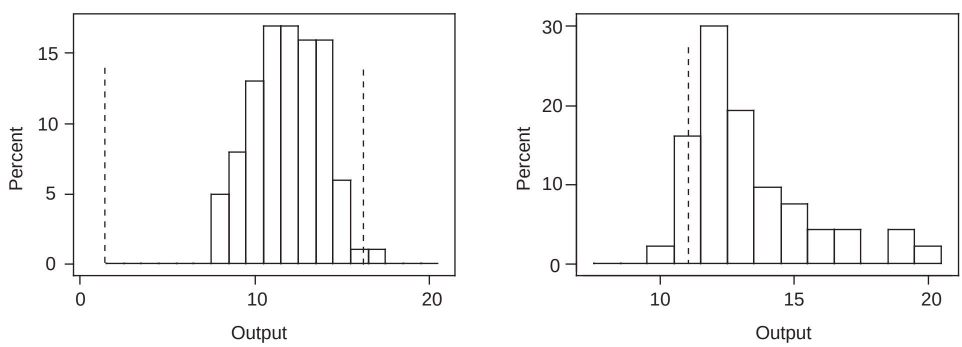


Figure 8.1 Two cases for shifting the process center.

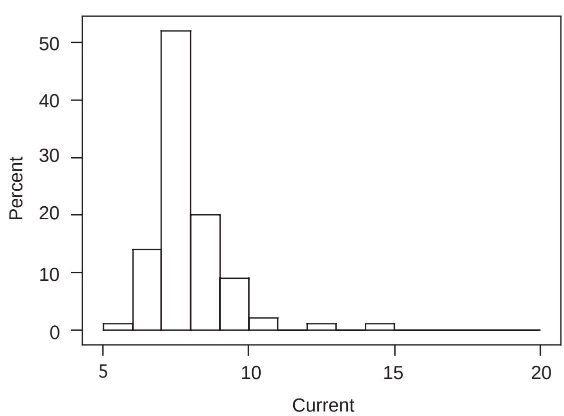


Figure 8.2 Baseline histogram of sand core strength.

The team knew they could increase average core strength by increasing the resin concentration in the core molding process. They could easily assess the cost of the change and maintain it in the future. The team did not know how much extra resin was needed and how well the stronger cores would behave in the casting process. They worried that it would not be possible to shake out the core sand if the cores were too strong. Residual core sand stuck in the casting is a serious defect.

Because of the available knowledge, the team adopted Move Process Center as their working approach. They planned an experimental investigation to quantify the effects of increasing the resin concentration on both core strength and casting quality.

Wheel Bearing Failure Time

A team was assigned a project to reduce warranty costs and customer complaints due to failure of a wheel bearing in a light truck model. The failures occurred within the first three years of use. Using the warranty database and a record of customer complaints, the team focused the problem to the failure of a seal under harsh driving conditions. They developed an accelerated test in the laboratory that could reproduce the field failure mode within a few hundred hours of testing. The results of five tests to failure for the current design are shown in Figure 8.3.

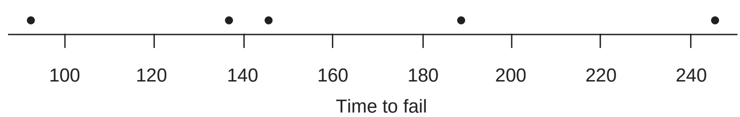


Figure 8.3 Wheel bearing seal failure times.

The team decided to look for a design change that would increase the failure time on the accelerated test to at least 200 hours for all seals. That is, they adopted Move Process Center as their working approach. They rejected the approaches based on finding the cause of the failure (what is different about the seals that explains the different failure times seen in Figure 8.3) because it was too expensive to collect the necessary data. Based on their knowledge of the failure, the team made a design change that they then tested by repeating the five-piece investigation. They accepted the risk that the proposed design change might prove ineffective in the field, where conditions were different from the accelerated test.

Camshaft Lobe Runout

We discussed the problem of base circle (BC) runout of a camshaft lobe in Chapter 1. The baseline histogram is shown in Figure 8.4.

The goal of the project was to produce a greater proportion of lobes with smaller runout. Recall that runout must be greater than zero and is a measure of variation itself. Defining the problem in terms of another output characteristic that is not a measure of variation would have been preferred. The team did not expect to directly discover a change to

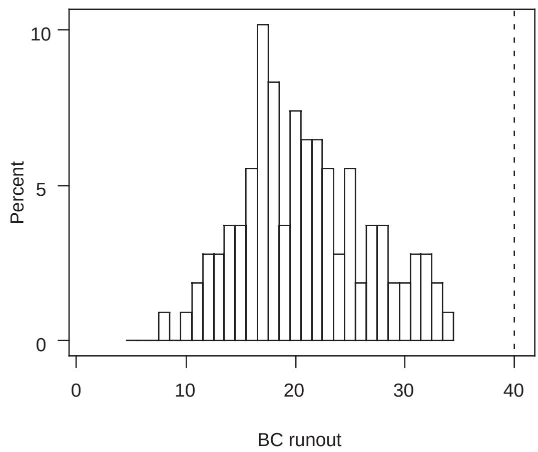


Figure 8.4 Baseline histogram of camshaft lobe BC runout (dashed line shows specification limit).

the process that would shift the histogram to the left. That is, they initially decided not to investigate one or more process changes that might meet the goal. This had been tried in the past without any substantial progress. Instead they decided to look for the dominant cause of lobe-to-lobe variation. Once the cause was found, they could reconsider which approach to select. They were willing to accept the investigation costs with no certainty that they could find the cause or that they would adopt one of the cause-based approaches.

8.3 CAN WE REDUCE VARIATION BY CHANGING ONE OR MORE FIXED INPUTS WITHOUT KNOWLEDGE OF A DOMINANT CAUSE?

If we answer yes, we adopt the Robustness approach. For this approach to work, we need to find a new setting for some fixed input or inputs that reduces the effect of the unknown dominant cause of variation. This approach is often adopted out of desperation after a search for the dominant cause has been fruitless. Without knowledge of the dominant cause, it is difficult to know what fixed inputs to consider.

Transmission Shaft Diameter

In a grinding process for an output shaft, a transmission component, the team set out to reduce shaft diameter variation. They established the baseline performance using process control data (every 50th part was measured) from the previous month. The team assessed the measurement system and found that it was not a major contributor to the overall variation. Having just returned from a course in designed experiments (taught by one of the authors), they decided to run an experiment with 16 runs in which six fixed inputs were varied. The inputs included feed rate, speed, coolant concentration, and so forth. Each experimental run consisted of grinding five shafts. That is, the team selected robustness as their working approach.

For each run the team calculated the logarithm of the standard deviation of the five diameters and then analyzed these 16 performance measures. They found the combination of levels of the fixed inputs that produced the smallest value of the performance measure.

The team then decided to confirm the improved settings of the fixed inputs by running these levels for one week under otherwise normal production conditions. They found no improvement in the process variation and abandoned the new settings because there was an increase in cost.

The team was disappointed in designed experiments as an improvement tool since they had spent a great deal of effort for no reward. What went wrong? The major problem was that the experiment set out to reduce short-term diameter variation (variation within five consecutive shafts), and it turned out that this was a small component of the baseline variation. The team adopted the working approach without sufficient knowledge. In this case, the project was doomed by this decision.

Speedometer Cable Shrinkage

In a famous positive example of making a process robust, Quinlan (1985) reported on a project to reduce both the average and variation in postextrusion shrinkage of a casing for a speedometer cable. The team did not find the cause of the shrinkage variation. Instead, they chose 15 fixed inputs and selected one new level for each (for each selected fixed input, there were two levels, existing and new). They then ran an experimental investigation with 16 runs, a highly fractionated design. In each run, approximately 3000 feet of casing was produced and four samples were selected haphazardly. The shrinkage was measured for each sample.

Here the team adopted the Move Process Center and Robustness approaches simultaneously, since reducing variation or shifting the process center could improve the process.

For each run, the team calculated a performance measure using the four shrinkage values. They then analyzed the performance measures to isolate the best combination of levels to shift the average shrinkage towards zero and reduce the variation. The new levels were confirmed and the process was much improved.

Comments

As shown by the speedometer cable shrinkage example, the robustness approach can be successful, but we feel it is better to first look for the dominant cause. The more you know about the dominant cause of variation, the greater the chance you will select fixed inputs to change that will mitigate the variation in the dominant cause.

In the transmission shaft example, the team would have been much better off if they had recognized that the dominant cause of variation acted over a longer time frame than five consecutive parts. They would not have planned the experiment as they did with this extra knowledge.

8.4 DOES THE PROCESS OUTPUT EXHIBIT A STRONG PATTERN OVER TIME?

If the answer is yes, we can predict future output values from present and recent values. If the team also knows of a cost-effective quick way to adjust the process center to compensate for the deviation between the prediction and the target (or feels they could find one), feedback control is feasible. A feedback controller does not require knowledge of a dominant cause.

Fascia Film Build

In the fascia film build example, introduced in Chapter 3, a team wanted to reduce variation in the paint flow rate. A baseline investigation showed the time pattern in the flow rate given in Figure 8.5.

The baseline data provided evidence that feedback may be feasible. The flow rate variation was much smaller over the short term than over the long term. Since the team also knew they could use a valve to adjust the flow rate, they adopted feedback control as the working approach. They planned to assess the benefit by simulating the effect of the feedback control scheme on the baseline data. At the same time they could assess the costs of the adjustments.

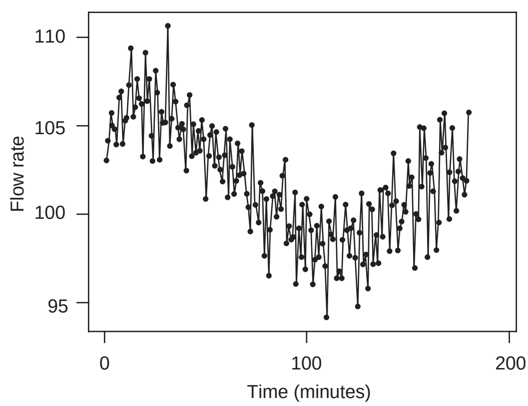


Figure 8.5 Variation in flow rate over time.

Parking Brake Tightness

In a vehicle assembly process, there was a problem with excess variation in parking brake tightness that led to considerable rework. A baseline investigation showed there were runs of high and low tightness values. The dominant cause acted slowly over time. Based on this knowledge, the team suspected that a dominant cause of tightness variation was the length of the front cable. The cables were delivered in batches and the average length differed substantially from batch to batch.

The team did not further explore front cable length; the cable arrived as part of an assembly, so it could not be measured in the plant. Because the dominant cause changed from batch to batch, they decided to consider feedback control.

The team knew that changing the depth of an adjustment nut would change parking brake tightness. To quantify the effect of this adjuster they planned an investigation where they would try different adjustment depths for a number of vehicles. In addition, the team still needed to decide on an adjustment rule, that is, when an adjustment should be made and by how much.

Comments

In the fascia film build and parking brake tightness examples, we saw a time pattern in the run chart of the output from the baseline investigation. We recommend a team use a sampling protocol in the baseline investigation that lets them see any systematic pattern in the output over time.

Patterns in the variation over time can have many forms. Some processes are setup dependent; that is, once a setup is complete, there is little variation from part to part. Many machining and stamping processes exhibit this behavior. If the team discovers such a pattern, and the output of the process can be measured immediately after setup and there is a low cost and quick way to adjust or redo the setup, then feedback control could be selected as the working approach.

In cases where feedback control is feasible, the output varies systematically over time. This implies that the dominant cause also exhibits the same time pattern. With this clue, it may be more economical to try to find the dominant cause than to adopt feedback control as the working approach. If the dominant cause is found, the team can address the cause directly, for example by applying feedback control to the cause. In the parking brake tightness example, feedback control was applied to the output despite partial knowledge of the dominant cause. This knowledge was used to help design the feedback control scheme.

8.5 SUMMARY

The thought process involved in choosing a working approach is summarized in Figure 8.6. Questions given at the same vertical height are addressed in either order or simultaneously depending on the problem context.

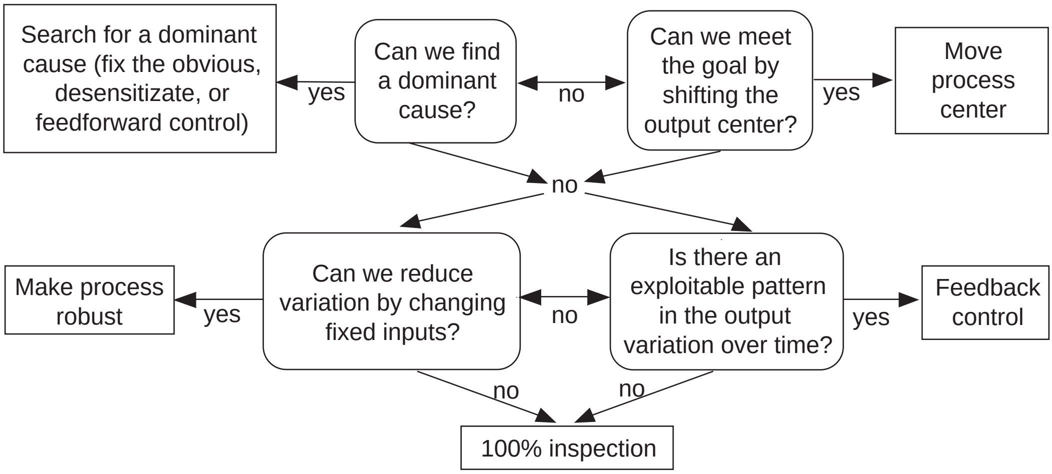


Figure 8.6 Flowchart to help choose a working approach before a dominant cause is known.

Key Points

- We recommend the team select a working approach from the seven possible variation reduction approaches to guide further efforts:
  –Fixing the obvious using knowledge of a dominant cause
  –Desensitizing the process to variation in a dominant cause
  –Feedforward control based on a dominant cause
  –Feedback control
  –Making the process robust
  –100% inspection
  –Moving the process center
- The following questions can help to select the working approach:
  –Can we find a dominant cause of the unit-to-unit variation?
  –Can we meet the goal by shifting the process center without reducing variation?
  –Can we reduce variation by changing one or more fixed inputs without knowledge of a dominant cause?
  –Does the process output exhibit a strong pattern in the variation over time?
- We strongly recommend the team search for a dominant cause of variation, unless there is clear evidence that one of the non-cause-based approaches is likely to be feasible.
- After choosing a working approach, we conduct further process investigations to determine whether the selected approach is feasible.
- In making a choice of working approach, we try to assess potential implementation costs.

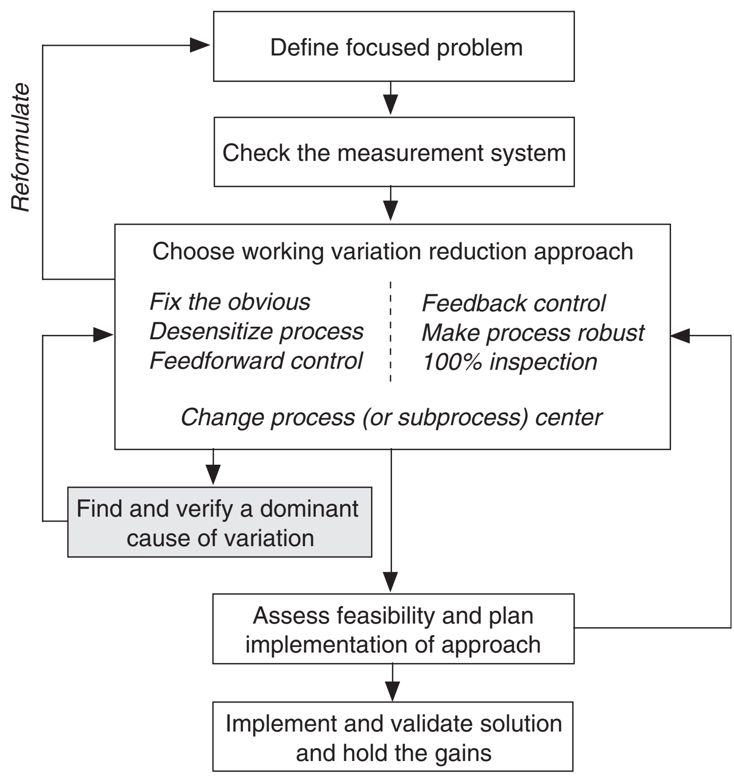
PART III

Finding a Dominant Cause of Variation

How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?

—Sir Arthur Conan Doyle (as Sherlock Holmes), 1859–1930

In many applications, identifying a dominant cause of variation leads to cost-effective process improvement. We recommend finding the dominant cause using the method of elimination and the idea of families of causes. With the use of observational plans and leveraging (comparing extremes), the search for a dominant cause of variation can be inexpensive and nondisruptive to the production process. We discuss the tools and methods used to support the search for a dominant cause. We introduce planned experiments to verify a suspect cause is dominant.



Statistical Engineering variation reduction algorithm.

9

Finding a Dominant Cause Using the Method of Elimination

We shall learn to be imaginative about what is causing product variability by going to the manufacturing area and figuratively talking to the product.

—Dorian Shainin, 1914–2000

In most problems, the team will decide to try to find a dominant cause before assessing the feasibility of a specific variation reduction approach. We recommend a search strategy based on the method of elimination to find a dominant cause. That is, we concentrate on ruling out possibilities rather than looking directly for the dominant cause. In this chapter, we discuss the principles behind this strategy; in chapters 10 through 13, we provide investigation plans, analysis tools, and numerous examples to demonstrate how it can be implemented.

The method of elimination can be explained using the familiar game 20 Questions. In one version of this game, we choose one of the 130,000 entries in the Canadian Oxford Dictionary (2002). We then give you the dictionary and ask you to determine the selected word using a series of yes/no questions. A poor strategy is to start asking about specific words. Unless you are lucky, you are quite likely to be exhausted before you get to the correct word. A much better strategy is to divide the dictionary in half and ask if the unknown word is in the first half. Whatever answer you get, you will have eliminated half the words in the dictionary with a single question. If you divide the remaining words in half at each iteration, you can find the unknown word with at most 17 questions (2^{17} is just greater than 130,000).

We apply the same idea to search for a dominant cause. We divide the set of all causes into families and then conduct an investigation to rule out all but one family. We repeat the exercise on the remaining family until a single dominant cause remains.

9.1 FAMILIES OF CAUSES OF VARIATION

In any process, there are many varying inputs that could have a large effect on the output. We partition the set of all such inputs into two or more families with common features such as the time frame or location in which they act. Then we use available data, investigations, and knowledge of the process to rule out all but one family as home of the dominant cause.

V6 Piston Diameter

Consider again the problem of reducing variation in V6 piston diameters introduced in Chapter 1. Figure 9.1 is a process map.

To start we divide the causes (varying inputs) into two families: measurement and the rest of the process. The measurement family includes all causes associated with the final gage and its operation and the rest of the process family includes all other causes. More specifically, the measurement system family includes those inputs that can change if we repeatedly measure the same piston: the operator, the position of the piston in the fixture, and the time since calibration of the gage, for example. The rest of the process family includes raw materials, the path through the machining process, the mold number, and so on. The dominant cause of diameter variation may be in either family.

To document progress in the search, we portray the families in a *diagnostic tree.* The top box describes the problem and the families of causes are listed below. The first part of the diagnostic tree for the V6 piston diameter problem is shown in Figure 9.2. As we proceed through the search, defining and eliminating families of causes, the diagnostic tree will grow.

To eliminate one family of causes, we carry out an investigation, as described in Chapter 7, to learn how much of the baseline variation can be attributed to the measurement system. In this investigation, two operators measured three different pistons three times each over two different days. The variation in the measurement system was relatively small since the discrimination ratio D was 8.3. We eliminate the measurement system family and concentrate on the rest of the process family as the home of the dominant cause of variation. We cross off the eliminated family on the diagnostic tree.

We always start with the two families: the measurement system and the rest of the process. We have explicitly included this partition of the causes in the Check Measurement System stage of the Statistical Engineering algorithm.

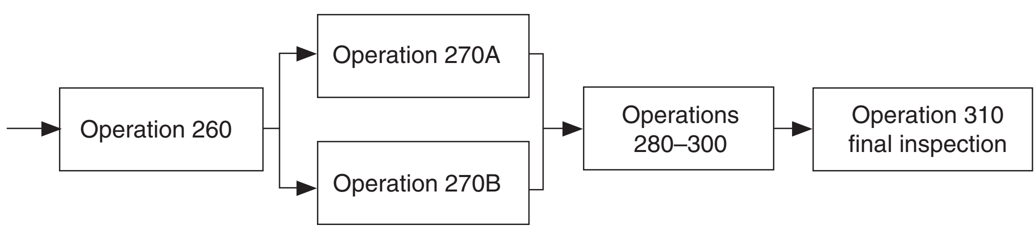


Figure 9.1 The piston machining process map.

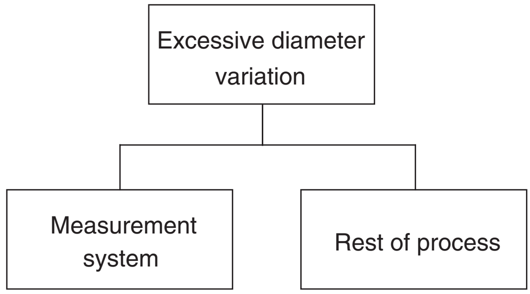


Figure 9.2 Diagnostic tree for piston diameter variation.

Bottle Label Height

In a bottling operation, management raised a concern over crooked and misplaced labels. The ideal label was a set distance below the bead on the neck of the bottle. The output characteristics were the average label height and the difference of the maximum and minimum heights as the bottle was rotated. The team assessed the measurement system that used a feeler gage and found that the measurement variation was small relative to the baseline for both characteristics.

Next, the team decided to proceed with the search for a dominant cause by partitioning the remaining causes into the bottle-to-bottle and the time-to-time families. The bottle-to-bottle family includes inputs that change quickly from one bottle to the next in the labeling process. The time-to-time family contains causes that vary more slowly and change substantially only from one hour to the next. To assess the relative effect of the two families, the team measured the label height for five consecutive bottles selected each hour for two shifts, a total of 16 hours.

We show the diagnostic tree for average height in Figure 9.3. Since there are two outputs in this problem, the tree may be different for each.

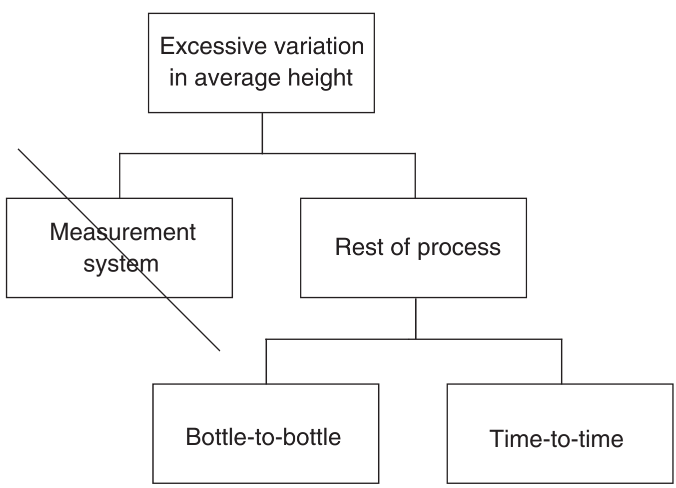


Figure 9.3 Diagnostic tree for average label height.

Engine Oil Consumption

A team was assigned the task of reducing warranty costs due to oil consumption in a truck engine. The engine was built at two different plants that had common equipment, common suppliers, and similar manufacturing processes. From the warranty database, the team noted that while the two plants had produced roughly the same number of engines, over 90% of the more than 1500 claims were associated with engines built in one of the two plants.

Here we divide the causes into two families: the within-plant and plant-to-plant families (see Figure 9.4). The within-plant family contains causes that vary in the same way for each plant, such as characteristics of components from a common supplier. The plant-to-plant family contains causes that have different values in the two plants, such as characteristics of components from different suppliers or differences in operating procedures.

Using the warranty data, the team ruled out the within-plant family. The dominant cause lived in the plant-to-plant family. To continue the search for a dominant cause, the team focused on the few varying inputs that were different in the two plants.

In this example, the team did not check the measurement system. The discrepancy in warranty claims was so great that they assumed that there was a real difference between the plants. The team made huge progress using the data available from the warranty system because the plant-to-plant family had relatively few causes.

Surprisingly, another team working on the same problem from an engineering perspective had decided (incorrectly) that changing the design of the PCV valve could solve the problem. The second team had implicitly adopted a robustness approach. By changing the PCV valve, they hoped to reduce the effect of the unknown dominant cause of oil consumption. Since there was a common supplier, no characteristic of the PCV valve could be a dominant cause because there was no systematic difference in the valves from plant to plant.

There are many ways to partition causes into families and different investigations and tools to eliminate all but one family. We look at numerous examples in the next three chapters.

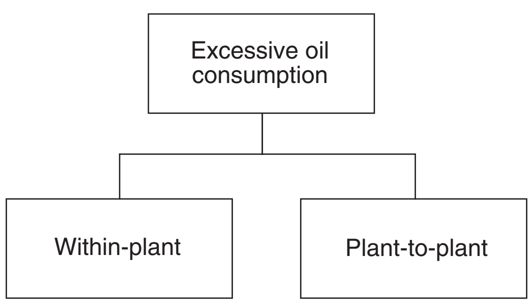


Figure 9.4 Partition of the causes of oil consumption.

9.2 FINDING A DOMINANT CAUSE USING THE METHOD OF ELIMINATION

To find a dominant cause, we apply the method of elimination repeatedly. There are three steps:

1. Divide the remaining suspect causes into two or more families.
2. Plan and carry out an investigation to eliminate all but one family as the home of the dominant cause.
3. Repeat steps 1 and 2 until only one suspect for the dominant cause remains.

The search for a dominant cause is complete if we find a cause or family that is specific enough to allow us to choose a variation reduction approach. In the engine oil consumption example, the dominant cause was in the plant-to-plant family. However, the team could not address this cause directly unless they were willing to shut down one plant! The team continued searching for a more specific cause in the identified family.

In some cases, we may stop the search when there are only a few possible causes left because it is more efficient to identify the dominant cause using an experimental plan (see Chapter 13).

V6 Piston Diameter

Consider again the V6 piston diameter example, discussed earlier in this chapter, where the problem was excess variation in piston diameter. The first investigation eliminated the measurement system family. Next, the team looked at two location-based families as shown on Figure 9.5.

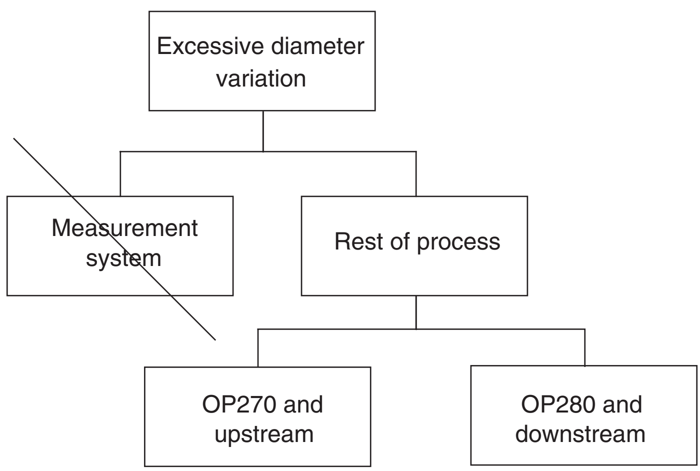


Figure 9.5 Diagnostic tree for V6 piston diameter variation.

The team planned an investigation in which they measured the diameter of 96 pistons after Operation 270 and at the final gage. This investigation eliminated the operations downstream from Operation 270 as the home of the dominant cause.

The team concluded that the diameter measured just after Operation 270 was the dominant cause of variation in the final diameter. This finding did not surprise the team, who understood the functions of Operation 280 through 310. However, by verifying that the dominant cause lived in Operation 270 or upstream, they had made substantial progress. The order in which the pistons are measured at the final gage was not the same as the order in which they were machined, whereas the order of production was preserved much better in the machining operations up to Operation 270. The preservation of order made it much easier to track pistons through the process and thus made further investigation easier.

The team was unable to directly address the diameter variation at Operation 270. They decided to reformulate the problem and to search for a dominant cause of diameter variation as measured after Operation 270. At the same time, they compared the measurement system at Operation 270 to the final gage. While the Operation 270 measurement system was not the home of a dominant cause of variation, the team discovered the relative bias between the Operation 270 and final gages as described in Section 5.2. They fixed this problem by changing an offset in the Operation 270 gage. They also knew that this obvious fix would not reduce diameter variation at either gage.

Next, the team decided to look at three families of causes of Operation 270 diameter variation as documented on Figure 9.6.

The team planned another investigation with 96 pistons. For each piston, the diameter was measured before and after Operation 270. Each parallel stream at Operation 270

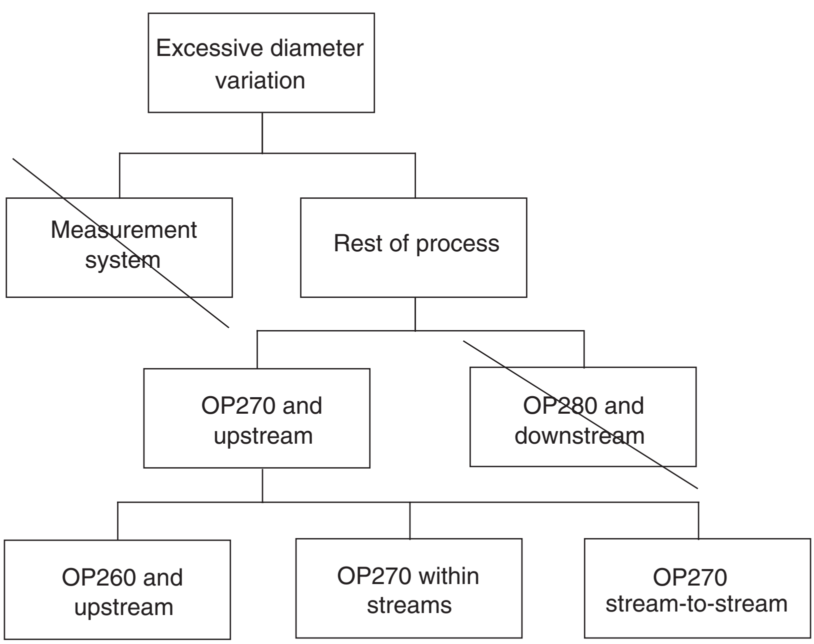


Figure 9.6 Diagnostic tree for V6 piston diameter problem after Operation 270 investigation.

processed half the pistons. Based on the observed data, the team concluded that the Operation 270 stream-to-stream family was the home of the dominant cause. They eliminated the upstream family. They concentrated their efforts on identifying what was different between the two streams that could explain the observed difference in process behavior.

At Operation 270, the grinders were of the same design, and there was a common source of unfinished pistons and a common control plan. However, there were different operators. The team interviewed the operators and found that each operator ran his machine differently; each was convinced that the method they used was superior to that in the written control plan.

Now the team suspected the dominant cause was a difference in the method of process control in the two streams. The final step was to verify this suspicion. We give the completed diagnostic tree in Figure 9.7.

We use the diagnostic tree to document the search for the dominant cause. We can add additional information to the diagram to support the logic of eliminating families. Remember that the tree is built up as we iteratively apply the method of elimination. We do not start by constructing the whole diagram.

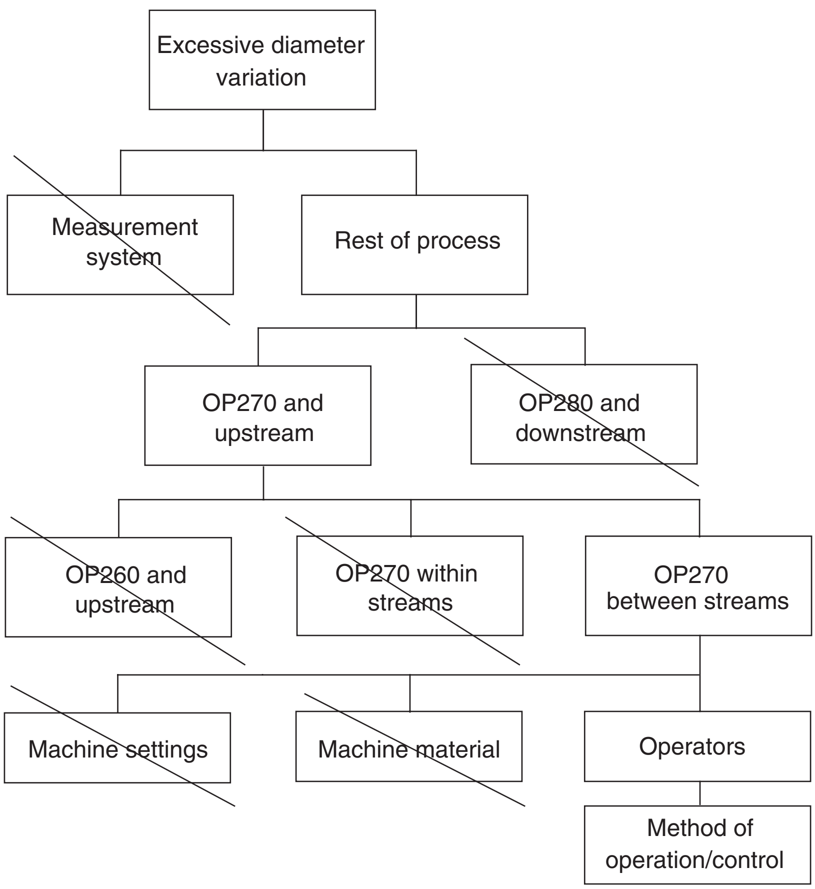


Figure 9.7 Final diagnostic tree for excessive variation in V6 piston diameter.

9.3 IMPLEMENTING THE METHOD OF ELIMINATION

In this section, we give some general suggestions for implementing the method of elimination that we illustrate with the examples in subsequent chapters.

Using the method of elimination, we do not start by listing possible causes and classifying them into families.[1] Instead, we start by considering broad categories that define the families. This is a major difference from other problem-solving algorithms. As we obtain more process knowledge, we subdivide the remaining families to further narrow down the suspects for a dominant cause. Our goal is to end up with a single cause or, in the worst case, a short list of possibilities. The final step is to verify (and select in the case of a short list) that we have indeed found a dominant cause.

For most problems, there are a large number of possible dominant causes and many ways to divide them into families. How do we start and proceed? We make the following suggestions.

Check the Measurement System First

To start, we always divide causes into those associated with the measurement system and those associated with the rest of the process. We can use a measurement system investigation as described in Chapter 7 to eliminate one of these two families.

Use the QPDAC Framework

The search for a dominant cause involves many process investigations. To help plan, execute, and draw conclusions from such empirical investigations, we use the five-step QPDAC framework as described in Chapter 5.

The investigations used to support the search for a dominant cause share many features. Using the QPDAC terminology, the *target population* is all units produced by the process now and in the future. The target population stays the same unless we reformulate the problem. The attributes of interest are the components of the output variation attributable to the different families or individual causes.

In the Plan step, we design the investigation to avoid large study and sample errors. We are unconcerned with measurement error since we have previously assessed the measurement system and determined that it is adequate. We use the results from the baseline investigation to help make the choice of study population. We choose this population so that the dominant cause has the opportunity to act within this population. We then select the sampling protocol so that we can estimate the variation attributable to each possible family. As always, we balance the statistical considerations against the constraints of cost and time.

Consider the problem of excess variation in average label height introduced earlier in this chapter. In the investigation to compare the bottle-to-bottle and time-to-time families of causes, the team decided to choose a study population defined by all bottles produced over two shifts. They made this choice because in the baseline investigation, where they had looked at historical records covering production from over a month, they had seen close to the full extent of variation in average label height within each shift. If, instead, the baseline investigation had shown that the dominant cause acted more slowly, for example, week to week, they would have needed a longer time frame for the study population. To obtain information about the bottle-to-bottle family, the sampling protocol must involve selecting some bottles made consecutively. To help avoid sample error, the team selected five consecutive bottles every hour throughout the study population.

In the Find and Verify a Dominant Cause stage of the algorithm, we assume the baseline investigation captured the performance of the current and future process with little study or sample error. With this assumption, if we see the full extent of variation in the output in any subsequent investigation, we are confident the dominant cause has acted. See further discussion of the full extent of variation later in this chapter.

Think About Families of Causes and the Possible Types of Investigations

In chapters 10 through 12, we present a variety of process investigations useful in the search for a dominant cause. Each type of investigation aims to compare different families of causes. We may decide to compare:

- Location-based families (stratification)
- Time-based families
- A combination of time- and location-based families (multivari investigation)
- Upstream and downstream families (variation transmission investigation)
- Assembly and component families (component swap)
- The effects of individual inputs (group comparison, input-output investigation)

The best choice of investigation depends on the problem, the process, and current process knowledge. At the start of the search, we want to compare large families of causes with the hope of eliminating many causes quickly. Later in the search, we are more willing to look at individual causes.

Use Available Data

There are often data available that have been collected for other purposes such as routine process control, scrap accounting, or warranty cost assessment. We also have the data from the baseline investigation. We suggest the team examine the available data and how it was collected by running through the steps of QPDAC. If the team has confidence that there is little study and sample error, they can use the data to rule out families of causes. Here, we start with the data and then think about the families.

The use of the warranty data in the engine oil consumption problem is a good example. For another example, in Chapter 3, we briefly discussed a problem where there was a high frequency of rejected engines at the valve train test. Every engine was tested and the results were recorded in a database. There were three test stands that operated in parallel and engines were assigned to a stand haphazardly. The stands checked a number of characteristics and an engine was rejected if one or more characteristics was out of specification. Rejected engines were torn down and repaired. There were many “no fault

Table 9.1 First-time reject rate by test stand.

Test stand	First-time reject rate
1	12.8%
2	4.2%
3	2.8%

found" rejects. The test stand operators repeatedly measured a rejected engine without teardown because they knew it would often pass with a second or third test.

A team was assigned the goal of reducing the reject rate. The baseline first-time reject rate was about 6.6%. Before conducting an assessment of the measurement system, the team used the available data to stratify the first-time reject rate by test stand over a one-week period. The results are given in Table 9.1.

The dominant cause of the high reject rate lived in the measurement system family. The team proceeded by looking at causes that could explain the stand-to-stand differences.

Use Knowledge of the Process

The method of elimination relies on process knowledge both in determining the plan for an investigation and interpreting the results. For example, in the V6 piston diameter example discussed earlier, the team suspected that the dominant cause acted upstream of Operation 280. They chose to split the process after Operation 270 so that they could rule out all the downstream operations in a single investigation.

Knowing what can be measured and where, how easy it is to trace parts through the system, when production order is preserved, and so on, is essential in selecting families and developing investigation plans.

Knowledge of the process can help the team interpret the data and determine what families are eliminated. For example, in a problem to reduce scratches on a painted fascia, the team made a concentration diagram using 100 damaged parts to show where on the part the damage occurred. Combining the new knowledge about the location of most of the scratches and existing knowledge of the handling system, the team was able to rule out large parts of this system as the source of the scratches.

Observe the Full Extent of Variation

We assume that there is a single dominant cause[2] that explains most of the output variation seen in the baseline investigation. When we carry out an investigation to eliminate one or more families, we need to see most of this baseline variation in the observed output values. Otherwise, we may decide the dominant cause is in the wrong family.

Consider the V6 piston diameter example with some hypothetical data. Suppose we start by considering the two families shown in Figure 9.8. Recall from the baseline investigation that the full extent of diameter variation was 582 to 602 microns.

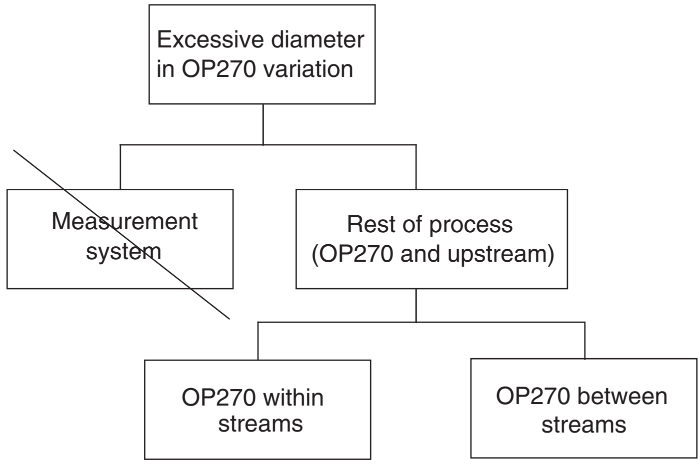


Figure 9.8 Diagnostic tree for hypothetical search.

To eliminate one of the families, we track 30 consecutive pistons from each stream at Operation 270 and measure the diameters at the final gage. What can we conclude if the results are as given by Figure 9.9?

If we are careless, we conclude that there is a large difference stream to stream and rule out the within-stream family. However, in the 60 pistons, the diameter varies only from 588 to 596 microns, about half the full extent of variation in the baseline investigation. The dominant cause has not acted fully during this investigation, and we may be incorrect in ruling out either family. We need to revisit the plan for the investigation. Perhaps we should have collected 30 pistons from each stream spread out over a day rather than 30 consecutive pistons to see the full extent of variation.

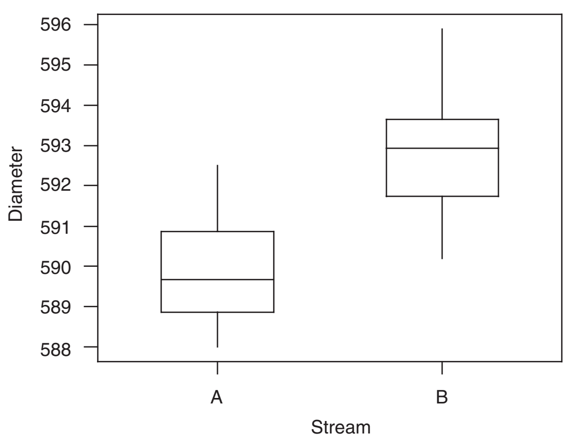


Figure 9.9 Box plot of diameter by Operation 270 stream from hypothetical investigation.

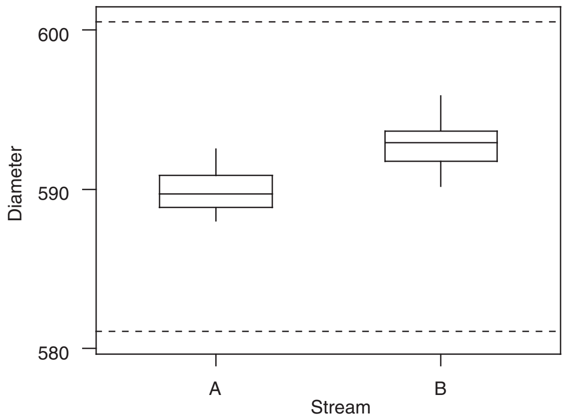


Figure 9.10 Box plot of diameter by Operation 270 stream from hypothetical investigation showing full extent of baseline variation.

To ensure correct interpretation, we recommend showing the full extent of variation in the baseline investigation on all graphical displays of data used in the method of elimination. So rather than use Figure 9.9, we present the results as in Figure 9.10. With Figure 9.10, we cannot conclude that the stream-to-stream family is home of a dominant cause.

The message here is that we need to consider the steps in QPDAC carefully even for these simple studies to rule out families of causes. We need to specify the study population and sampling protocol so that we can be confident that we will see the full extent of variation in the output.

Use Leverage Where Possible

Leverage means that we select units with output values at both ends of the full extent of variation. Assuming there is a single dominant cause, the values of this cause will be very different among these extreme units. This guarantees that we will see the full extent of variation in the output.

Consider again the problem of excessive variation in V6 piston diameters. The full extent of variation in the baseline was 582 to 602 microns. After checking the measurement system, suppose we select 10 pistons with diameter larger than 600 and 10 with diameter less than 584 microns. If the dominant cause acts in the stream-to-stream family at Operation 270, we expect almost all of the large pistons will come from one stream and almost all the small pistons from the other stream. By selecting pistons that are very different, we get leverage in comparing families of causes.

In the V6 piston diameter example, the team could not conduct such an investigation because they could not determine the stream of the finished pistons.

Leverage was used in the problem of noisy windshield washer pumps introduced in Chapter 6. The team established a reliable in-plant measurement system for noise and a baseline. They then selected five noisy and five quiet pumps relative to the full extent of variation. Next they disassembled and reassembled each pump and noticed that there were dramatic changes in noise levels. The noisy pumps became much quieter and the quiet pumps somewhat noisier when the pump housing was rotated relative to the motor housing in the reassembly. The dominant cause of noise was in the assembly family.

If we exploit leverage, we can use small sample sizes in many investigations. We need to ensure that the extreme output values are due to the dominant cause. If the extreme values are due to a different failure mode, leveraging may mislead us. By defining a focused problem and avoiding outliers, we hope to avoid this occurrence.

Take the Simplest Path

There are many ways to divide causes into families. For a given partition, you may or may not be able to think of a way to investigate the process that will allow you to rule out all but one family. In the next chapters, we describe numerous plans for investigations and the corresponding analysis tools. We recommend that you consider each of these plans in the context of your problem and where you are on the diagnostic tree. Then pick a simple plan that can help make progress.

In theory, you can look at many families simultaneously using a complex plan. This seems like a good idea because you can greatly narrow the search for the dominant cause with a single investigation. However, we strongly recommend using a small number of large families to keep the investigation simple. We know that this recommendation generates more iterations in the search for the dominant cause. We also know from bitter experience that complex plans often fall apart due to logistical difficulties and production pressures.

Most of the plans for ruling out families of causes are observational. We strongly recommend observational plans, where we "listen to the parts," over experimental plans that require an intervention in the process. Experimental plans are difficult to manage and expensive to execute in the production environment.

Be Patient

To use the method of elimination, we need to plan, execute, and analyze the results of a number of simple investigations. To be successful, we need to avoid the temptation to jump to a specific cause too soon. As consultants, we have met strong resistance from process managers when we suggested that we plan to conduct several investigations to find the cause of variation. The usual reaction is to ask that we get it over with quickly using a single investigation. In management review meetings, we have seen problem-solving teams struggle with the question,

"Well, have you found the cause (or solution) yet?"

The answer, "No, but we have eliminated a lot of possibilities," can be hard to defend if the problem is urgent and substantial resources have been committed to its solution. On the other hand, trying but failing to find the cause with a single complicated investigation is even harder to defend.

Reconsider the Variation Reduction Approach

Sometimes the method of elimination fails. We may have identified a particular family as home of the dominant cause, but we cannot see how to split this family further. We are left with two choices. We can begin testing the remaining individual suspects using the methods described in chapters 12 and 13. Alternately, we can abandon the search and adopt a variation reduction approach that does not require the identification of a dominant cause. We still use any knowledge gained about the family containing the dominant cause.

Key Points

- To find the dominant cause of variation, partition the possible causes into families and use the results of an investigation to eliminate all but one family as the home of the dominant cause. Iterate to further subdivide the remaining families until a dominant cause is found.
- "Let the process or product do the talking." Observational data can provide strong clues about a dominant cause of variation.
- Document the search for a dominant cause using a diagnostic tree.
- Implementation suggestions:

  –Check the measurement system first
  –Use the QPDAC framework to help plan the investigations
  –Use available data
  –Use knowledge of the process
  –Observe the full extent of variation
  –Use leverage where possible
  –Take the simplest path
  –Be patient
  –Reconsider the variation reduction approach

Endnotes (see the Chapter 9 Supplement on the CD-ROM)

1. The traditional strategy to find the cause of a problem starts by listing all the possible causes in a cause and effect diagram. We compare this strategy to the use of families and the method of elimination in the chapter supplement.
2. The method of elimination relies on there being one or two dominant causes. In the chapter supplement, we discuss what happens when this assumption does not hold.

Exercises are included on the accompanying CD-ROM

10

Investigations to Compare Two Families of Variation

The world is full of obvious things which nobody by any chance ever observes.

—Sir Arthur Conan Doyle (as Sherlock Holmes), 1859–1930

In this chapter, we describe some plans and the corresponding analysis to compare two families of causes. We concentrate on a single iteration of the method of elimination where the goal is to eliminate one of the two families. For each example, we assume that the team has successfully completed the Define Focused Problem and Check Measurement System stages of the variation reduction algorithm and has decided to search for a dominant cause of variation.

We use plans that meet the criteria discussed in Chapter 9:

- Keep it simple.
- Use available data.
- Observe the full extent of variation in the output.
- Exploit leverage if possible.

10.1 STRATIFICATION

With stratification (Kume, 1985), we divide parts into distinct groups based on their source. For example, we may stratify by machine, mold, line, gage, plant, location within the part, supplier, operator, and so on.

Using stratification, we can assess whether the dominant cause impacts each group in the same way or each group in different ways. A dominant cause that impacts different groups differently is said to act in the group-to-group family. If such a cause is present, we will see large differences among the group averages and low variation within each group.

Casting Scrap

The baseline scrap rate for a casting defect was 2.4%. The castings were molded four at a time with cavities labeled *A, B, C,* and *D*. The cavity label was molded into the casting.

The team decided to stratify the causes into the cavity-to-cavity family (also called the *among-cavity* or *between-cavity* family in other references) and the within-cavity family. We have:

- Cavity-to-cavity family: causes that affect different cavities differently
- Within-cavity family: causes that affect each cavity in the same way

The team classified 200 scrap castings by cavity as shown in Table 10.1. Because the proportion of scrap castings from each cavity was about the same, the team ruled out the cavity-to-cavity family. The dominant cause affected all cavities in the same way.

In Table 10.1, we compared the number of scrap castings from each cavity directly because there was equal volume from all four cavities. If the cavity volumes were not the same, we would have compared the percent scrapped from each cavity.

Table 10.1 Scrap stratified by cavity.

Cavity	Number of castings
A	58
B	51
C	43
D	48
Total	200

Rod Thickness

In Chapter 6, we looked at an example where the goal was to reduce thickness variation in a connecting rod at a grinding operation. The team established a baseline and found the full extent of thickness variation was 2 to 59 thousandths of an inch (recorded as a deviation from a particular value).

In the baseline investigation, the team measured thickness of each sampled rod at four positions. The data are given in the file *rod thickness baseline*. With these data, we can compare the family of causes that produce different thickness within each position to the family of causes that act from position to position.

In Figure 10.1, we plot thickness by position. Since we are using the baseline data, we are guaranteed to see the full extent of variation for the output, given by the dashed horizontal lines. Position 3 has a smaller average thickness than the other positions. There is a cause in the position-to-position family that produces this difference. However, we can see that this cause is not dominant by imagining the plot with the average thickness

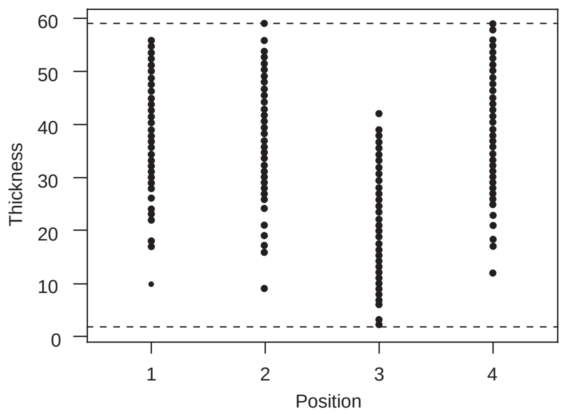


Figure 10.1 Rod thickness stratified by position (dashed horizontal lines give the full extent of variation).

aligned for each position.[1] The variation within each position is a substantial part of the overall variation. We cannot eliminate all the out-of-specification rods without reducing the within-position variation. Of course, we will implement an obvious fix to better align the thickness center at Position 3 if available.

Camshaft Lobe Runout

There are 12 lobes per camshaft in the problem with excess base circle (BC) runout described in Chapter 1. In the baseline investigation, the team measured BC runout on all lobes for nine different camshafts selected each day for 12 days. We can stratify the causes into the lobe-to-lobe and within-lobe families.

- Lobe-to-lobe: causes that affect different lobes differently
- Within-lobes: causes that affect each lobe in the same way

We use these data, given in *camshaft lobe runout baseline*, to try to eliminate one family of causes. Here again we are guaranteed the full extent of variation since the data came from the baseline investigation.

We can see the variation due to the within-lobe and lobe-to-lobe families in Figure 10.2. The center lobes (numbers 5–8) have much lower averages and standard deviations than do the end lobes.

We cannot eliminate either family based on the data. However, we get some valuable clues. The dominant cause is an input that acts in both families or involves inputs from both families.[2] We redefine the families. Since it is difficult to think of good labels, we proceed based on a diagnostic tree as given in Figure 10.3.

In the subsequent search, the team looked for causes that could explain the difference between the middle and end lobes and eliminated any that could not.

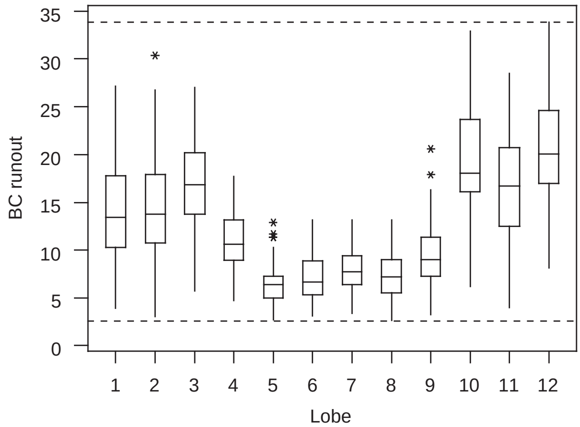


Figure 10.2 Camshaft journal runout by lobe (dashed horizontal lines give the full extent of variation).

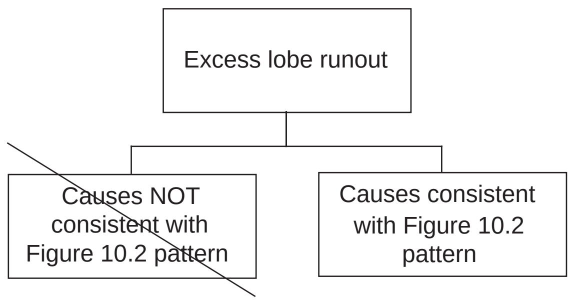


Figure 10.3 Diagnostic tree for camshaft lobe runout.

Comments

We can often use stratification with existing data for any multistream process if we can identify the stream associated with each part. Some examples are parallel processing steps, multiple suppliers, gages, product sources, operators, and so on. We saw two examples in Chapter 9: first, when oil usage warranty claims were stratified by the plant where the engines were built, and second when rejected engines were stratified by the test stand.

For a continuous characteristic, a box plot (using either the box summary or individual values depending on the data volume) is a convenient way to present the data to display the variation within groups and group to group. Adding dashed horizontal lines shows if we have seen the full extent of variation. The easiest way to assess the contributions of the two families is to shift the boxes (or individual values for each group) in your mind so that they all have the same centerline. The variation that remains is due to the within-group family. If this variation is large relative to the full extent of variation, we eliminate the group-to-group family as the home of the dominant cause.[3]

10.2 COMPARING TWO TIME-BASED FAMILIES

For many problems, we can separate causes based on the time frame within which they vary. Some causes, such as properties of raw materials and tool wear, change slowly, while others, such as position in a fixture and contamination by dirt, can change quickly from part to part. We use the term *part-to-part* to label the family of causes that vary from one part to the next in time sequence. An investigation to compare two time-based families is a simple special case of a multivari—see Chapter 11.

In Figure 10.4, we show a schematic of a sampling plan where we measure three consecutive parts once per shift. We plan to sample parts until we see the full extent of variation established in the baseline investigation.

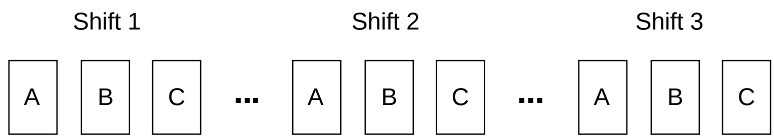


Figure 10.4 Part-to-part and shift-to-shift families.

Camshaft Lobe BC Runout

The plan of the BC runout baseline investigation was to collect nine camshafts per day for 12 days. To compare the effects of the within-day and day-to-day families, the team plotted the runout values by day as given in Figure 10.5. Since the pattern across days was similar, they were tempted to eliminate the day-to-day family.

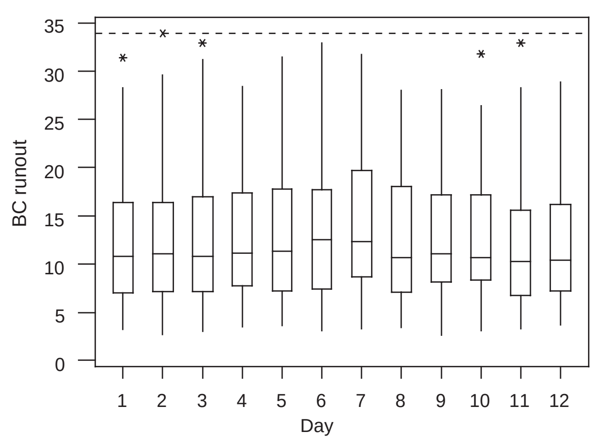


Figure 10.5 Box plots of BC runout by day (dashed horizontal lines give the full extent of variation).

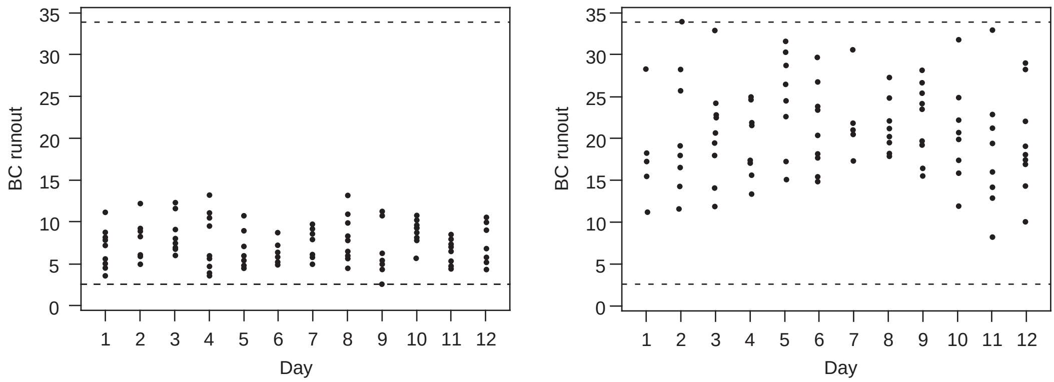


Figure 10.6 Box plots of camshaft journal runout by day for Lobe 8 (left) and Lobe 12 (right) (dashed horizontal lines give the full extent of variation).

However, recall that the dominant cause must explain the differences between end and middle lobes as discussed in Section 10.1. To see if the lobe-to-lobe differences depended on the day, the team examined BC runout by day for each lobe separately. Figure 10.6 shows the plots for lobes 8 and 12. In the baseline, Lobe 12 exhibited the most runout variation, while Lobe 8 gave the least.

Figure 10.6 suggests there was no interaction between causes that act lobe-to-lobe and day-to-day since for each lobe, the pattern in runout is similar across days. The team concluded the dominant cause acted within each day. Using this conclusion, they could plan subsequent investigations over shorter time periods (i.e., one or two days) and expect to see the full extent of variation.

Engine Block Porosity

We introduced a problem of reducing scrap due to porosity in cast-iron engine blocks (Figure 10.7) in Chapter 3. Porosity occurs because gas bubbles in the molten iron do not escape before the casting hardens. The scrap rate was about 4%, with half detected in the foundry (the porosity was visible on the surface) and the rest found after the initial machining of the bank faces of the block.

The team created a new measurement system to determine porosity on a continuous scale. The system measured the total hole size across the bank faces based on a classification system of hole sizes in 1/64 square inches. The team checked the measurement system and judged that it was adequate. With this measurement system, there was not a direct translation to the decision to scrap a block since that decision depended critically on the location of the porosity, not just the total size.

Although a new output was defined, the team did not establish a new baseline. The full extent of variation in the new porosity measure could have been easily determined by measuring porosity using the new continuous scale on a large number of blocks. The team decided not to take the time to establish the baseline for the new measure of porosity. This decision introduced the risk that they would make the wrong decision about the identity of a dominant cause.

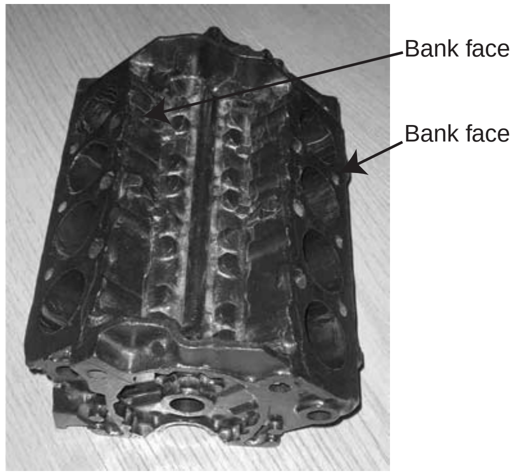


Figure 10.7 Plastic scale model of the engine block.

The team decided to compare block-to-block and time-to-time families as shown in Figure 10.8. The block-to-block family includes causes that change quickly from one block to the next as they are poured. The time-to-time family includes causes that change more slowly.

The team marked five consecutive blocks at the start of every hour for eight hours. They located the 40 blocks after the bank faces were machined and measured the porosity. Of the 40 blocks, 3 would have been rejected due to porosity had they been subject to the regular production inspection. This exceeded the historical reject rate, so the team was confident that the dominant cause was acting during the investigation. The data are stored in the file *engine block porosity multivari*.

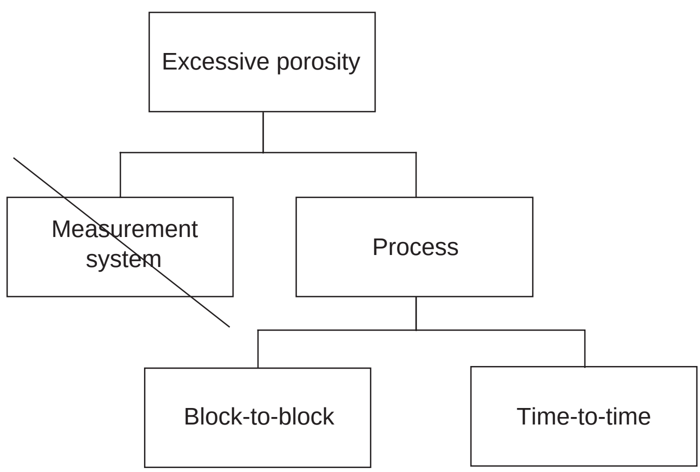


Figure 10.8 Diagnostic tree for excess porosity.

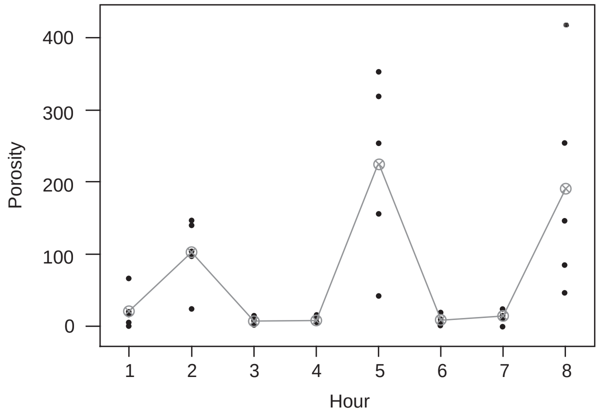


Figure 10.9 Multivari chart of block porosity versus hour.

We created the plot in Figure 10.9 using the multivari routine in MINITAB (see Appendix C). The dots on the plot correspond to the individual porosity values. The five block porosity averages are joined from hour to hour. We see large differences in the average porosity over time relative to the variation of the porosity within most hours. Based on this observation, we eliminate all causes, such as mold dimensions, that act from block to block. The average porosity (and the variation) is highest in hours 5 and 8. The team noted that these were times when the process was neglected due to lunch and preparations for the end of the shift. Combining this process knowledge with the patterns seen in Figure 10.9, the team looked for causes that behaved differently during these special times.[4]

10.3 COMPARING UPSTREAM AND DOWNSTREAM FAMILIES

We can divide causes into families based on where in the process they act. If we represent the process as a series of operations, we can split the causes in terms of whether they act upstream or downstream from a particular point.

When comparing upstream and downstream families we use a *variation transmission* investigation. We want to determine whether variation observed at an intermediate operation is transmitted through the downstream operation or if the downstream operations add substantial variation. We consider variation transmission investigations involving more than two process steps in Chapter 11.

Camshaft Lobe BC Runout

The team had discovered that for Lobe 12, the dominant cause of BC runout variation acted within days. From the baseline investigation they knew that the full extent of runout variation

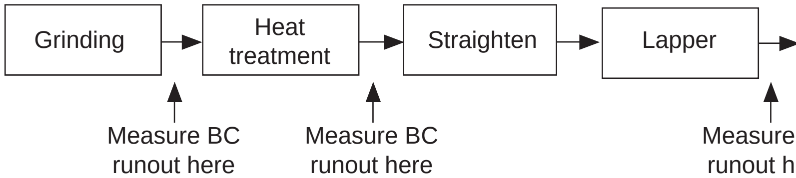


Figure 10.10 Camshaft process map.

was 2.6 to 33.9. Using the process map (Figure 10.10), they first split the remaining causes into two families:

- Causes that act downstream of heat treatment
- Causes that act in heat treatment or upstream

The team selected 32 parts over the course of one day. There were four camshafts from each of the eight lobe grinders. One camshaft from each grinder was processed on each heat treatment spindle. BC runout was measured after the heat treatment and after the final step of the process on each lobe. The data for Lobe 12 are in the file *camshaft lobe runout variation transmission*.

Figure 10.11 shows a plot of the Lobe 12 final BC runout versus the BC runout after heat treatment. Similar plots were produced for the other lobes. Across all the lobes, the final runout variation matched the baseline variation. Using Figure 10.11, we can separate the two families:

- Downstream family—causes that act downstream of heat treatment
- Upstream family—causes that act in the heat treatment step or upstream

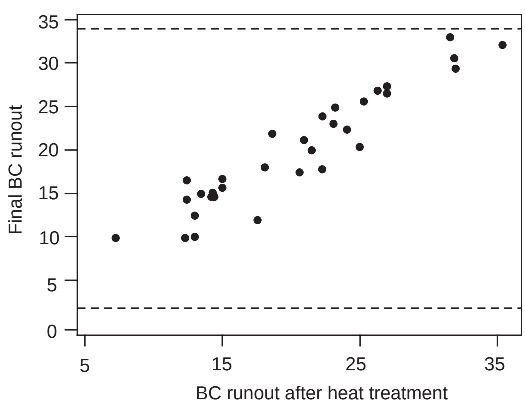


Figure 10.11 Lobe 12 BC runout after heat treatment and final (dashed lines give full extent of variation).

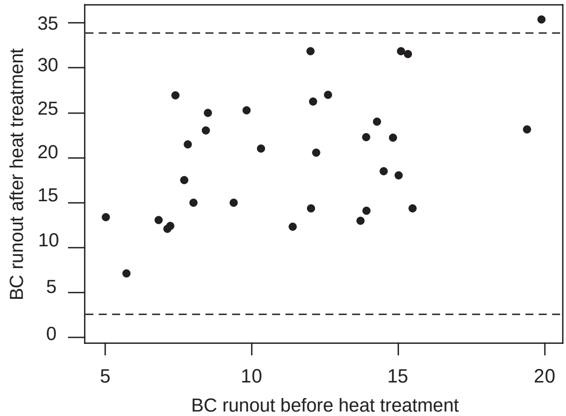


Figure 10.12 Lobe 12 BC runout before and after heat treatment (dashed lines give full extent of variation).

The plot shows a strong relationship between runout after heat treatment and at the end of the process. If we hold fixed the runout after heat treatment, there is little variation in final runout. We eliminate the downstream family that includes the straightening and lapping operations as the home of the dominant cause. The variation in runout is coming from the upstream family.

In the investigation, the team also measured runout for all lobes before the heat treatment step. Figure 10.12 shows very little relationship between the BC runout before and after heat treatment. If we hold fixed the runout before heat treatment, we see most of the full extent of variation after heat treatment. The team concluded that the dominant cause acts in the heat treatment. They eliminated all causes that act upstream of the heat treatment.

Cylinder Head Rail Damage

In the casting of cylinder heads, the problem was excessive scrap due to damaged head rails. The damage occurred somewhere in the shakeout and cleaning operations. The heads were 100% inspected.

To eliminate some causes, the team recorded the location of the damage for 100 scrap rails on a concentration diagram, a schematic of the part. See Figure 10.13.

The team noticed that the damage did not occur at random on the rails but was concentrated at particular locations. The nonrandom pattern ruled out several of the steps in the cleaning process as the home of a dominant cause. As a result of this simple investigation, the process owners canceled the purchase of new equipment ordered to deal with the damaged rails. The team concluded the dominant cause did not act in the processing step where the new equipment would have been installed.

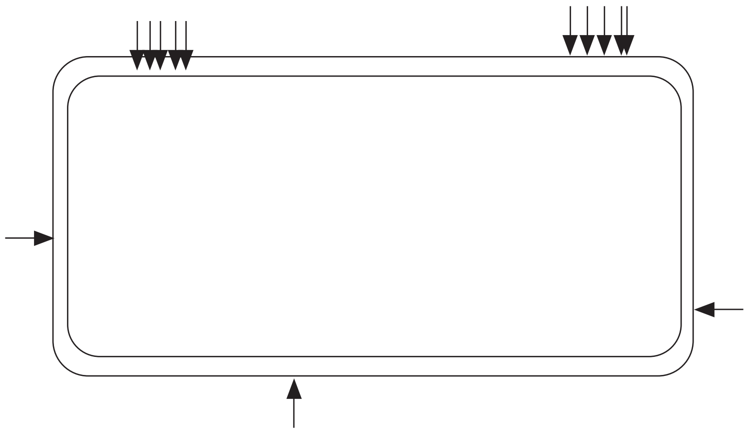


Figure 10.13 Concentration diagram showing location of rail damage.

Comments

We can use plans similar to that used in the camshaft lobe runout example to partition the causes upstream and downstream from an intermediate operation whenever we can measure the output after the intermediate operation.

In the chapter supplement, we look at other plans and analysis tools for comparing upstream and downstream process step families that have more limited application.[5]

10.4 COMPARING ASSEMBLY AND COMPONENT FAMILIES

For an assembled product, we can partition causes into those that reside in the assembly operation and those that live with the components. To rule out one of the families, we observe whether the output characteristic of a unit changes substantially when it is reassembled using the same components. This is only possible if we can disassemble and reassemble without damaging or changing the components. This type of investigation works well for rare defects.

To exploit leverage, we start with two assemblies with opposite and extreme performance in terms of the output.[6] In other words, the two selected units show the full extent of variation. Next we disassemble and reassemble the units at least three times, each time measuring the output. If the output does not change substantially for either unit, we eliminate the assembly operation as the home of the dominant cause.

Power Window Buzz

A team was charged with resolving a "buzz" problem with a power window regulator. Buzz was measured on a seven-point scale. In the baseline investigation, the team found regulators with scores ranging from 1 to 7. The regulator was assembled from four major components. The team split the causes into two families:

- Components: causes acting within the four components
- Assembly: causes acting in the process that puts the four components together

To eliminate one of these families, the team found two regulators, one that was extremely noisy with a score of 7 and a quiet one with a score of 2. They disassembled and reassembled the two regulators three times and measured the noise score each time. They found that the noise score did not change in either regulator. The team eliminated the assembly family and proceeded on the basis that the dominant cause acted in the components. We discuss this example in more detail in Chapter 11.

Door Closing Effort

High door closing effort was a relatively frequent complaint in a new vehicle owner survey. A team worked on the rear doors where there were more complaints. They measured door closing effort using a velocity meter that determined the minimum velocity (meters per second) necessary to close and latch the door. A baseline investigation showed there was considerable door-to-door variation in the velocity and that doors with high velocity were difficult to close.

Using the baseline data, the team selected two cars with extreme velocity values, one high and one low. The task of complete disassembly and reassembly of the door was difficult, so instead the team decided to remove and replace only the striker. For each car, this was repeated five times using production operators and tooling. The data are shown in Figure 10.14. The initial values are shown by the horizontal lines and indicate the full extent of velocity variation.

There was a large cause acting in both the assembly and components families. Within each car, there was considerable variation as the striker was removed and replaced but not the full extent of variation. The door-to-door average velocity differed considerably due to differences in the components of the two door systems. The team kept both families and worked on each separately. Since the assembly family consisted of only the position of the striker, they first took the striker assembly variation as one large cause and designed a new fixture to control this variation. Using the new fixture, there was a significant reduction in both the velocity average and standard deviation, but not enough to meet the problem goal. They also searched for the dominant cause of the remaining variation in the components family.

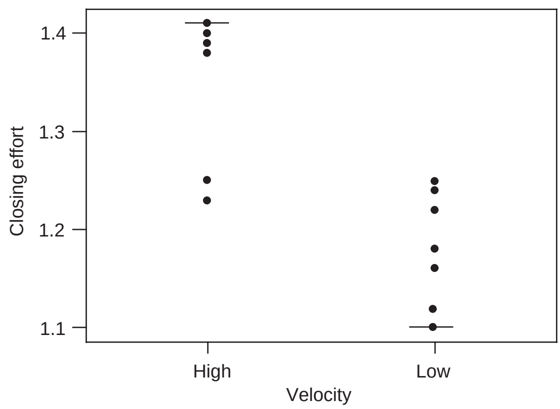


Figure 10.14 Closing effort by velocity (initial values from the baseline given by horizontal lines).

Comments

Figure 10.15 shows three possible plots of the results of a disassembly-reassembly investigation.

In the left panel of Figure 10.15, the dominant cause acts in the assembly family and in the right panel it acts in the component family. In the middle panel, there is an interaction between causes in the assembly and component families. For example, the presence of a burr on a component may make it difficult to assemble the unit so that it performs consistently. If the burr is absent, it is easy to assemble the product to get consistent performance.

There are several cautions when using this plan to compare assembly and component families. It is best to use the original assembly process, tools, and people to avoid study error. Otherwise there is a risk that any conclusions will not be relevant for the actual process. Also, we must ensure that parts are not damaged or changed in the disassembly-reassembly.

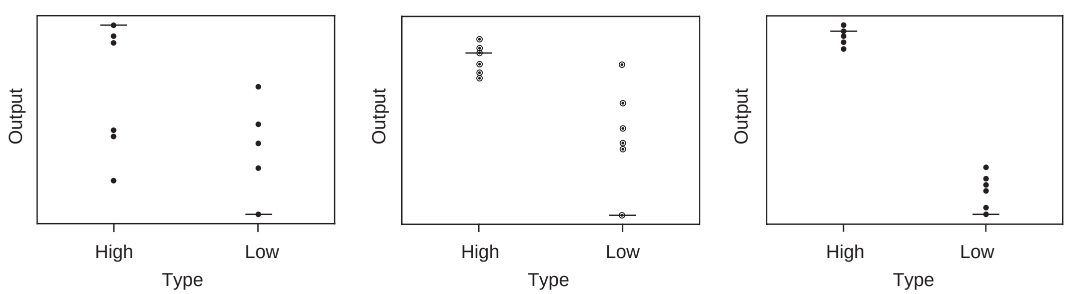


Figure 10.15 Hypothetical results for disassembly-reassembly investigation (initial value from the baseline given by horizontal lines).

Comparing Assembly and Component Families Summary

Question

In the current process, does the dominant cause act in the assembly or component family?

Plan

- Select two assemblies with opposite and extreme output values relative to the full extent of variation.
- Disassemble and reassemble each part three or more times. On each occasion measure the output.

Data

Record the original output values and each new measurement together with the corresponding part number, one measurement per row.

Analysis

- Plot the output by part. Use a special symbol to denote the original output values.
- Use the plot to compare the assembly variation within each part to the difference between the two original output values.

Conclusion

If the assembly variation within each of the two parts is relatively:

- Large, the dominant cause acts in the assembly family
- Small, the dominant cause acts in the component family

Otherwise the dominant cause involves inputs in both the assembly and component families.

10.5 COMMENTS

Binary Output

With a binary output characteristic and a low rate of defectives, we cannot look at the part-to-part family. In the block leaker problem described in chapters 1 and 6, the team made a false start. They sampled five consecutive blocks for 20 hours. They tested each block for leaks and found two leakers in separate hourly periods. From these data they could not eliminate either the block-to-block or hour-to-hour family because of the low failure rate. With a high-volume process, we can separate the hour-to-hour and day-to-day families by plotting the hourly rate of defectives by day.

Traceability and Order Preservation

In general, to stratify a process by time or process step we must be able to trace the parts measured to a specific time period or processing step. We must also ensure that the sampling plan matches the actual production order. If we call parts consecutive, then they should be processed consecutively as far as possible.

In the engine block porosity problem, the team had to specially mark the molds so that they could find the five consecutively molded blocks after the bank face machining. Once the blocks were poured, they were not necessarily finished and machined in pouring order. In a process that produced brake lines, the problem was a crack in a formed flange that led to leakage when the line was coupled to a valve. The crack was found at final inspection. The output characteristic was the size of the crack. The team considered an investigation to compare the part-to-part and hour-to-hour families. They abandoned this plan because the sequence of parts changed so many times in the process. Consecutive parts at final inspection were not processed consecutively through earlier process steps. The team could not isolate causes that acted in the part-to-part family because of this loss of order.

Formal Analysis

In this chapter we have described simple investigations supported by graphical displays such as scatter plots, box plots, multivari charts, and defect concentration diagrams. Since we are searching for a dominant cause, we can almost always draw the conclusion to eliminate a family without formal analysis procedures. Rarely do we need more formal methods such as the analysis of variance (ANOVA)[7] or regression analysis (see Chapter 12).

What If We Do Not Observe the Full Extent of Variation?

We must observe close to the full extent of variation in the output to draw conclusions from any investigation used in the search for a dominant cause. Otherwise, the dominant cause may not change during the investigation and we may only see the effects of nondominant causes. In QPDAC language, we have a study or sample error.

For example, suppose we are comparing part-to-part and time-to-time families. We measure five consecutive parts produced at the start of each hour and stop sampling after one day, as in the block porosity problem. We plot the data with two different vertical scales in Figure 10.16.

From the left panel of Figure 10.16, we might mistakenly conclude that the dominant cause lives in the hour-to-hour family. This conclusion is premature when we examine the right panel of Figure 10.16, where the dotted lines show the full extent of variation. Whatever the dominant cause and wherever it resides, we have not yet seen its full effect. We need to sample for a longer period or in a different way.

In the engine block porosity example discussed earlier, the team failed to establish a baseline in terms of their new porosity measure. As a result, they did not know the full extent of porosity variation. They fortunately were able to link the results of the multivari with the baseline, established in terms of reject rate. They then assumed the full extent of porosity variation was captured in the multivari investigation.

We can guarantee the full extent of variation by either using baseline data or using the idea of leverage where we compare units that are extreme with respect to their output values. We use leverage, for example, in investigations to compare the assembly and component families.

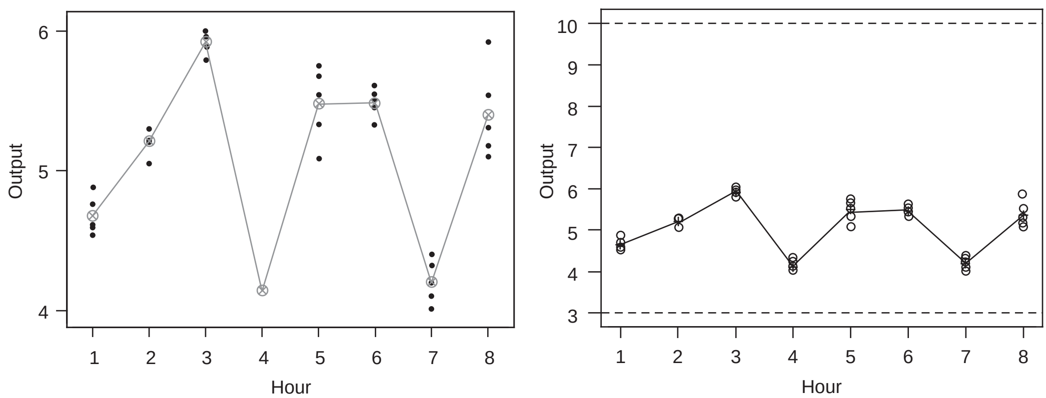


Figure 10.16 Two multivari charts of the same data with less than full extent of variation (dashed horizontal lines give the full extent of variation from the baseline).

Some Investigations Show Changes in Variation

Figures 10.2 and 10.9 showed that output variation changed by lobe and by time, respectively. Such patterns provide valuable clues about the dominant cause. For instance, in Figure 10.9, the difference in porosity variation of the output at hours 5 and 8 may be due to:

- A difference in the average or range of a dominant cause at different time intervals
- A different relationship between the cause and output in different time intervals
- A combination of these two reasons[8]

In both the examples, the change in variation was accompanied by a change in the average output level. To examine a situation where only the output variation changes, consider a hypothetical example. Suppose in the V6 piston diameter example discussed in Chapter 9, we collect 30 pistons from each stream over one shift so that we see the full extent of variation. We plot the results in Figure 10.17.

The average diameter is roughly the same for the two streams, so stream-to-stream differences are not a dominant cause of variation. However, stream *B* shows substantially greater diameter variation than does stream *A*. This pattern of unequal variation tells us a lot about the dominant cause. The dominant cause must act within the stream *B* family. We have eliminated all causes in the stream-to-stream and within-stream-*A* families. Note that this conclusion is more specific than the usual interpretation from stratification where we are able to rule out either the within-stream or stream-to-stream families.

What kinds of causes can explain the behavior seen in Figure 10.17?[9] Two possibilities are that there was a worn fixture in stream *B* or that the stream *B* operator made more frequent adjustments.

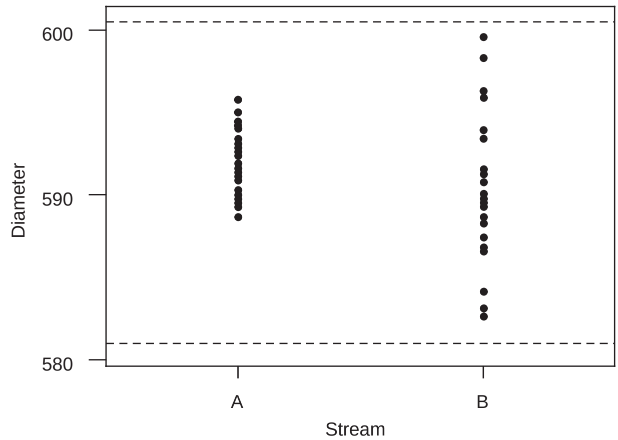


Figure 10.17 Box plot of diameter by Operation 270 stream from hypothetical investigation (dashed horizontal lines give the full extent of variation from the baseline).

Key Points

- Simple investigations and graphical analysis tools can provide valuable clues about the home of a dominant cause of variation.
- We often find it useful to divide the remaining possible causes into two families of causes such as:

  –Short-time and long-time
  –Group-to-group and within-group
  –Upstream and downstream
  –Assembly and components

  and then carry out a simple investigation to eliminate one family.
- To avoid study and sample errors, we must observe the full extent of variation in the investigation.

Endnotes (see the Chapter 10 Supplement on the CD-ROM)

1. If we have two families and stratified data, we can use analysis of variance (ANOVA) to partition the overall standard deviation into two pieces, one associated with each family. We look at the details in the supplement.

2. By understanding the possible reasons for an observed pattern in the results of an investigation, we can eliminate many suspects. In the supplement, we look more closely at possible explanations for the changing average and variation pattern seen in figures 10.2, 10.9, and 10.17.
3. We identify a dominant cause primarily using graphical displays. We do not use a formal hypothesis test. In the chapter supplement we elaborate on why hypothesis tests are not appropriate in the search for a dominant cause.
4. See note 2.
5. We used a simple plan to compare upstream and downstream families. We examine operations swap and randomized sequencing, two more restrictive plans for separating these families.
6. By selecting units with extreme performance, we can compare the assembly and component families with a very limited number of disassembly-reassembly cycles. We examine this application of leverage in more detail in the supplement.
7. See note 1.
8. See note 2.
9. See note 2.

Exercises are included on the accompanying CD-ROM

11

Investigations to Compare Three or More Families of Variation

Things which matter most must never be at the mercy of things which matter least.

—Johann Wolfgang von Goethe, 1749–1832

In Chapter 10, we compared two families of causes using simple plans and analysis tools. In this chapter, we consider investigations for simultaneously comparing three or more families to identify the home of a dominant cause of variation. We also consider the repeated application of investigations based on two families.

11.1 MULTIVARI INVESTIGATIONS: COMPARING TIME- AND LOCATION-BASED FAMILIES

For many processes, we can construct families of causes based on time and location. Suppose the goal of a project is to reduce variation in the weight of a molded part. The process has two four-cavity molds that operate in parallel, as shown in Figure 11.1.

We consider the following families:

Family	**Description**
Mold-to-mold	Causes that explain differences in weight from one mold to the other
Cavity-to-cavity	Causes that explain differences in weight among the cavities
Part-to-part	Causes that explain differences in weight among consecutive parts
Hour-to-hour	Causes that explain differences in weight from one hour to the next

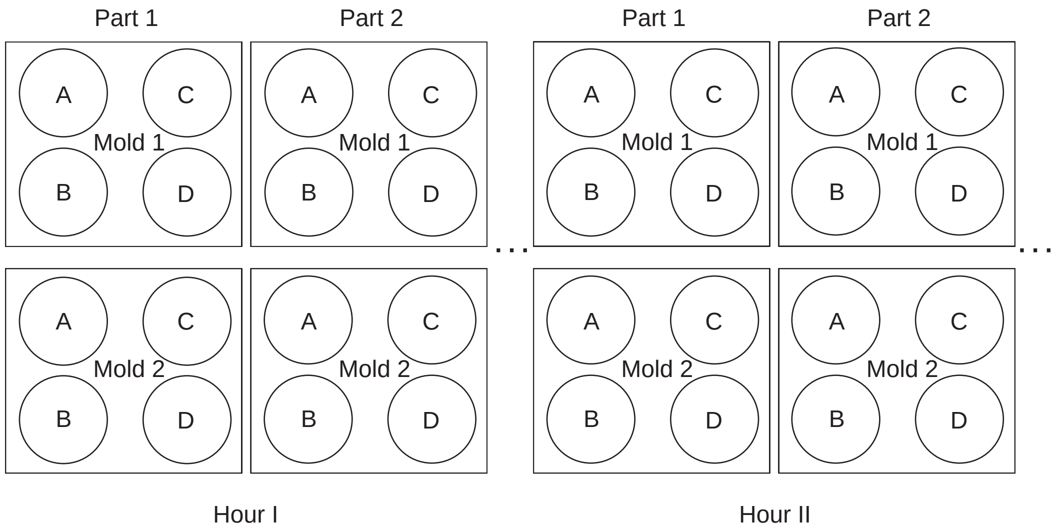


Figure 11.1 Cavity-to-cavity, mold-to-mold, part-to-part, and hour-to-hour families.

To isolate the effects of each family, we sample from the process in a systematic manner. We include parts in the sample that are produced consecutively and over hours from all molds and cavities. For example, we might select and measure five consecutive parts from each mold and cavity (40 parts in total) every hour for several hours.

With the proposed sampling plan, we can also detect interactions between the families. For example, we can see if mold-to-mold differences change from hour to hour.

We call this a multivari investigation (Seder 1950a, 1950b, 1990). See also Zaciewski and Nemeth (1995) and Snee (2001). We design the sampling protocol to see the variation due to the causes within each family of interest. We must be able to trace the parts according to the order and location in which they are produced. In the example, each part was labeled by cavity and mold. We may have more difficulty finding parts that were molded consecutively at the end of the process.

We use multivari charts (modified run charts) to display the results of a multivari investigation. Using MINITAB, we can choose to display the effects of up to four families on the multivari chart at the same time. Since charts with many families are difficult to interpret, we prefer to use a series of charts, each with fewer families. We can use analysis of variance (ANOVA) to quantify the effects of each family. ANOVA is useful when we cannot easily determine the dominant family from the multivari charts.[1]

We expect the causes in some families to act systematically. For example, if the dominant cause acts in the cavity-to-cavity family, we will see large differences in the average output across the cavities. In other families such as the part-to-part, we expect the causes to act haphazardly. We would be surprised to see large differences between the output averages (averaged over all cavities, times, and molds) of the first and second part sampled in different hours.[2] The effects of causes that change from part to part will likely produce haphazard variation. We need to plot the data carefully to avoid masking this type of variation.

We provide three examples of multivari investigations to illustrate the sampling plan, the presentation of the data, and the interpretation of the results.

Cylinder Head Scrap

Iron cylinder heads were cast in a green sand process using a four-cavity mold. In the jargon of the foundry, each cavity is called a pattern. The project was to reduce scrap at the machining operation due to a defect called *shift*. Shift occurs when the top and bottom parts of the sand mold are not properly aligned. Excessive shift causes problems with part location in the subsequent machining operation.

To measure shift, there are locator pads on the top and the bottom of the casting. Shift is determined by measuring the relative distance between the two pads in the x (side shift) and y (end shift) directions. The team recorded both measurements in inches for each part. Here we look at side shift only. The target value is 0.

The team carried out the baseline investigation by haphazardly selecting and measuring 20 castings per day from each of five days' production. The data are given in the file *cylinder head scrap baseline*. Since it was difficult to trace parts through the process, they recorded only the day of production.

We plot side shift stratified by day in Figure 11.2. The average side shift was –0.009 inches with a standard deviation of 0.017. The values varied from –0.050 to 0.035 inches. The team noted that the within-day variation was large.

The team goals were to better center the process and reduce side shift standard deviation by at least 50%. They expected these changes to reduce scrap substantially at the machining operation.

The team used a coordinate measuring machine (CMM) to make the measurements. They checked the system with a gage R&R investigation, carried out over a short time period, contrary to the recommendations in Chapter 7. The observed measurement variation was small relative to the baseline.

The team decided to look for a dominant cause and planned a multivari investigation to rule out a large number of suspects. They considered the four families shown on the diagnostic tree in Figure 11.3.

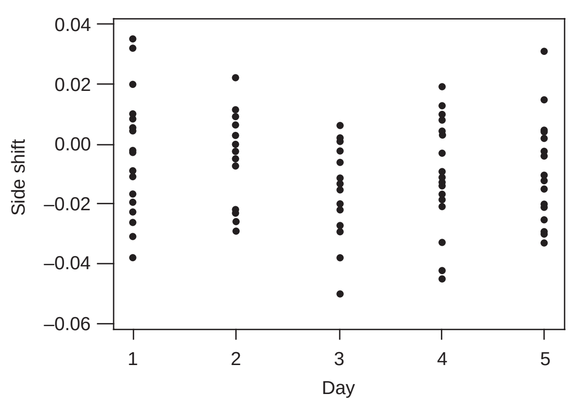


Figure 11.2 Plot of side shift by day.

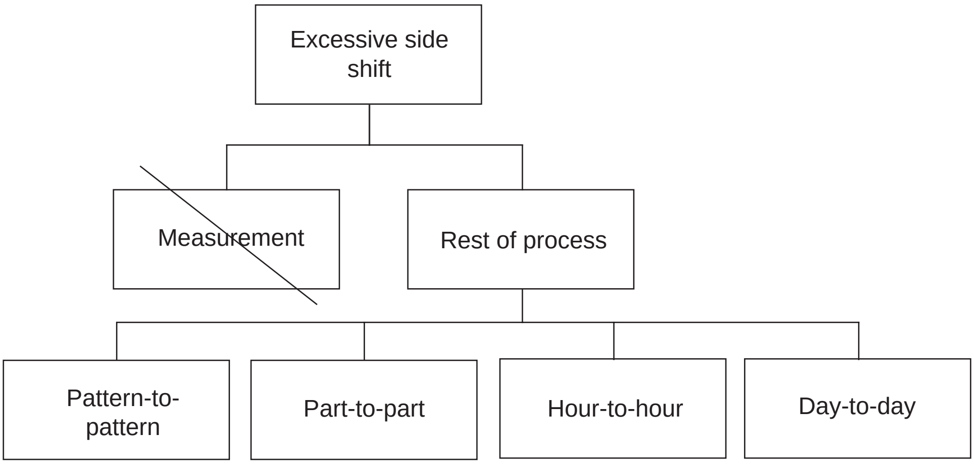


Figure 11.3 Diagnostic tree for side shift.

In order to assess the variation contributed by each of these families, they used the following sampling protocol that required tracking and measuring 96 parts:

1. Sample and measure four parts cast consecutively from each pattern.
2. Repeat step 1 for castings poured at hours one, three, and five of the day shift.
3. Repeat step 2 on two consecutive days.

With this sampling protocol, the team could isolate the contributions to the baseline variation from each family in the diagnostic tree. For example, sampling four consecutive parts showed whether the variation was dominated by a cause that changed quickly from one part to the next, such as the relative positioning of the two halves of the mold onto the guiding pins and bushings.

From the baseline investigation, the team expected the full extent of variation to occur within the two days. The major difficulty in implementing the plan was tracing castings molded consecutively. The team marked the sand molds so that they could find these castings at the end of the shakeout and cleaning process.

Side and end shift for the 96 measured parts are stored in the file *cylinder head scrap multivari*. For side shift, we see close to the full extent of variation in the histogram shown in Figure 11.4, so we know the dominant cause acted during the investigation.

We look at a variety of plots. To simplify the data display, we start by creating a new input called time that combines hour and day. See the section on multivari charts in Appendix C.

In Figure 11.5, the lines join the average side shift at the six times and the dots give the 16 side shift values at each time. We recommend that you always select the option (unfortunately not the default) to plot the individual values on MINITAB multivari charts. From Figure 11.5, we see:

- The time averages are almost constant; that is, there are no systematic differences in the side shift between days or hour to hour within days.
- The variation of side shift within each time is close to the full extent of variation.

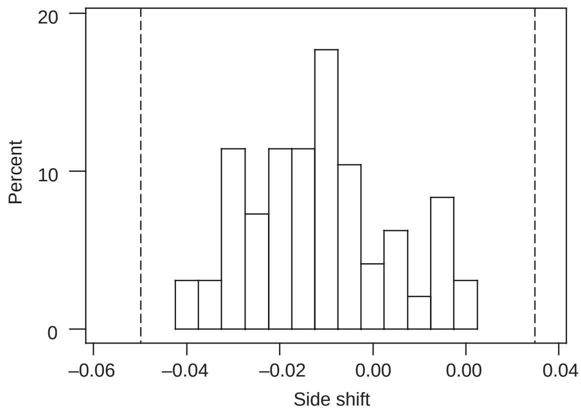


Figure 11.4 Histogram of side shift values from the multivari investigation (full extent of variation in the baseline given by dashed lines).

The observed pattern cannot be explained by a cause that changes slowly from time to time so we are tempted to eliminate the hour-to-hour and day-to-day families. We wait to do so until we look for possible interactions with the other families.

In the left panel of Figure 11.6, we look at a multivari chart with side shift stratified by pattern. From the chart, we see large systematic differences among the patterns. The variation within each pattern is only about 60% of the full extent of variation. The differences in the pattern averages contribute substantially to the baseline variation. We conclude that the pattern-to-pattern family contains a dominant cause.

To deal with the part-to-part family, we need to recognize that variation from consecutive molds is likely to be haphazard. We create a new input called *group* with a different value for each combination of the four patterns and six times. There are four side shift values within each of the 24 groups corresponding to the consecutive castings in that group.

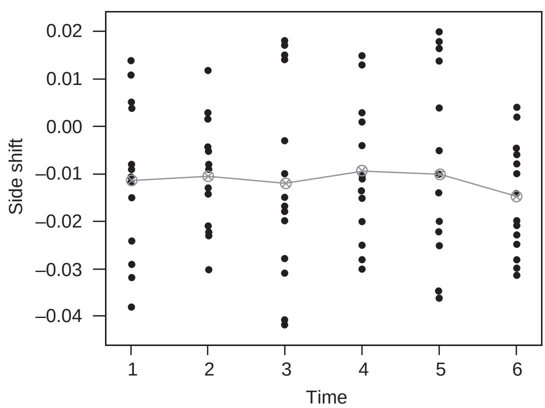


Figure 11.5 Multivari chart of side shift by time.

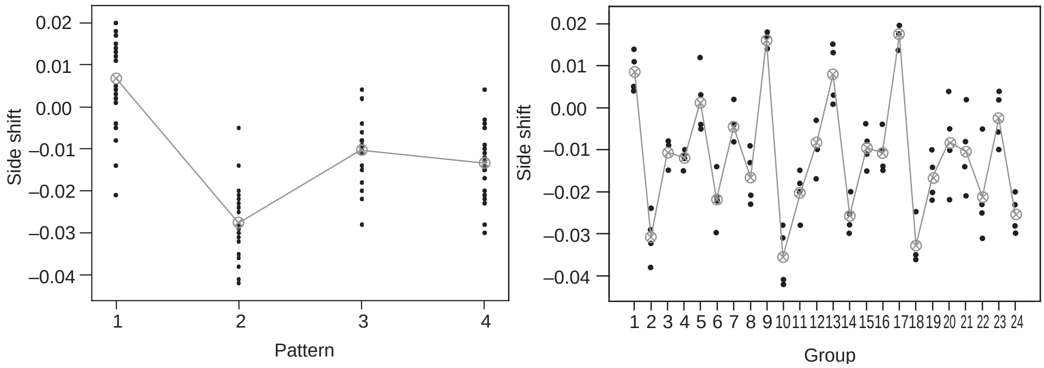


Figure 11.6 Multivari charts for side shift versus pattern on left, side shift versus group on right.

In the multivari chart in the right panel of Figure 11.6, the first six groups correspond to the six time points for the first pattern. The next six correspond to the second pattern, and so on. The chart shows that the variation within each group, as represented by the four dots, is small relative to the baseline variation. Alternatively, we see that the side shift averages vary substantially from group to group. The pattern-to-pattern family is included in the group-to-group family, which explains the group-to-group differences in side shift average seen in the right panel of Figure 11.6. We also see that the within-group variation is roughly constant over all sampling periods, which means that there is no evidence of interaction between the part-to-part family and the other families of causes. We eliminate the part-to-part family from further consideration.[3]

We now look for possible interactions with the pattern-to-pattern family by constructing multivari charts with pattern and time. In Figure 11.7, we show the side shift averages and individual values for each time and pattern. The effects of the time family are small for each

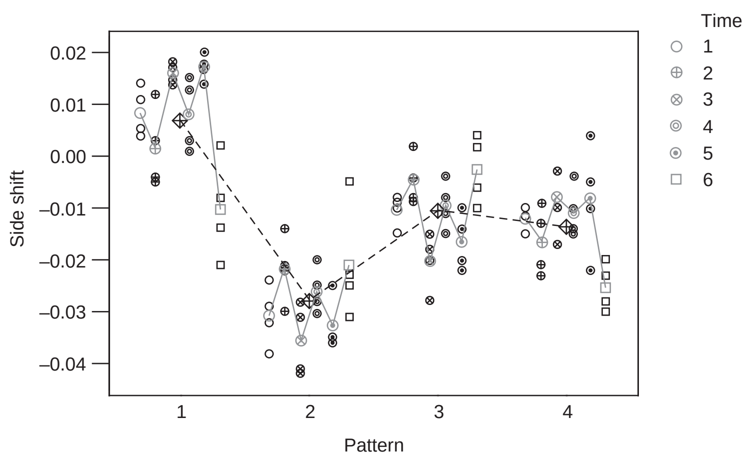


Figure 11.7 Multivari chart of side shift versus pattern and time.

pattern. There is no evidence of an interaction between causes in the pattern-to-pattern and time-to-time families.

We conclude that the dominant cause lives in the pattern-to-pattern family. To reduce the variation, we need to focus on causes within this family.

The team knew how to change the pattern averages by making a one-time adjustment to the dies. From the left panel of Figure 11.6, they saw that if they aligned all the patterns they would almost meet the project goal in terms of reducing side shift standard deviation. Based on the side shift averages in the multivari investigation, they recommended that the dies for the top half of the mold be moved to better center the process and remove the pattern-to-pattern differences. They also planned a weekly capability study to examine side shift by pattern to ensure that the averages did not drift apart again.

Camshaft Journal Diameter

In Chapter 7, we introduced a project to reduce scrap and rework due to excess variation in the diameter of camshaft journals. At final 100% inspection, the four journal diameters (the maximum diameter as the part was rotated) were measured at two locations, front and rear. The diameter specifications were ± 12.5 microns, measured from a target value. There were also specifications on out-of-round and taper for each journal. Camshafts that did not meet the specifications at the final gage were sent to a rework station. There the parts were remeasured and either scrapped or reworked. At the beginning of the project, the average monthly reject rate at the final gage was 4.7% of which 1.5% was scrap. Management set a goal to cut both these rates in half.

The team focused on journal diameter variation. There was no data storage at the final inspection gage. To establish a baseline, the team had two or three camshafts per hour set aside before being measured at the final gage to get a sample of 20 parts per day. At the end of the day, an operator measured and recorded the diameters for the four journals and two locations on each camshaft. This process was repeated for five days. The 800 measured values (from 100 camshafts) are stored in the file *camshaft journal diameter baseline*. The average diameter was 2.41 microns, the standard deviation was 5.00 microns, and the corresponding capability (P_{pk}) was 0.67. The operators had centered the process above zero to avoid scrap at the expense of rework. There were six parts needing rework and one scrapped in the baseline sample. The run chart of diameter over time did not show any clear patterns. The full extent of variation was –12.5 to 17.5 microns. The team set a goal to improve the process capability (P_{pk}) to more than 1.20.

The investigation of the final gage, described in Chapter 7, revealed that there was a significant bias in the head that measured the front diameter on the first journal. The team removed this bias with the expectation that the process variation would be reduced. They decided to repeat the baseline investigation before proceeding. The data are stored in the file *camshaft journal diameter baseline2* and the performance is summarized in Figure 11.8. The average, standard deviation, and P_{pk} for the second baseline investigation were 2.55, 4.53, and 0.73, respectively. Five out of the 100 parts required rework because of oversize diameters. The full extent of variation in the new baseline was –11.8 to 15.9 microns.

Before choosing a working approach, the team decided to look for a dominant cause of journal diameter variation. We give a high-level process map in Figure 11.9.

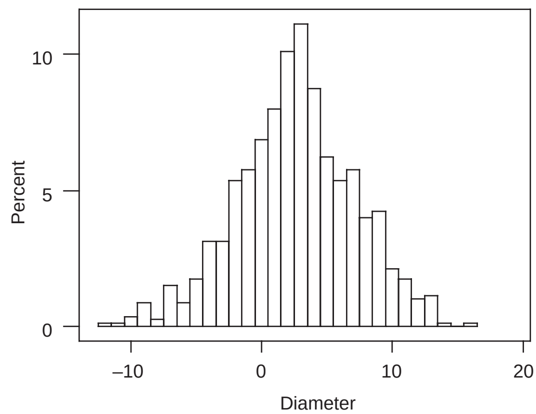


Figure 11.8 Histogram of journal diameter for second baseline investigation.

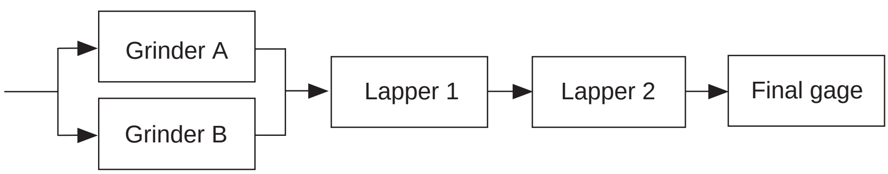


Figure 11.9 Camshaft production process map.

Blank parts arrived in batches from a supplier and were haphazardly assigned to one of the two grinders. The grinders used a feedback control system in which 1 part in 10 was measured (journal 1 only) and the grinders were adjusted if the diameter was out of specification. The within-process specification limits at the grinder were set based on the assumption that the lappers would reduce the diameter by 24 microns. Each grinder had its own gage.

The team planned a multivari investigation to examine the following families:

Family	**Description**
Position-to-position (within-part)	Causes that explain differences in diameter from one position to the other within the same camshaft
Grinder-to-grinder	Causes that explain differences in diameter between grinders
Part-to-part	Causes that explain differences in diameter among consecutive camshafts
Hour-to-hour	Causes that explain differences in diameter from one hour to the next
Batch-to-batch	Causes that explain differences in diameter from one batch of incoming blanks to the next

They decided to use three batches (about one day's production) of blank parts and to sample parts once every two hours. They selected five camshafts from each grinder within

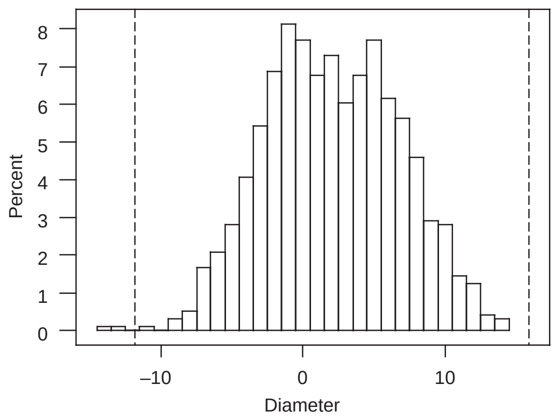


Figure 11.10 Diameter variation in the multivari (dashed lines give the full extent of variation).

an adjustment cycle (part numbers 1, 3, 5, 7, and 9). These parts were specially marked so that they could be found at the final gage. They planned to collect 120 camshafts and make 960 diameter measurements in total. The team carried out the plan without any difficulties.

The data are recorded in the file *camshaft journal diameter multivari*. We see the full extent of variation in the histogram shown in Figure 11.10. We know the dominant cause acted during the investigation.

We give the multivari charts for position, grinder, and the combined hour and batch families in Figure 11.11. We expect these families to show systematic differences in the averages if they contain a dominant cause. There is a large difference between the two grinder averages indicating a possible dominant cause. There is no evidence of a dominant cause acting in the other families.

Since a dominant cause appears to act in the grinder-to-grinder family, we look for an interaction with the position or batch/hour families using the multivari charts in Figure 11.12. Since the effect of grinder does not appear to depend on position or batch/hour, we see no evidence of any interactions.

We expect the part-to-part variation to be haphazard.[4] To display the magnitude of this variation, we create a new input called group that indexes the 192 sampling points (8 positions by 2 grinders by 3 batches by 4 hours). Since there appears to be a large cause in the grinder-to-grinder family, we define the group input so that the first 96 values correspond to grinder *A* and the second 96 values correspond to grinder *B*.

We give the multivari chart of diameter by group in Figure 11.13.

This chart is difficult to interpret because there are so many groups. However, we can see that the variation around the averages is large, which indicates substantial part-to-part variation. The magnitude of this variation is roughly the same for each group of consecutive diameter measurements at a particular position and from grinder to grinder, which indicates a lack of interaction. We find ANOVA useful here to quantify the effects and augment the multivari charts.[5] We conclude there are dominant causes within the grinder-to-grinder and part-to-part families.

The team concentrated their efforts on these two families in the search for the dominant causes. The team was surprised by the grinder-to-grinder differences. They had not seen this

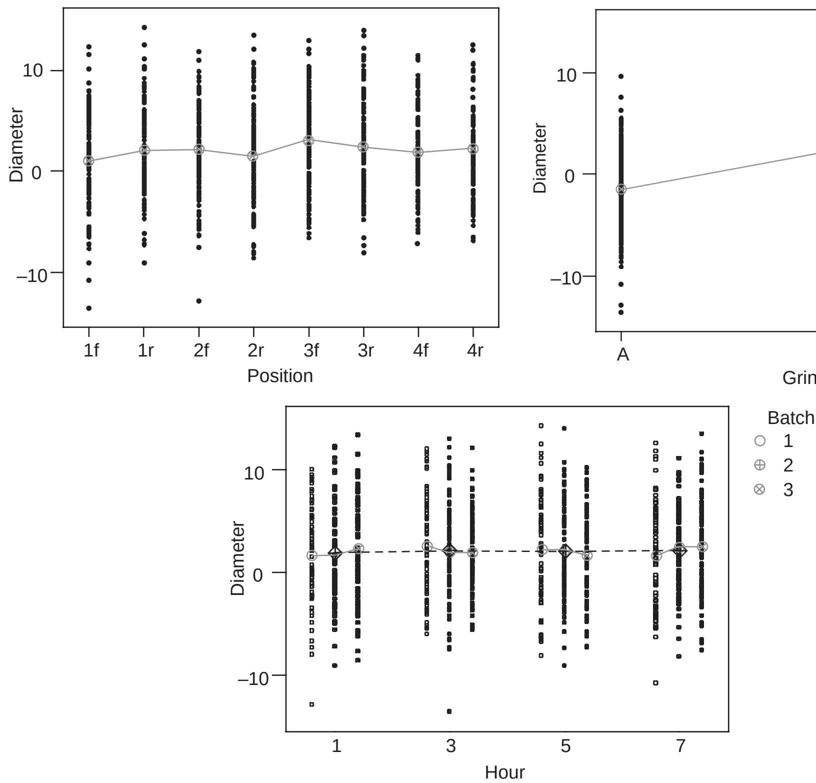


Figure 11.11 Multivari charts of diameter by position, grinder, and batch/hour.

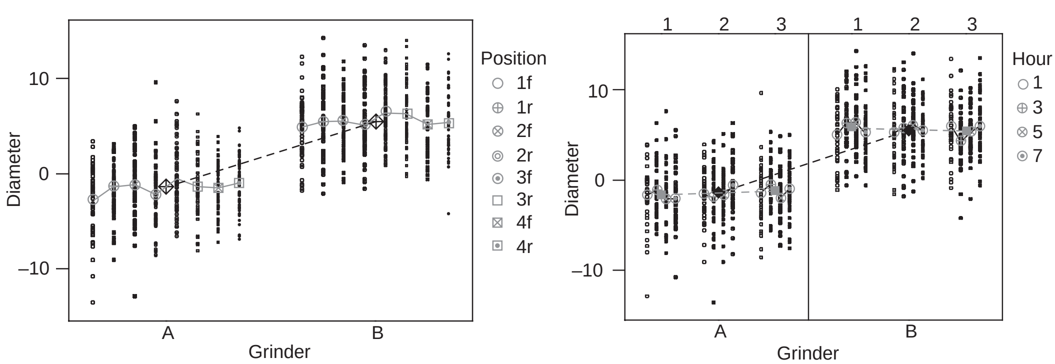


Figure 11.12 Multivari charts of diameter by grinder versus position and batch/hour.

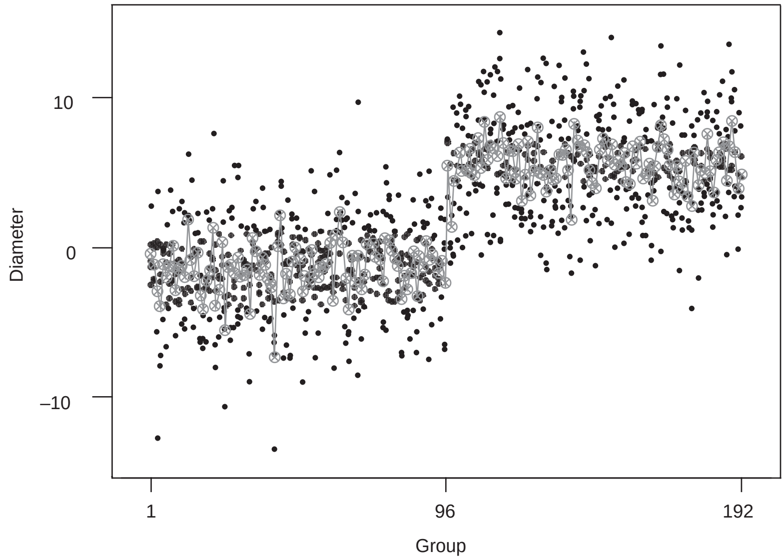


Figure 11.13 Multivari charts of diameter by group.

difference in regular production data because normally there was no traceability to the grinder.

Fascia Cratering

Management assigned a team to reduce scrap and rework on front and rear fascias because of cratering. This was a new product, and during the startup period the reject rate was 8.9%. The goal was to reduce this rate to less than 1%. The crater defects were not visible directly after molding but they could be clearly observed after the fascia was primed. The crater rejects occurred over all shifts and were haphazardly distributed over time. The team developed and checked a measurement system to count the number of craters after priming.

In a baseline investigation, the team found that about 80% of the fascias had no craters, 10% had fewer than 25, and the remaining 10% had more than 25 craters. They decided to search for a dominant cause of the craters. From the baseline sample, the team constructed a defect concentration diagram that revealed that 75% of the craters were located in the front half of the fascia. However, the team could not see how to use this knowledge to rule out any family of causes.

The fascias were produced on two different molding machines. The team planned a multivari investigation to examine the fascia-to-fascia, machine-to-machine, and time-to-time families. They sampled five consecutive fascias from each of the two machines every four hours. They continued the sampling for three shifts (24 hours) resulting in a sample of 60 fascias. To keep track of the consecutive parts and machine numbers, they marked each fascia on the inside immediately after molding.

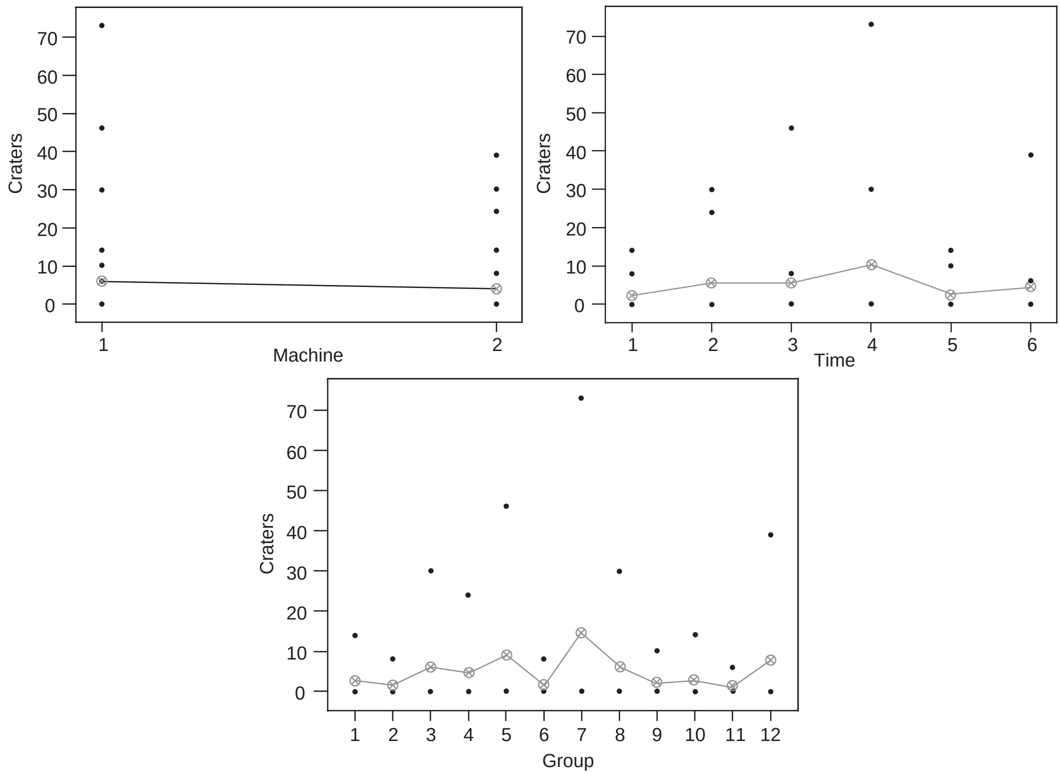


Figure 11.14 Multivari chart from first fascia cratering multivari investigation.

The data are given in *fascia cratering multivari*. The distribution of the number of craters matches the baseline closely. To look at the fascia-to-fascia (part-to-part) family, we create a new input group, with a different value for each combination of machine and time. We give multivari charts for the time-to-time, machine-to-machine, and fascia-to-fascia families in Figure 11.14.

There is no evidence that the dominant cause acts in the machine-to-machine or time-to-time families. There is large fascia-to-fascia variation in some of the groups and little in others.

An observant team member noticed that there were only two values plotted for each group of five fascias. When the team examined the data more closely, they saw that there was only one fascia in each group with a value different from 0. This was a very strong clue about the dominant cause because the team could only think of one process step where something happened only once in every five parts. The operator sprayed mold release every 10 pieces. After five pieces he applied a minor spray. Mold release was used to prevent tearing when the fascia was removed from the mold.

In a follow-up investigation, the team sampled 10 consecutive fascias in the molding sequence at three different times in the morning shift. The data are given in the file *fascia cratering multivari2*. The team confirmed mold release as the dominant cause of cratering as shown in the Figure 11.15. With knowledge of the dominant cause, the team needed to choose a working approach. We continue the story in Chapter 14.

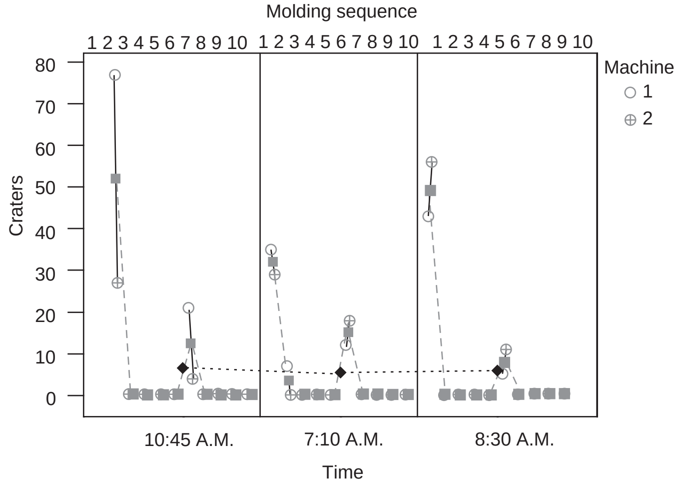


Figure 11.15 Multivari chart from second fascia cratering multivari investigation.

Note that in this example the fascia-to-fascia family of causes results in systematic variation and not haphazard variation as in most applications where we have a part-to-part family.

Multivari Investigation Summary

Select families to investigate by grouping the remaining suspect dominant cause.

Question

For the current process, which of the selected families is home to a dominant cause?

Plan

- Specify a study population long enough to capture the full extent of variation.
- Select a sample using a systematic sampling protocol designed to capture the variation due to the chosen families.
- Spread out over the study population.

Data

Record the output value and the corresponding families, one output per row.

Analysis

- Plot a histogram of the data. Check that the range of output values covers most of the full extent of variation.
- For families likely to have systematic effect, construct the single-family multivari charts, plotting individual output values.

- For the part-to-part family (or other families expected to have haphazard effect), create a new input group so that the effect of the family is seen in the variation within each value of group. Construct the multivari chart with *group*.
- For each family identified as having a large effect, construct two family multivari charts to look for evidence of interaction.

Conclusion

- Identify the dominant family (or families if the dominant cause acts in two or more families).

Comments

A multivari investigation is a powerful tool for assessing the contribution of families defined based on time or location.

In planning a multivari investigation, keep the number of families to five or fewer so that the sampling plan and analysis are not too complex. The camshaft journal diameter multivari illustrates the difficulties.

In a multivari investigation, we do not use random sampling; instead we select parts deliberately to estimate the effects of the specified families. We may need to sample additional parts beyond those collected for the normal control or monitoring of the process.

For any family in a multivari investigation, we need traceability. In the cylinder head example, we were interested in the mold-to-mold family, that is, variation in consecutive pieces from the same pattern. To assess this family, we must be able to identify the heads produced consecutively from each pattern when we measure the output. If we cannot trace parts through the process easily, then for time-based families, we need first-in, first-out discipline at each process operation so that the time sequence is not lost when we select parts at the end of the process.

Multivari investigations are not effective for binary output characteristics unless the rate of defectives is high. In the block leaker project, introduced in Chapter 1, the baseline rate of leaking blocks was 2.2%. The team initially conducted a multivari investigation in which five consecutively poured blocks were tracked through the finishing process every two hours for several days. After 160 blocks, they found only one leaker. There was no useful information in the multivari investigation for this rare defect.

Sometimes the recommended analysis process fails because we do not find a dominant family when we look at one family at a time. In this case, we recommend constructing all two-family multivari charts and looking for interactions.

If the multivari charts fail to reveal the dominant family, we suggest a formal analysis of variance (ANOVA) to quantify the relative contributions of each family.[6] We may then decide to search for causes within one or more of the families with the largest contributions, or we may abandon the partition of causes in these families and start over with a new set of families.

11.2 COMPARING FAMILIES DEFINED BY PROCESSING STEPS

Most manufacturing processes consist of several steps. We can form families of causes based on inputs that change within each of these steps. In Chapter 10 we looked at some examples where the process was split into two families, causes acting upstream and downstream of a particular point in the process. Here we consider applying this idea repeatedly.

We consider two types of plans:

- Measure the output characteristic on the same parts after each processing step.
- Allow some parts to skip processing steps and measure the final output characteristic.

The first type of plan can be applied more commonly than the second. Consider the following examples.

V6 Piston Diameter

In Chapter 5, we described a problem to reduce variation in the diameters of machined aluminum pistons. In the baseline investigation, the final diameter after Operation 310 varied between 581 and 601 microns, measured from a fixed value. In searching for a dominant cause, the project team decided to define families based on a number of processing steps as shown in Figure 11.16. The processing steps in Figure 11.16 include all those that affect diameter.

To eliminate as many processing steps as possible, the team tracked 96 pistons through the process, measuring the diameter after Operation 200 and after each of the steps shown in the process map. In total they measured each piston diameter six times. The team carried out the investigation over three days, 32 pistons at a time. During this period, the process operators collected the marked pistons for measurement after each operation.

From the baseline investigation, the team expected that they would see close to the full extent of variation over the three days. They decided to use only 32 pistons each day to limit the disruption to the normal operations. The team used the in-process gages for the measurements after Operation 200 to Operation 260 and the Operation 310 gage for the final three measurements.

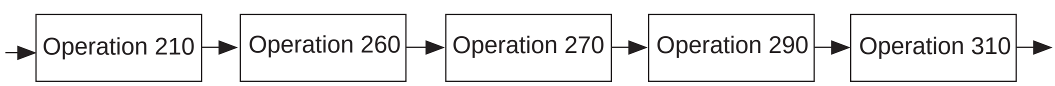


Figure 11.16 Piston machining process map.

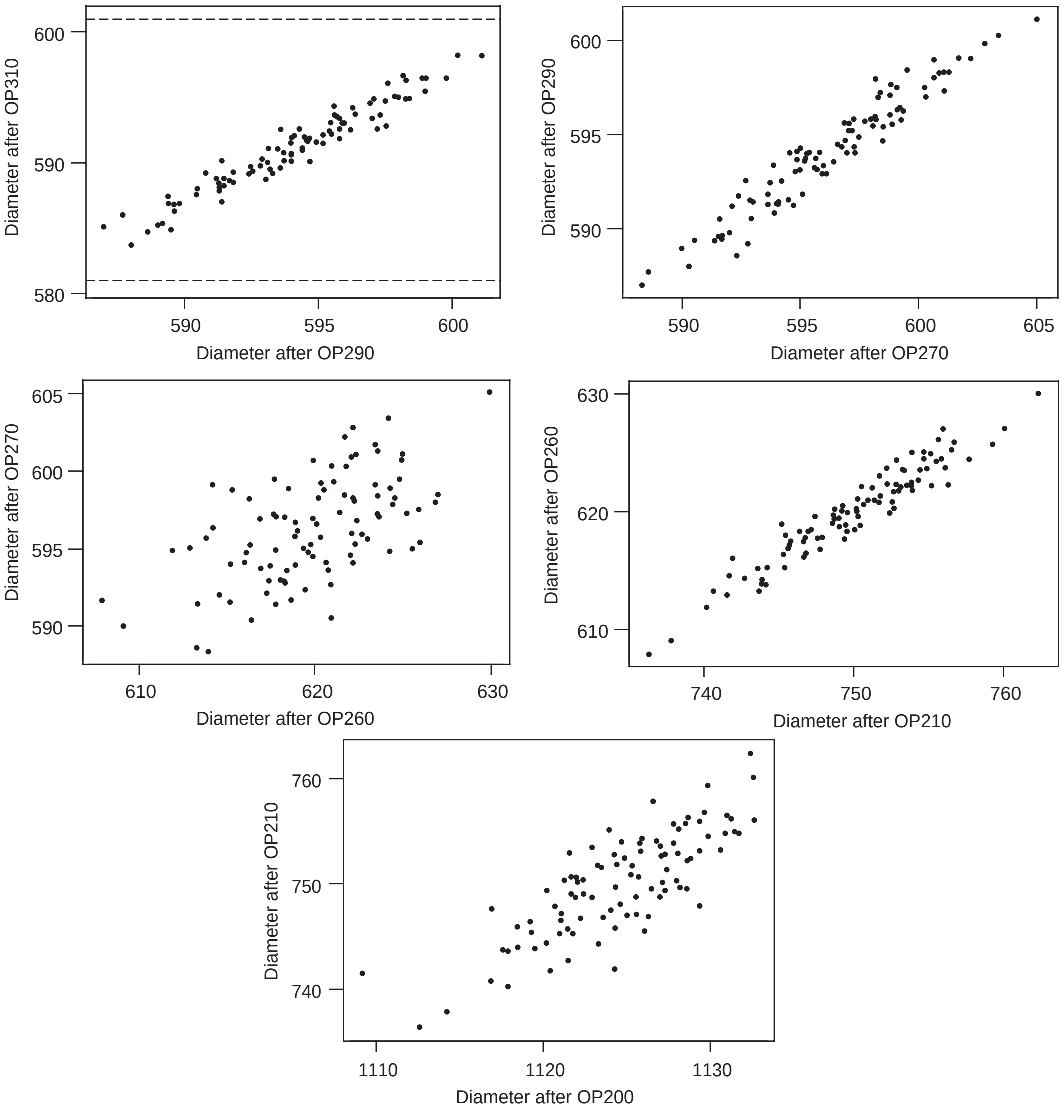


Figure 11.17 Scatter plots of outgoing versus ingoing diameter (by operation; dashed horizontal lines give the full extent of variation).

The team converted all measurements, given in the file *V6 piston diameter variation transmission*, to the same units as at the final gage. The team produced the scatter plots of the incoming and outgoing diameter for each operation, as shown in Figure 11.17. They added horizontal lines to show the full extent of variation to the plot that shows the final diameter.

We interpret the scatter plots in Figure 11.17 starting from the end of the process. From the top left diagram, we see the full extent of variation in the final output. We also see that most of the variation in the final diameter is transmitted through Operation 310. If we could eliminate the variation in the diameter after Operation 290 so that the output is, say, 595, then from the graph we see there would be very little variation in the diameter after Operation 310. That is, little diameter variation is added in Operation 310; most of the variation is transmitted from upstream. We can eliminate Operation 310 as the home of a dominant cause. From the top right plot, we see similarly that most of the variation in the diameter after Operation 290 is due to variation at Operation 270 and, hence, we can also eliminate Operation 290.

The left plot in the second row of Figure 11.17 has a different pattern. At Operation 270 there is substantial variation added to the process. If the diameters after Operation 260 were fixed at, say, 620, there would still be substantial variation after Operation 270. Hence we eliminate all operations upstream of Operation 270 and conclude the dominant cause of the variation in the final diameter is in the Operation 270 family. Next, the team focused their efforts on Operation 270 to isolate the dominant cause of variation.

In this example, it was possible to measure the diameter of each piston at various points throughout the process. The pistons were traceable through the process by implementing special measures during the investigation. Also, in this example, the team measured the diameter after each operation in a single investigation. They could have adopted another strategy and, for example, measured the diameter after Operation 270 and after Operation 310 only. Then, they would repeat the investigation, splitting the remaining family. This way they would compare only two families at a time, as in Chapter 10. In the example described here, it was easier to track pistons through in a single investigation process.

We can use regression analysis to quantify the amount of variation added and transmitted at each stage if the pictures do not tell a clear story.[7]

Roof Paint Defect

In a painting process, there was a defect, called a *line*, on the roof of a finished car. The defect rate was about 10% and the repair costs were high. In an earlier application of the algorithm the team discovered that by reducing the film build (thickness) of the clear coat, the last step in the multistep painting process, the defect rate was reduced to less than 1%. However, this process change resulted in an inferior overall appearance of the painted surface. The team decided to reapply the algorithm to find a solution to the line defect problem that would allow a higher film build.

The team used the baseline defect rate of 10%. The goal was to achieve a defect rate of less than 1% with the original specifications for the clear coat film build.

The team decided to look for a dominant cause of the defect. They could not restore the clear coat film build specification in the production process because of the high cost of rework. As a consequence, the team decided to use test panels instead of car roofs as the study units because they could apply a high film build to the panels as in the original process. They mounted the panels on the roof of a dummy vehicle where the line defect normally occurred. To check this procedure, the team ran a set of panels through the process with the clear coat film build above its original level. Under these conditions, the defect occurred on almost every panel. The team was confident that knowledge acquired using the panels could later be applied to vehicles. In our language, they felt that there would be little study error.

The painting process has five major steps as seen in Figure 11.18.

Figure 11.18 Painting process map.

The team decided to search for a dominant cause by first isolating where in the process the defect occurred. They divided the causes into five families corresponding to the major processing steps.

In the first investigation, the team masked three panels until after the e-coat step. Three other panels were processed normally. The defect occurred on all six panels. The team eliminated the incoming metal, the phosphate step, and the e-coat step from consideration. In a second investigation, they processed another six panels normally up to the end of the e-coat step. Three of the panels were processed through the primer stage but not the color coat. The other three were not primed but color coated as normal. Clear coat was applied to all the panels. Again the defect was present in all six panels. The team concluded that the primer and color coat steps could be eliminated and hence the clear coat process was the home of the dominant cause. This was not surprising given the first solution to the problem.

The two major substeps within the clear coat process were paint application and oven cure. To split these two steps, three panels were removed from the test car after painting and were cured in a laboratory oven. None showed the defect. The remaining three panels were processed normally and as expected, all showed the defect. The dominant cause lived in the clear coat cure oven. The team continued to split the process using similar trials. Eventually, they isolated the airflow from exhaust ducts within a particular zone of the oven as the dominant cause of the line defect.

The team then carefully tested a minor modification to the ducts that would change the airflow pattern over the roof of painted vehicles. Twenty vehicles were painted in the modified process with high clear coat film build. There were no line defects and no other noticeable negative side effects. The change was made permanent. The clear coat film build was increased with a marked improvement in appearance. There were no occurrences of the line defect in the new process.

In this problem, the team repeatedly split the process and homed in on the dominant cause. Because of the nature of the process and the defect, they could skip or alter steps of the process and still produce painted panels with or without the defect. This would not be possible in many processes.

Variation Transmission Investigation Summary

Question

For the current process, which process step is the home of a dominant cause?

Plan

- Specify a time frame over which we expect to see the full extent of variation in the output.
- After the first process step, select a sample of 30 or more parts spread over the time frame.

- Measure the characteristic corresponding to the output on each part in the sample after each process step.

Data

Record the measured output values, one row for each part.

Analysis

- For each process step (after the first), plot the output after the step versus the output before that step.
- For the plot showing the output from the final process step, add horizontal lines showing the full extent of variation.

Conclusion

- If the full extent of variation is not observed, the process step containing the dominant cause cannot be identified.
- Starting at the last process step, eliminate any step where most of the variation is transmitted. The first process step where the variation is not transmitted is the home of the dominant cause.

11.3 COMPARING COMPONENT FAMILIES

In Chapter 10, we looked at problems defined in terms of the function of an assembled product. We separated the contribution of the assembly and the component families by repeatedly disassembling and reassembling two assemblies. If the dominant cause lies with the components, we can conduct further investigations to pinpoint the particular component family that is home to the dominant cause. In this section we look at such investigations.

We call the individual pieces that make up the assembly *components*, a specified set of components a *group*, and all components known not to contain the dominant cause the *housing*. To start, there may be no housing if all components are suspect. The number of components in the housing grows as we eliminate components.

To compare component families, we find two assemblies with opposite and extreme performance in terms of the output of interest; that is, we use leverage as discussed in Section 9.3. With this plan the output of the two assemblies covers the full extent of variation. Then, we repeatedly swap groups of components between the two assemblies and remeasure the output.

We demonstrate the conduct, analysis, and interpretation of a component swap investigation with two examples.

Headrest Failure

Management initiated a project to reduce the frequency of customer complaints about seat headrests that would not stay in a set position. In focusing the problem, the team demonstrated that the problem could be solved by eliminating headrests that required a force of less than 35 Newtons (N) to move through their full range of motion.

The team carried out a baseline investigation. To meet their goal, they wanted to reduce variation in force rather than increase the average force to avoid headrests that would be very hard to move; that is, they wanted to change only those headrests with low output. They assessed the force measurement system and found that there was little contribution from this system to the baseline variation. They decided to search for a dominant cause of force variation.

The team selected two seat assemblies, one with high force (70.3 N) labeled *H*, and one with low force (25.8 N) labeled *L*, relative to the baseline. The seat assembly consists of a headrest and a seat. To assess the assembly process (as in Chapter 10), the team removed and reinstalled each headrest three times in its original seat. The force was measured after each reinstallation. The data are:

Original force	After first reassembly	After second reassembly	After third reassembly
H (70.3)	72.5	74.2	71.9
L (25.8)	25.9	24.3	25.4

Because there were only small changes after reinstallation, the team ruled out the assembly process and concentrated on the components. Next, they swapped the two headrests between the seat assemblies and measured the force. For this and subsequent component swaps we display the data in a table where:

- Diagonal cells contain the force measurements from the assembly-reassembly phase.
- Off-diagonal cells contain the force measurements after the components are swapped.

For the example the data are:

	Headrest	
Seat	**H**	**L**
H	70.3, 72.5, 74.2, 71.9	74.6
L	26.7	25.8, 25.9, 24.3, 25.4

Changing the headrest does not change the force. We say the *performance follows* the seat, since the seat from the original high (low) force assembly results in high (low) force even with a different headrest. The team eliminated the headrest family and looked at the seat in more detail.

The seat was assembled from three components: guides, springs, and the actual seat. The team disassembled and reassembled the two seats into the three components three times and measured the force. The data are shown in the diagonal cells of the following table. The team eliminated the assembly process for the three components as the home of the dominant cause. During the seat disassembly-reassembly investigation, the team noticed differences in the shape of the springs in the two assemblies. They next swapped the springs in the two seat assemblies and measured the force. The data are given in the table.

	Spring	
Other components	**H**	**L**
H	71.4, 72.5, 72.3	25.5
L	75.2	25.9, 25.5, 24.3

The dominant cause acts in the spring family of causes. The diagnostic tree in Figure 11.19 summarizes the search for a dominant cause.

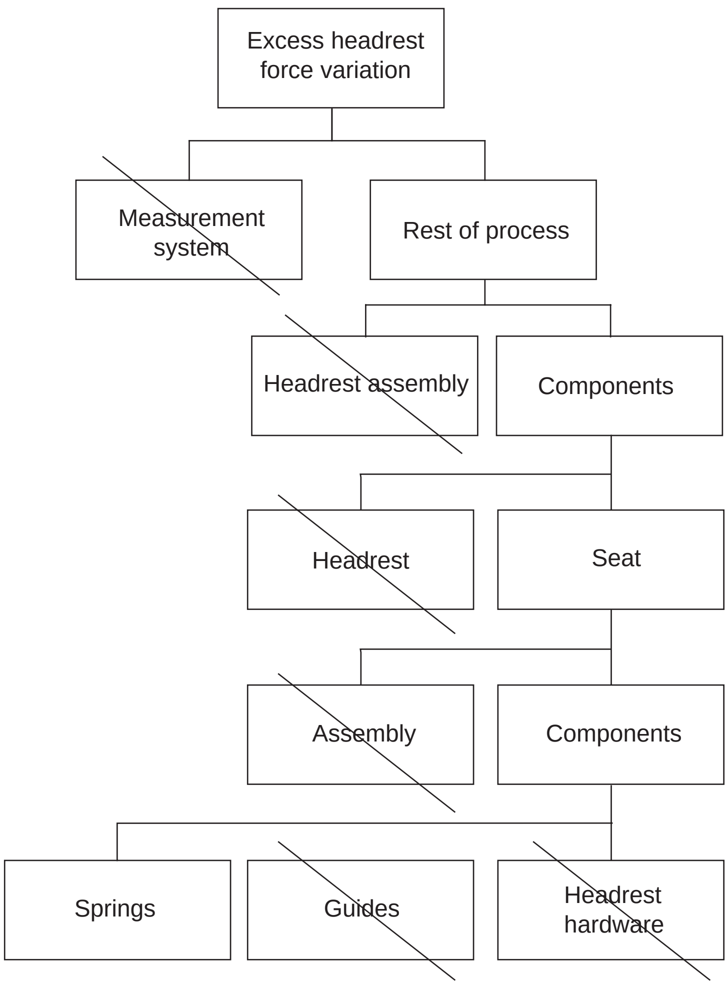


Figure 11.19 Diagnostic tree for headrest failure example.

The team worked with the supplier to understand the differences in the springs. They discovered that the supplier had reworked the springs in the low-force seats. Together, the supplier and customer developed a change of the rework process so that the force associated with the reworked springs exceeded 35 N. Customer complaints about the headrest moving were virtually eliminated.

Power Window Buzz Noise

There were frequent complaints about a noise fault called *buzz* in power window regulators (the motor and linkage that opens and closes the window). The customer, an assembly plant, demanded that the supplier implement 100% inspection to prevent noisy regulators from reaching the assembly plant. Management assigned a team the goal of reducing or eliminating the defect so that the costly inspection could be removed.

Trained listeners measured buzz subjectively on a scale of 1 to 7. They convinced their customer that if the buzz score was less than 4 on all regulators, then they could eliminate the 100% inspection.

The team did not carry out formal baseline or measurement system investigations. They had informal baseline information from past experience since they knew the buzz noise level varied between 1 and 7. They could ensure the dominant cause was acting in any investigation by selecting regulators with noise levels at the extremes. The decision not to check the measurement system was more risky. Large differences in measured noise for equally noisy regulators would make finding a dominant cause difficult. To address this concern, the team used a single listener to assign the noise levels for all regulators in the subsequent investigations.

They decided to search for the cause of the buzz. They divided the causes into two families, assembly and components. They selected a noisy regulator (score 7) and a quiet regulator (score 1). They could take the regulator apart without damaging or changing any of the major components. There was no change in buzz score when they disassembled and reassembled the two regulators three times each using the normal assembly sequence. The team eliminated the assembly process and concentrated on the components family.

The regulator has six components: arm one, arm two, spring, motor, back-plate, and sector. The team felt the motor was the likely home of the dominant cause of noise. They swapped the motors in the two regulators and measured the noise. The motor with the components from the originally quiet regulator had a score 7 and the motor with the other noisy components had a score 1. Indeed, the dominant cause of buzz lived in the motor.

The motor was a purchased part and comprised 18 components that could be disassembled without damage. The team formed two groups of nine components each, labeled G1 and G2, based on their limited knowledge of how the motor worked and the ease of disassembly. They did not know the details of the assembly process used by the motor supplier, so they ignored the motor assembly family for the moment. They next swapped all of the G1 components simultaneously and remeasured the buzz. The results were:

	G1	
G2	**Noisy**	**Quiet**
Noisy	7	7
Quiet	2	1

The entries 7 and 1 on the diagonal are the initial buzz measurements from the two original regulators. The off-diagonal entries are the buzz measurements for the regulators with all components in G1 swapped. Note that to measure buzz, the motors were assembled together with the other five components from the original quiet regulator. Since all components other than the motor had been eliminated as possible homes of the dominant cause, the choice of housing for the motor was not important.

The buzz followed G2 so all components in G1 were eliminated from consideration. Now G2 was split into two groups, G21 and G22, and the team found that the dominant cause lived in G22, a group of four components, three gears in the drive train, and the armature shaft.

When the first and second gears were swapped together, the buzz only occurred for the combination of all the G22 components from the originally noisy motor. The data were:

	First and second gears	
Gear 3 and armature	**Noisy**	**Quiet**
Noisy	7	3
Quiet	2	1

Now the picture is less clear. This pattern suggests a dominant cause involving one (or both) of the first and second gears, and one (or both) of the third gear and the armature. In other words, the dominant cause involves one or more components in each of the two remaining families.

The team returned the components to their original motor. They next simultaneously swapped the second and third gears with the results:

	Second and third gears	
First gear and armature	**Noisy**	**Quiet**
Noisy	7	1
Quiet	7	1

The buzz followed the second and third gears together. The dominant cause of buzz is an interaction between these two gears. Swapping only one of these two gears does not produce a complete switch in the buzz. We show the diagnostic tree for the search to this point in Figure 11.20.

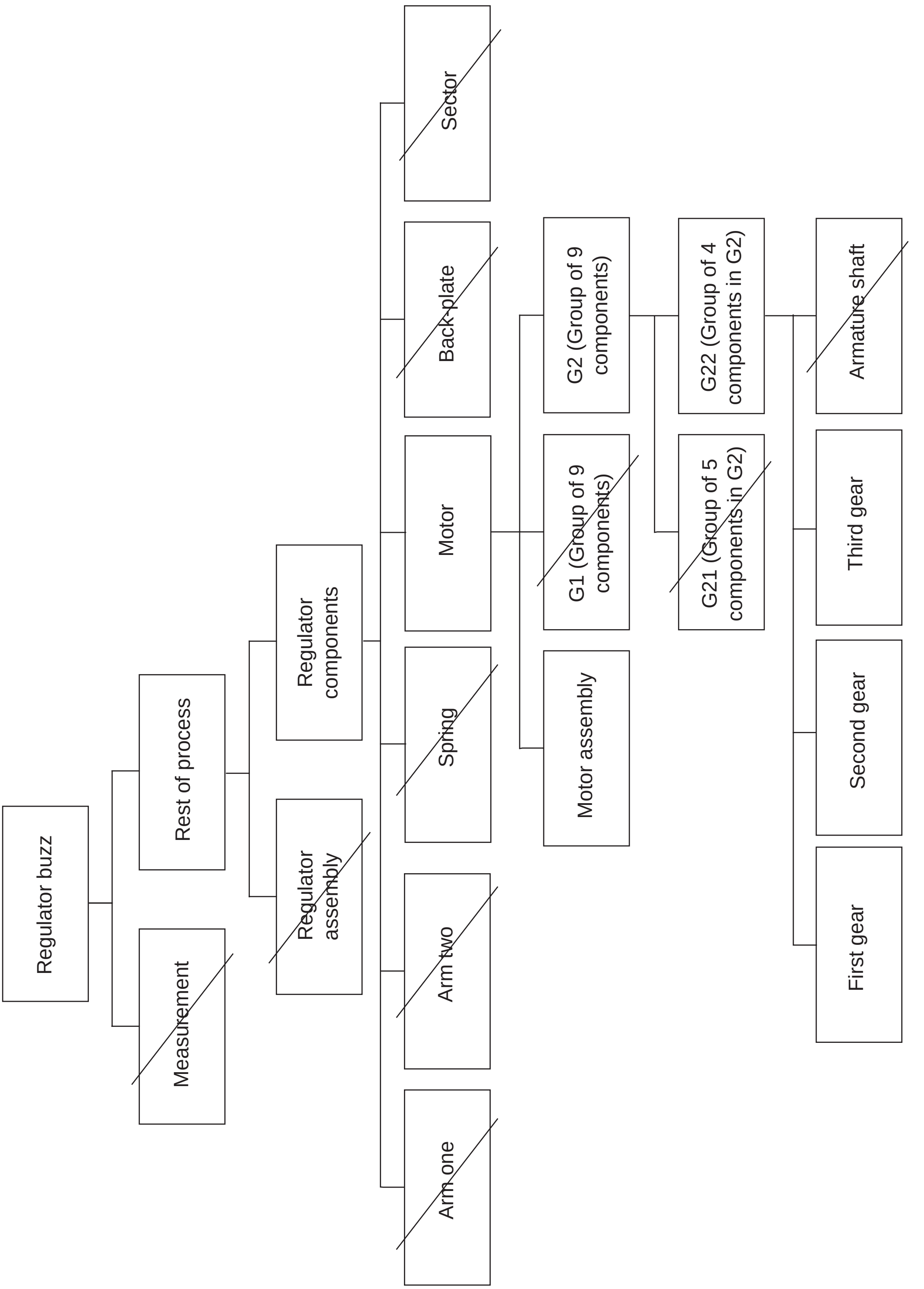


Figure 11.20 Diagnostic tree for power window buzz noise.

The team focused their search for the dominant cause on the second and third gears. They approached the motor supplier with their findings. The supplier provided a new quieter type of motor to solve the problem. Given that this solution does not require an understanding of the cause of motor noise, the team would have been better off talking with the supplier much earlier.

Interpreting the Results

In any component swap investigation where we divide an assembly into two groups of components denoted by G1 and G2, there are a number of possible results when we swap G1 and G2. Consider the following hypothetical example based on the window buzz problem. We always start with the original measurements and the results from the repeated disassembly and reassembly, if they are available. Suppose we have:

	G1	
G2	**Noisy**	**Quiet**
Noisy	7, 6, 7, 6	
Quiet		1, 2, 2, 1

The first numbers in the diagonal cells correspond to the original values of the extreme assemblies that we selected for the investigation. When we repeatedly disassemble and reassemble, we see the variation due to the assembly process. In this case, we eliminate the assembly process as the home of the dominant cause.

We obtain the off-diagonal elements by swapping the components in G1 between the two assemblies. If the performance follows one of the two groups of components, the interpretation of the results is straightforward. For example, we might have the data:

	G1	
G2	**Noisy**	**Quiet**
Noisy	7, 6, 7, 6	1
Quiet	7	1, 2, 2, 1

or

	G1	
G2	**Noisy**	**Quiet**
Noisy	7, 6, 7, 6	7
Quiet	2	1, 2, 2, 1

In the left table, G1 is the home of the dominant cause; in the right table, the dominant cause acts in G2. In other cases the performance does not follow either group. For instance, consider the following four possible results:

	G1	
G2	Noisy	Quiet
Noisy	7, 6, 7, 6	1
Quiet	1	1, 2, 2, 1

	G1	
G2	Noisy	Quiet
Noisy	7, 6, 7, 6	4
Quiet	4	1, 2, 2, 1

	G1	
G2	Noisy	Quiet
Noisy	7, 6, 7, 6	7
Quiet	4	1, 2, 2, 1

	G1	
G2	Noisy	Quiet
Noisy	7, 6, 7, 6	4
Quiet	1	1, 2, 2, 1

In all of these cases the dominant cause involves (at least) one component from both G1 and G2. We need to re-form the groups to determine where the dominant cause lies. We discuss the situation where a dominant cause involves two or more components in more detail later in this section. We also explore the general issue of dealing with a dominant cause involving two (or more) inputs in Chapter 14.

Summary

Suppose we have:

- Two assemblies with opposite and extreme output relative to the full extent of variation
- Assemblies that can be disassembled and reassembled without damage
- Elimination of the assembly family as the possible home of a dominant cause (see Chapter 10)

We search for the component that is the home of the dominant cause with a series of small experiments where we divide the components into two groups and swap one group. We analyze the data by looking at two-way tables. We assume that a dominant cause lives with one or at most a pair of components. If we swap this component or pair, the output will move across most of its full extent of variation.

The component swap procedure has two parts. In the first we assume the dominant cause is in a single component family.

Component Swap Procedure

1. Divide the (remaining) components into two groups.
2. Swap one of the two groups of components, as shown in Figure 11.21, and measure the output for the two new assemblies.
3. Interpret the results. If the performance:
   - Follows one of the two groups, eliminate the other group. Stop when a single component remains; otherwise go back to step 1.
   - Does not follow one of the two groups, the dominant cause lives with a pair of components, one from each of the two groups. Go to the component swap add-on procedure.

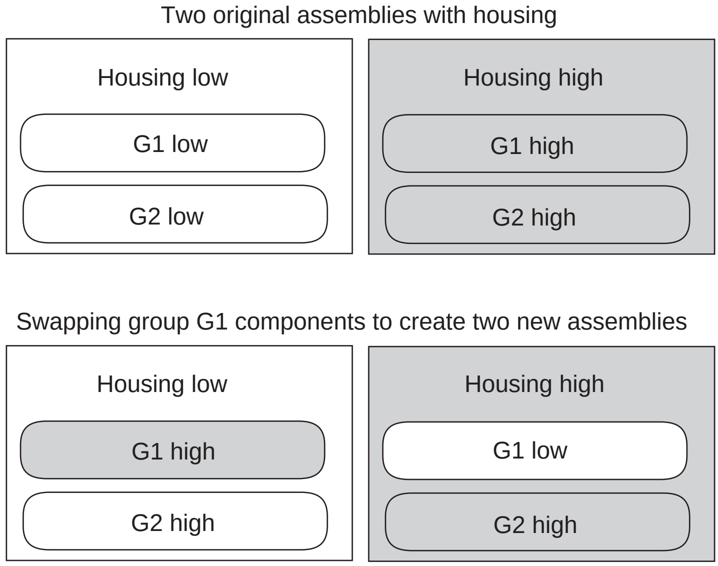


Figure 11.21 Illustration of swapping the group of components labeled G1.

The procedure is more complicated if the dominant cause involves two components.

Component Swap Procedure Add-on

A1. Divide the remaining components into two groups in a new way.

A2. Swap one of the two groups of components, as shown in Figure 11.21, and measure the output for the two new assemblies.

A3. Interpret the results. If the performance:

- Follows one of the two groups, eliminate the other group. Stop when a single pair of components remains; otherwise divide the remaining components into two groups and go back to step A2.
- Does not follow one of the two groups, go back to step A1.

This procedure will fail if there is no dominant cause or if the dominant cause involves three or more components.

Comments

A key requirement for component swap plans is that the product can be disassembled and reassembled without damaging the parts. To avoid study error, the reassembly process should match the assembly process in normal production as closely as possible. Once we have ruled out the assembly process, we carry out the component swap investigation offline to avoid interference with regular production.

By exploiting leverage, we use only two assemblies chosen to reflect the full extent of variation. There is a risk we may select assemblies that are extreme due to a different failure mode, and hence with a different dominant cause, than that of the problem we are trying to address. To alleviate this risk we recommend confirming the conclusion with an experiment that uses several extreme assemblies. For instance, to confirm the conclusion in the headrest failure example, we can put good springs in five seat assemblies that originally required less than 35 Newtons force to move through their full range of motion. If the new springs increase the force above 35 Newtons, we confirm the spring as the dominant cause.

Since all components that make up the housing have been eliminated as a possible home of the dominant cause, we can use a single housing for further disassembly-reassembly and component swap investigations. We employed this idea in the power window buzz noise example. Note that at the start of an investigation there is often no housing because all components are suspect.

In the recommended component swap procedure, we assume the assembly family has been eliminated using an investigation where we disassembled and reassembled down to the individual components. One advantage of this plan is that we can use the production assembly process. Alternatively, suppose we divide the assemblies into a number of subassemblies (groups of components). In that case, we start by disassembling and reassembling down to the subassembly level. Then, if this assembly process is eliminated, we apply the component swap procedure to the subassemblies. A complication arises if the proposed component swap procedure indicates the dominant cause is an interaction between two groups of subassemblies. This interaction could be due to either components or the assembly process for the subassemblies.

To illustrate, consider the window buzz noise example. The team found the dominant cause acted in the motor, which had 18 individual components that could be grouped into two subassemblies. Suppose the team started by disassembling and reassembling each extreme motor into the two subassemblies and that this part of the assembly family was eliminated. Next, the team would swap one of subassemblies between the two motors. Suppose the dominant cause was found to act within one of the subassemblies. Then the next step would be to check the assembly family for that subassembly before proceeding to swap components from the identified subassembly.

There are many other component swapping plans.[8] For example, we can use more than two groups of component groups at each stage.[9] We strongly recommend dividing the remaining components into only two groups at each stage to keep the procedure simple and to maximize the number of components eliminated with each swap.

Key Points

- We use a multivari investigation to compare time and location families. We design a systematic sampling protocol to isolate the contribution of each family to the baseline variation.
- We use a variation transmission investigation to eliminate process step families. We trace parts through the process and measure the output characteristic after each process step.
- We use component swap to eliminate families related to an assembly process and the corresponding components. We repeatedly disassemble, reassemble, and swap components in an organized way.

Endnotes (see the Chapter 11 Supplement on the CD/ROM)

1. If we cannot isolate a dominant family or families from the multivari charts, we can use analysis of variance (ANOVA) to quantify the contribution of each family to the overall variation.
2. In multivari investigations, we expect families such as part-to-part to show haphazard rather than systematic variation. We need to plot the data carefully to detect the effect of such families. We provide more detail on assessing this type of variation in the supplement.
3. See note 2.
4. See note 2.
5. See note 1.
6. See note 1.
7. We can use regression analysis to separate the variation transmitted and added at each operation if we can measure the output characteristic before and after the operation. We elaborate in the supplement.
8. We can conduct investigations to compare component families in different ways. We discuss a specialized tool first introduced by Dorian Shainin and some other related methods in the supplement.
9. We propose component swap investigations that involve dividing all remaining components into only two groups. In the supplement, we consider plans that divide components into three groups.

Exercises are included on the accompanying CD-ROM

12

Investigations Based on Single Causes

I walk slowly, but I never walk backward.

—Abraham Lincoln, 1809–1865

In this chapter, we look at investigations to search for a dominant cause based on single causes, that is, individual varying inputs. These plans are not particularly useful early in the search because they do not result in the elimination of families with a large number of causes. However, we can use these investigations for available data or when the family of remaining suspects is small. With these plans, we measure the output and selected suspects for a number of parts.

Large variation in a measurement system for an input can mask a dominant cause. Ideally, we would assess the measurement systems used to determine these inputs—see Chapter 7 for details. This is often not done, either because of prior experience with the measurement system or due to lack of resources.

In this chapter, we distinguish between plans for continuous and binary outputs because they lead to different analysis tools.

12.1 GROUP COMPARISON: COMPARING PARTS WITH BINARY OUTPUT

We use a *group comparison* to examine the effects of individual causes when the output is binary. To start, we select a number of parts with each of the two output values. Then we measure as many inputs as possible as suspects. Recall that a suspect is a cause that has not been ruled out in the search for a dominant cause. The values for a dominant cause will be substantially different for the two groups of parts. We look at two examples.

Engine Block Leaks

In the engine block leak example discussed in Chapter 1, the project was divided into three separate problems to address different failure modes. The blocks that leaked from the left rear intake wall had a visible defect. Close inspection with a microscope revealed that the defect was a sand inclusion. The team also observed that wall thickness varied from block to block.

The team planned a new investigation. Whenever they found an intake wall leaker, they set aside a nonleaking block cast in the same hour. They collected 100 blocks for each group. Then, for each of the sampled blocks, they measured thickness in inches at six locations in the left rear intake wall. The data are given in the file *engine block leaks comparison*.

To analyze the data, we construct box plots of wall thickness at each location for leakers and nonleakers. We show the results for two locations in Figure 12.1.

The right panel of Figure 12.1 shows a difference in average wall thickness between leakers and nonleakers at location four. There was little difference for the other locations as illustrated in the left panel of Figure 12.1. The team concluded that wall thickness at location 4 was a dominant cause of rear intake wall leaks.

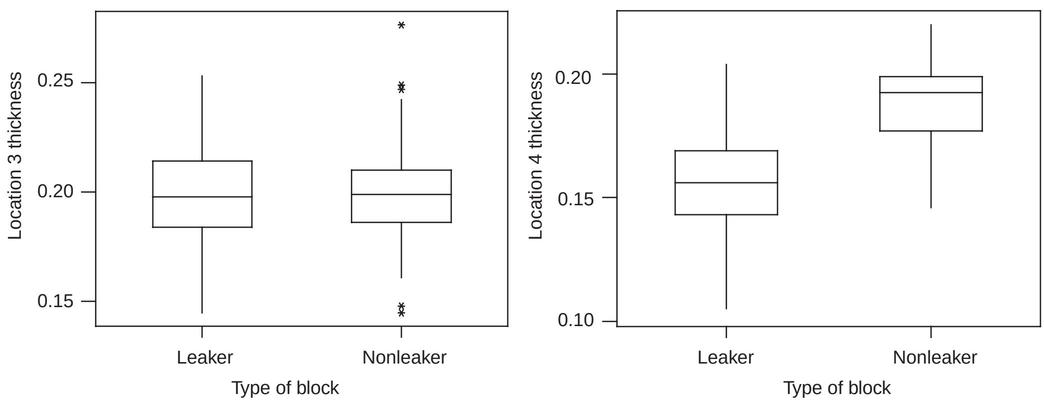


Figure 12.1 Box plots of wall thickness by block type at locations 3 and 4.

Window Leaks

Based on the results of customer surveys, the management of a truck assembly plant identified rear window water leaks as a substantial customer concern and assigned a team to address the problem. In the plant, 10 trucks per day were leak-tested in a special chamber that simulated extreme conditions. Using one month's data to establish a baseline, the team found that the rear window leak rate was 8% in the aggravated test. The team assumed that if they reduced this rate there would be substantial reduction in customer complaints. Using Pareto analysis by location, the team identified the upper ditch as the source of about 50% of the leaks. They set the problem goal to eliminate upper ditch rearindow leaks.

The team showed that the measurement system was acceptable for classifying trucks as either leaking or not and for determining the location of the leak. However, the system could not consistently measure the severity of the leak. For this reason, the team used the measurement system only to classify a truck as a leaker or a nonleaker of the upper ditch rear window.

The team decided on a group comparison with eight leakers and eight nonleakers. They selected eight trucks that had failed the leak test with an upper ditch leak. Obtaining trucks that did not leak was more difficult since there was great pressure to immediately ship any good truck. The team found eight nonleakers from trucks that had been set aside for other problems.

They measured nine input characteristics thought to be related to water leaks. The data are given in the file *window leaks comparison*. The team plotted the data by group for each of the nine suspects. We show the result for primary seal fit in the left panel of Figure 12.2. There was no clear separation between leakers and nonleakers for any input. They also created scatter plots for all pairs of suspects with a different plotting symbol for leakers and nonleakers.[1] The plot for primary seal gap and quality of plastisol application, in the right panel of Figure 12.2, shows that all leaking trucks had both a large primary seal gap and poor plastisol application, while that combination never occurred for nonleaking trucks.

The team concluded that the dominant cause of upper ditch leaks was the combination of a large primary seal gap and poor plastisol application. In addition, they eliminated the other seven inputs from consideration. This conclusion was based on a small number of trucks and required verification.

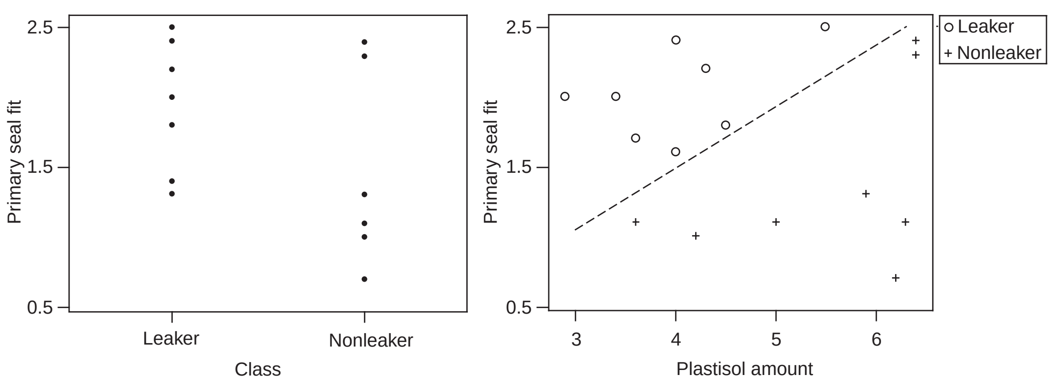


Figure 12.2 Plot of primary seal fit by group (left) and by plastisol amount (right).

Group Comparison Summary

Select as many suspects as possible that can be determined after measuring the output.

Question

In the current process, which, if any, of the suspects is a dominant cause?

Plan

- Select a group of at least six parts for each of the two output values.
- Measure all suspects for each part.

Data

Record the output (group) and input values, one row for each part.

Analysis

- Plot the values of each suspect stratified by group.
- If no clear dominant cause is evident, create scatter plots of all possible pairs of inputs using different plotting symbols for each group.

Conclusion

- If there is clear separation of the suspect values between the groups, the suspect is the dominant cause.
- If there is a separating line on a scatter plot that divides the groups, a combination of the two suspects is the dominant cause.

Comments

Group comparison is especially useful if the problem is defined in terms of a rare defect. Low-frequency problems are difficult to solve because it is hard to get information about the dominant cause of the defect. We can collect the two groups of parts over time until we have sufficient numbers to make the comparison.

We sometimes try a group comparison early in the search for a dominant cause where we measure many continuous characteristics on each part in the two groups. If we can find one characteristic that separates the defectives and the good parts, we can reformulate the problem in terms of the continuous characteristic. The reformulated problem may be much easier to solve because we can now see variation in the output on each part.

We can use group comparison if the output is binary or if the output is continuous and we select parts from the extremes to get the full extent of variation. In this instance, we create two groups and ignore the measured values of the output. We need to be careful if we use leverage in this way. One danger is that the extreme parts used in the investigation may be due to different failure modes and thus have different causes. This highlights the importance of focusing the problem so that there is only a single dominant cause. Suppose, in the engine block leaks example, the team had not focused the problems based on the location of the leak. If they had then carried out the group comparison described previously, they would not have seen large differences in wall thickness between the two groups, because the leakers group would likely contain blocks that leak at other locations.

In a group comparison, the groups are formed using the output. As a result, we can only compare inputs that can be measured or determined after the part is produced. For instance, in the window leaks example, it was not possible to compare process inputs such as window installer or machine settings because these cannot be determined after a truck is leak tested.

We require at least six to eight parts per group to avoid falsely identifying a dominant cause. We examine some alternate plans and analyses in the supplement to this chapter.[2]

12.2 INVESTIGATING THE RELATIONSHIP BETWEEN INPUTS AND A CONTINUOUS OUTPUT

For an *input/output relationship investigation*, we measure the output and several inputs on each part. We present three examples.

Crossbar Dimension

There was excess variation in a key crossbar dimension of an injection-molded part. In the baseline investigation, the team estimated the standard deviation and the full extent of variation of the dimension to be 0.46 and –0.3 to 2.0 thousandths of an inch. The problem goal was to reduce the standard deviation to less than 0.25.

The team showed that the measurement system was highly capable. They decided to search for a dominant cause. They conducted a multivari investigation where five consecutive parts from the single mold were sampled every 30 minutes for four hours. See the exercises for Chapter 10. The team found that the time-to-time family contained the dominant cause.

Next they planned an input/output relationship investigation. Forty shots of the process were selected haphazardly over a two-day period. For each shot, the team measured the crossbar dimension of the part and recorded five inputs: die temperature, nozzle temperature, barrel temperature, hydraulic pressure, and cavity pressure. All these suspects were thought to exhibit time-to-time variation.

The data are given in the file *crossbar dimension input-output*. The first step in the analysis is to check that the output varied over the full extent of variation as in the baseline. We see from the summary that this is the case.

```
Variable        N       Mean      Median      TrMean       StDev     SE Mean
dimension      40     0.8340      0.7618      0.8353      0.5467      0.0864

Variable       Minimum    Maximum         Q1          Q3
dimension      -0.1611     1.9154     0.4282      1.3649
```

The second step is to fit a regression model[3] (Montgomery et al., 2001) that includes all of the inputs simultaneously. See Appendix E for the MINITAB directions. The residual standard deviation is 0.2515, which is substantially less than the baseline standard deviation. The residual standard deviation is an estimate of the variation in crossbar dimension if we could hold all of the inputs in the regression model fixed. The small value of residual standard deviation indicates that one or more of the inputs is a dominant cause of the variation.

In the third step, we look at the scatter plots of crossbar dimension versus each of the inputs.[4] Two of the plots are shown in Figure 12.3.

There is a strong relationship between barrel temperature and crossbar dimension. If we could hold barrel temperature fixed at 77°, for example, we see from the plot that crossbar dimension would vary only by about 0.5. There is strong evidence that barrel temperature

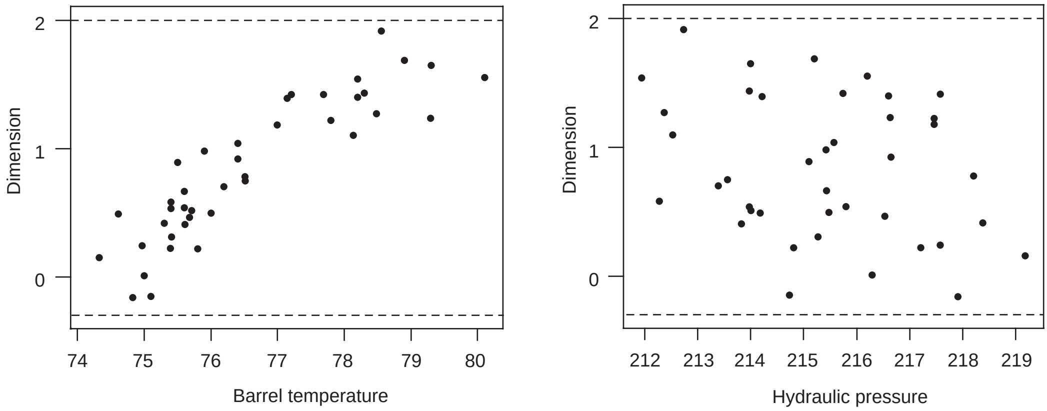


Figure 12.3 Plot of crossbar dimension versus barrel temperature and hydraulic pressure.

is a dominant cause of crossbar dimension variation. In Chapter 13, we describe an experiment to verify this conclusion. There is no strong association of crossbar dimension with the other inputs (as illustrated for hydraulic pressure in the right panel of Figure 12.3) and we eliminate them from further consideration.

If the relationship between the input and output is roughly linear, as is the case here, we can quantify the contribution of the input to the output variation using a regression model. The residual standard deviation represents the remaining (unexplained) variation in the output if we could hold the input fixed. We compare the residual standard deviation to the baseline to assess the contribution of the input.

Using MINITAB, we fit a regression model for crossbar dimension as a function of barrel temperature. The results of the analysis are:

```
The regression equation is
dimension = -23.9 + 0.323 barrel temp

Predictor        Coef     SE Coef          T        P
Constant      -23.898       2.075     -11.52    0.000
barrel temp   0.32293     0.02709      11.92    0.000

S = 0.2544      R-Sq = 78.9%     R-Sq(adj) = 78.3%

Analysis of Variance

Source           DF          SS          MS        F        P
Regression        1      9.1979      9.1979   142.13    0.000
Residual Error   38      2.4591      0.0647
Total            39     11.6570
```

The residual standard deviation is 0.25, much smaller than the baseline value 0.46. If the team could eliminate the effect of barrel temperature they would meet their goal.

Truck Pull

Consider again the truck pull problem introduced in Chapter 1. The team decided to focus on right caster since caster variation had a much larger effect on pull than camber variation. The baseline standard deviation for right caster was 0.24°. Previous investigations ruled out the measurement system. The assembly process was difficult to investigate so the team set that family aside for the moment and concentrated on the component families. A feed-forward controller compensated for variation in the frame geometry. This left the other components of the alignment system, namely the knuckle and upper and lower control arms, as possible homes of the dominant cause. The search for the cause is illustrated in the diagnostic tree in Figure 12.4.

At this point the team had several choices. They could have tried assessing the assembly and component families, as discussed in Chapter 11, but they could not use the production assembly process. Moreover, at the time of the investigation, there was a proposal to bar-code the control arms with dimensional data that could be fed into the feedforward controller already in use to compensate for the frame geometry (see Chapter 16). With this proposed process change, all control arms would be measured and bar-coded at the supplier's plant. Before implementing such an expensive proposal, the team wanted assurance that it would be worthwhile. If the control arm inputs were not dominant causes, there would be little reduction in pull variation from this costly change.

The team decided to explore specific characteristics of the control arms and knuckle. Based on the proposal for the feedforward control scheme, they selected dimensional characteristics of the components, one for each control arm and two for the knuckle, thought to affect caster.

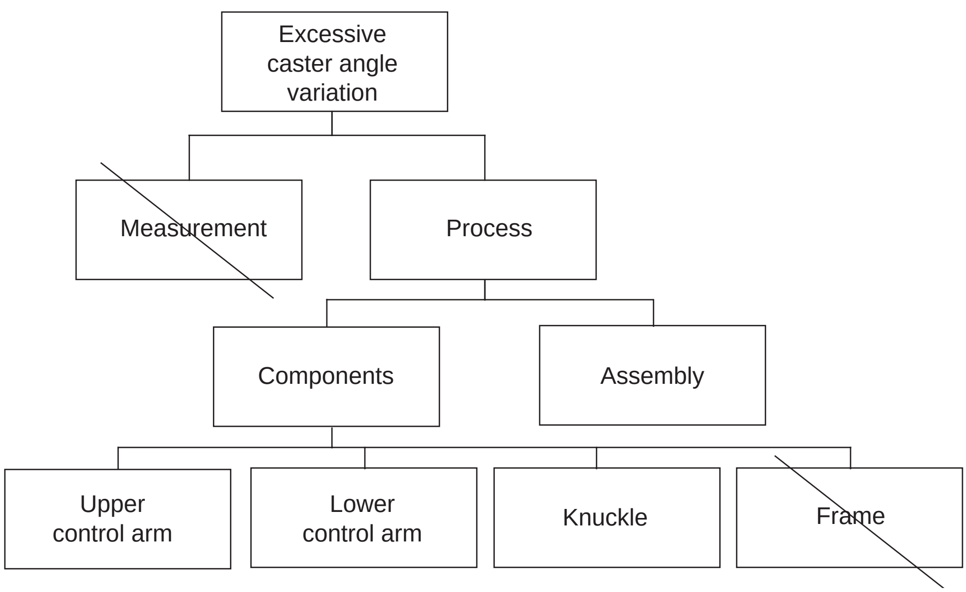


Figure 12.4 Diagnostic tree for search for the cause of excessive caster variation.

For the investigation, the team selected 30 sets of components (two control arms and a knuckle) from regular production over six days. From the baseline investigation, the team expected to see the full extent of variation in caster angle with this sampling scheme. They measured the four inputs on each set of components and the right caster angle on the assembled truck. The data are given in the file *truck pull input-output*. The standard deviation of the caster angle in the investigation is 0.22, somewhat smaller than the baseline standard deviation.

We fit a regression model to describe the relationship between caster angle and the four inputs. The results from MINITAB are:

```
The regression equation is
caster = 68.5 + 1.57 lower ball + 1.48 upper ball + 0.242 U reading
            -0.184 L reading

Predictor        Coef     SE Coef         T        P
Constant       68.546       9.922      6.91    0.000
lower ball     1.5710      0.6905      2.28    0.033
upper ball     1.4779      0.6839      2.16    0.042
U reading     0.24187     0.02282     10.60    0.000
L reading    -0.18446     0.02844     -6.49    0.000

S = 0.1620      R-Sq = 87.8%     R-Sq(adj) = 85.6%
```

The residual sum of squares is $s = 0.162$, a moderate reduction from the baseline variation 0.24. Next we look at the scatter plots of the inputs by caster angle. We cannot see any strong relationships between caster angle and the individual component dimensions in Figure 12.5.

To quantify the contribution of each component dimension, we fit regression models with one input at a time. We rank the inputs based on the residual standard deviation in Table 12.1.

Table 12.1 Component dimensions ranked by residual standard deviation.

Input	Component	Residual standard deviation
U reading	Upper control arm	0.19
L reading	Lower control arm	0.21
Upper ball	Knuckle	0.21
Lower ball	Knuckle	0.22

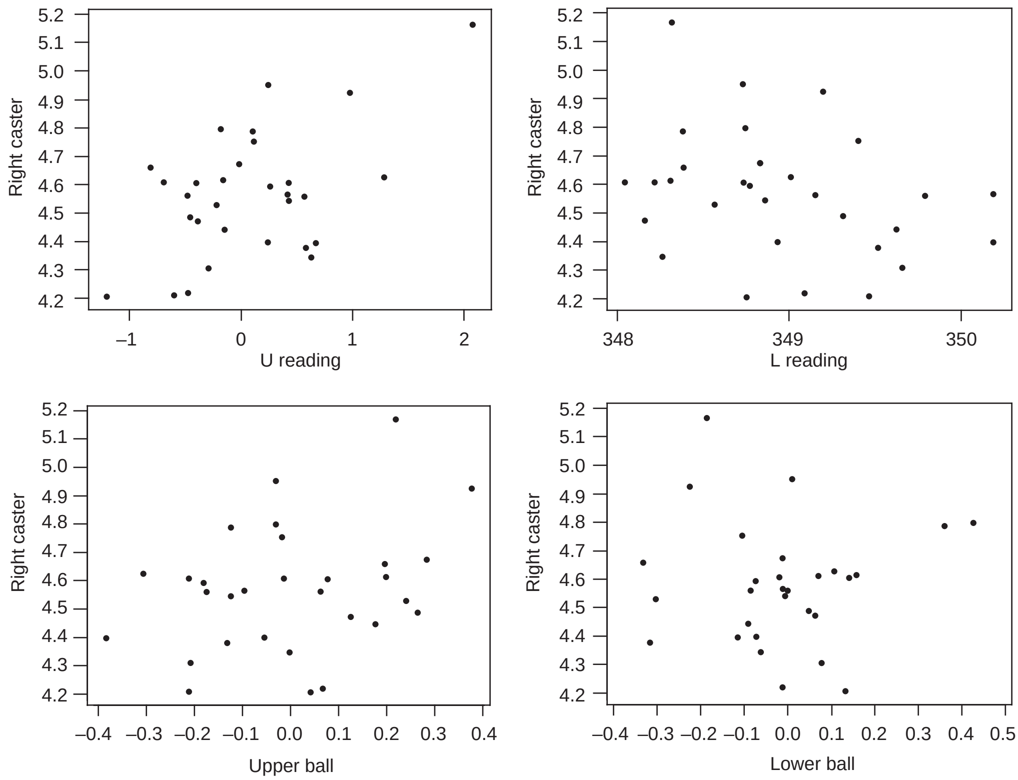


Figure 12.5 Scatter plots of right caster versus the component characteristics.

None of the component dimensions is a dominant cause when considered singly. When we fit regression models with pairs of inputs, we see in Table 12.2 that the two control arm dimensions together produce the smallest residual standard deviation.

We conclude that if we could completely remove the effects of the two control arms by using the feedforward controller, we would reduce the caster variation by about 20%. This reduction could be achieved only if the feedforward controller worked perfectly. Also, the sample size in this investigation is small and there is considerable uncertainty in the estimates of the standard deviations.

The team used these results to argue that the proposed bar-coding of the control arms would not be cost-effective. They decided instead to investigate the feasibility of feedback control on caster angle (see Chapter 17).

Table 12.2 Pair wise component dimensions ranked by residual standard deviation.

Input pairs	Residual standard deviation
U reading, L reading	0.174
U reading, Upper ball	0.177
U reading, Lower ball	0.195
L reading, Upper ball	0.212

Manifold Sand Scrap

In the production of cast-iron exhaust manifolds, a foundry attributed many defects to the sand system used to create the molds. Historically, the sand-related defect rate was 2% with substantial variation shift to shift. Management assigned a project team to reduce sand-related scrap.

The team used a run chart of the scrap rate by shift as a baseline. This chart was produced daily as part of the management information system. They did not formally investigate the measurement system used to determine whether a manifold had a sand-related defect. They believed from past experience that this system was reliable. They decided to search for a dominant cause of the defects.

Many characteristics of the molding sand such as temperature, compactness, permeability, moisture level, green strength, and percent friability were routinely measured during production. The team decided to use the available data to determine if any of these individual sand characteristics was a dominant cause. They chose not to use broad families of causes as is usually done early in the search because the data were already available and the cost was low.

It was not easy to get the data into a usable form. The sand characteristics were not measured for each casting. There were substantial and varying time lags between the measurement of a sand characteristic and the use of the sand to make a mold. For each casting, the time of casting was known only up to the nearest hour. Because of these traceability difficulties, the team used the hourly scrap rate as the output and the average of the sand characteristics over the hour as the inputs.

The team selected 91 hours of production and made the linkages between hourly scrap rate and average sand characteristics. In each hour, the plant produced between 30 and 180 castings. The data are given in the file *manifold sand scrap input-output*. The average and standard deviation of the hourly scrap rate were 0.017 and 0.023. The average rate 1.7% matches the baseline well.

To analyze the data, the team ignored the changes in volume and focused solely on the proportion of sand scrap in each hour. Fitting a regression model to the hourly scrap rates with all of the sand characteristics, the residual standard deviation was 0.022, a very small reduction from the baseline value 0.023. None of the sand characteristics was a dominant cause. This is confirmed by looking at scatter plots of the sand scrap proportion versus the inputs. Some of the plots are given in Figure 12.6.

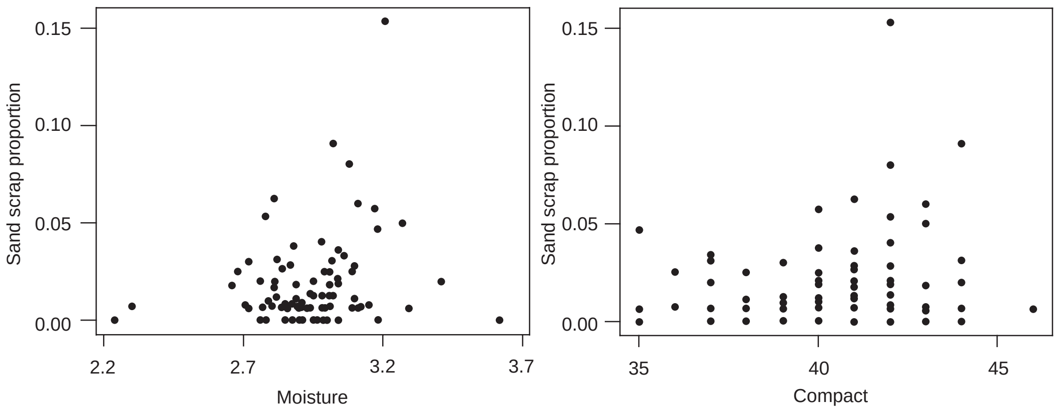


Figure 12.6 Scatter plot of sand scrap proportion versus moisture and compactness.

The team noticed in the plot of hourly scrap rate versus temperature, shown in Figure 12.7, that there was a nonlinear relationship. Using MINITAB, they fit a quadratic regression model that included the square of the temperature. Part of the results are shown as follows:

```
The regression equation is
sand scrap proportion = 6.16 – 0.121 temperature +0.000596 temp. squared

Predictor             Coef     SE Coef          T        P
Constant            6.1647      0.9629       6.40    0.000
temperature       -0.12116     0.01911      -6.34    0.000
temp. squared   0.00059610  0.00009476       6.29    0.000

S = 0.01922     R-Sq = 33.0%     R-Sq(adj) = 31.4%
```

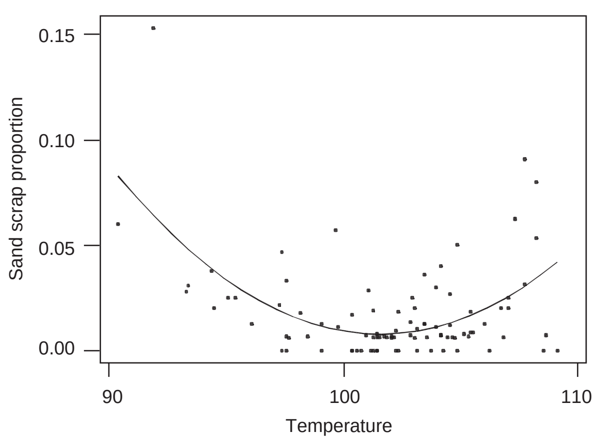


Figure 12.7 Sand scrap proportion versus sand temperature with quadratic fit.

While the observed pattern makes physical sense, temperature is not a dominant cause. Holding sand temperature fixed (which would be very expensive) would not reduce the sand scrap proportion substantially.

In the end the team was forced to conclude the investigation was a failure. They were not able to find a dominant cause, and they could not eliminate any of the suspects because the linkage between sand characteristics and the defect rate was poorly established.

This example demonstrates the risk of using an investigation based on individual causes early in the search for a dominant cause. Even if the team could have eliminated the measured sand characteristics as dominant causes, they would have made little progress, as there were a large number of remaining suspects.

Input/Output Relationship Investigation Summary

Select as many suspects as possible that can be traced to the output.

Question

For the current process, which, if any, of the suspects is a dominant cause?

Plan

- Select a time frame in which we expect to see the full extent of variation in the output.
- Select 30 or more parts spread across the time frame.
- For each part, measure the output and all suspects.

Data

Record the output and suspect values, one row for each part.

Analysis

- Use regression to model the output as a function of all suspects simultaneously. For categorical suspects, use indicator variables.[5]
- Plot the output versus each one of the suspects. For continuous suspects, fit the corresponding simple regression model to quantify any strong linear relationships. For categorical suspects, use one-way ANOVA.

Conclusion

- If the residual standard deviation in the first regression model is much less than the baseline standard deviation, one of more of the suspects is a dominant cause.
- If there is a dominant cause, identify it from the scatter plots.

Comments

In an input/output relationship investigation, we can select parts in one of three ways:

- Using some time- and/or location-based sampling scheme
- Based on the values of the inputs
- Based on the output values

In all cases, the goal is to get the full extent of variation so that we know the dominant cause has acted.

With the first option, we need to have a large enough sample spread across appropriate times and locations to meet the goal. We can use the results of the baseline investigation to help to define the sampling scheme. This option is the most common and was used in the examples discussed in this section.

The second option requires more effort. For example, in the truck pull example, the team had information about the historical variation in the four inputs. They could have selected components with values for these dimensions at both ends of their historical range. This would have required extra measurement to find the appropriate components. Using extreme values for the inputs, we employ leverage and can use a smaller sample size. If we do not see the full extent of output variation with this plan, none of the selected inputs is a dominant cause.

The final option corresponds closely to the group comparison plan. For example, in the window leaks problem, had the team been able to measure the severity of the leak, they could have fit a regression model to the data. By choosing parts with extreme output values, we are sure to see the full extent of variation. However, we may not be able to determine the values of many suspects due to lack of traceability.

We need to be careful interpreting the MINITAB regression results. In the truck pull example, each of the four inputs is *statistically significant* because the *p*-value is small (less than 5%). This does not mean that any of the input is a dominant, or even large, cause, as we saw in the example. The same difficulty occurs in group comparisons when the input averages for the two groups are statistically *significantly different*, and yet there is no evidence that the input is a dominant cause. Hypothesis tests are not a useful tool for identifying a dominant cause. See further discussion of this issue in the Chapter 10 supplement.

Regression models can accommodate binary outputs, discrete inputs, quadratic terms, interactions between inputs, and so on.[6] We can transform or combine the input values. We need to be careful using these models because we can be misled by nonlinear relationships, outliers, and influential observations.

Key Points

- Group comparison and input/output investigations assess single causes and are most useful in the latter stages of the search for a dominant cause.
- With a binary output, we can compare inputs measured on parts using a group comparison. We use plots of the individual values or box plots to look for large differences in the inputs for the two groups of parts defined by the output values.
- With continuous output, we use regression models and scatter plots to show the input/output relationships and isolate dominant causes. We calculate the residual standard deviation from a simple regression fit to assess whether a particular suspect is a dominant cause.

Endnotes (see the Chapter 12 Supplement on the CD-ROM)

1. We show in the supplement that for a group comparison or an input/output investigation, there is a quick way to make all the desired scatter plots in MINITAB.
2. In the chapter supplement we compare group comparison with paired comparison as suggested in Bhote and Bhote (2000).
3. Regression analysis is a flexible analysis tool that has been extensively studied in the statistical literature. In the supplement, we explore some of the useful extensions in more detail and give references to further work.
4. See note 1.
5. See note 3.
6. See note 3.

Exercises are included on the accompanying CD-ROM

Chapter 13
Verifying a Dominant Cause

Approach each new problem not with a view to finding what you hope will be there, but to get the truth, the realities that must be grappled with. You may not like what you find. In that case you are entitled to try to change it. But do not deceive yourself as to what you do find to be the facts of the situation.

—Bernard M. Baruch, 1870–1965

We recommended the method of elimination and a series of simple observational investigations to isolate a dominant cause. Before proceeding with the next stages of the Statistical Engineering algorithm, we want to be sure that the suspected cause, here called a *suspect,* is dominant. We call this *verification.*

In many applications of the algorithm, we have sufficient evidence from the search to be sure we have found a dominant cause and we require no further verification. For example, in the fascia cratering problem discussed in Chapter 11, the team found that craters occurred only on every fifth and tenth fascia taken from the mold. They concluded that the dominant cause was the application of mold release. They did not verify this conclusion because no other cause matched the observed pattern of craters. Similarly, in the V6 piston diameter example (Chapter 11), the team concluded that a dominant cause of diameter variation at the final gage was the diameter after Operation 270. The variation transmission investigation showed that pistons with large (small) diameters after Operation 270 were large (small) at the final gage. The team could explain the observed pattern in only one way and decided not to verify their conclusion.

Why do we need to verify? In the search, we might have inadvertently ruled out a family that contains the dominant cause. More commonly, we may have selected the suspect from the remaining family of causes using our best judgment. If we are wrong, there may be other causes in the family that are dominant. Consider the problem of excess crossbar dimension variation discussed in Chapter 12. There, the team concluded that barrel temperature was the dominant cause based on the results of an input/output relationship investigation. They decided that verification was necessary because it was possible that the actual dominant cause was another (unidentified) cause in the same family (time-to-time) as barrel temperature.

To verify that a suspect is a dominant cause, we use an experimental plan, often called a *designed experiment,* where the value of the suspect is deliberately manipulated. Good references on experimental plans include Box et al. (1978), Ryan (1989), Wheeler (1990), and Montgomery (2001).

Experimental plans are also important tools, as we will see in later chapters, for helping to assess the feasibility and determine how to implement several of the variation reduction approaches. In this chapter, we introduce the language and principles of experiments. We start with plans to verify a single suspect dominant cause. We then introduce more complex plans used to isolate a dominant cause from a short list of suspects.

13.1 VERIFYING A SINGLE SUSPECT DOMINANT CAUSE

To verify a single suspect, we plan a simple experiment where only the suspect is deliberately varied. For illustration, we use the engine oil consumption problem discussed in Chapter 10. Using stratification, the team had determined that a dominant cause acted in the plant-to-plant family. Based on knowledge of the engine, one of the few plant-to-plant differences that could explain the variation in oil consumption was a different supplier of valve lifters. The valve lifters were suspects because changing the valve lifters on a returned engine eliminated the oil consumption. Further application of the method of elimination pointed to a clearance dimension in the lifter as a primary suspect. The team decided to use an experimental plan to verify that the valve lifter clearance dimension was the dominant cause of oil consumption.

We introduce some terminology used in designed experiments. We plan to deliberately change one or more inputs. We call the different values the *levels* of the inputs. In the oil consumption example, clearance was the single suspect. The team decided to use two levels for clearance, corresponding to the high and low ends of its known historical range. We call these the *low* and *high level* of the suspect input.

The suspect is a dominant cause if the output moves across its full extent of variation when we change the level of the suspect from low to high.

An *experimental run,* usually shortened to a *run,* corresponds to setting the level of the suspect, running the process and measuring the output characteristic. In the example, a run corresponds to installing a set of lifters with low (or high) clearance into an engine, putting the engine through the accelerated dynamometer test, and measuring the oil consumption. We carry out one or more runs for each level of the suspect. If there is more than one run for a given level, we call this *replication.* To achieve *balance,* we use an equal number of runs for each level.

We normally require a minimum of three or four *replicates* at each level of the suspect. More replicates will give more reliable conclusions but increase the cost of the experiment. Within a run, the process may produce a number of parts, some of which are measured. We call these parts *repeats.* We recommend the same number of repeats per run to preserve the balance. The connection between replicates, runs, and repeats is illustrated in Figure 13.1.[1]

In the example, the team found three sets of lifters (each engine requires a set of eight lifters) with low clearance and three sets with high clearance by measuring clearance on incoming lifters from the poor supplier. In turn, they installed each set in an engine and measured the oil consumption. The experiment had six runs with no repeats. There were three replicates for each of the two levels of clearance.

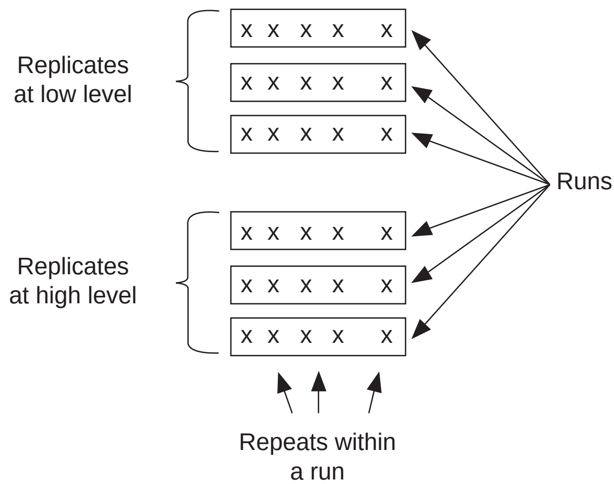


Figure 13.1 Runs, replicates, and repeats for an experiment with a single suspect at two levels.

To protect against some other input changing systematically as we change the suspect, we randomize the order of the runs, if feasible.[2] This use of randomization is one of the main differences between observational and experimental investigations. We can randomize the order of replicates but not repeats. For example, we give the random order and the oil consumption for each run in Table 13.1.

Table 13.1 Valve lifter clearance experiment and results.

Level	Average lifter clearance	Run order	Oil consumption (grams per hour)
Low	5.0	4	23
Low	7.0	3	24
Low	8.5	6	29
High	21.0	1	76
High	24.0	2	113
High	25.0	5	120

To analyze the results, we rely on tables of averages and graphical summaries such as box plots or scatter plots. If the suspect is a dominant cause of variation, the output characteristic should vary over most of its full extent of variation when the value of the suspect changes from its high to low level. In Figure 13.2, we see that changing the valve lifter clearance has a large consistent effect on the oil consumption. The team concluded that valve lifter clearance was a dominant cause of the variation in oil consumption and that low clearance values led to less oil consumption.

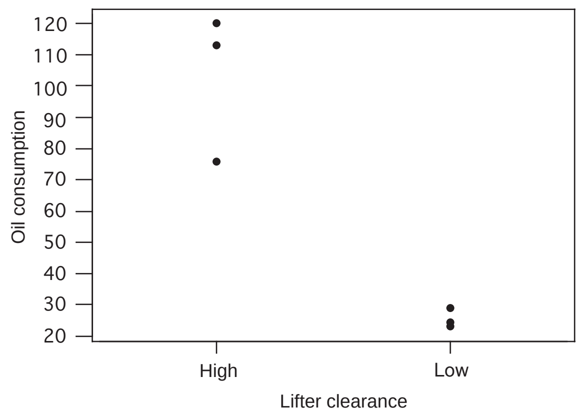


Figure 13.2 Oil consumption by lifter clearance level.

There were several limitations to this conclusion. First, the team did not have a good idea of the baseline in terms of the accelerated dynamometer test. They had measured only one engine from the field failures. Repeated testing of this engine in the dynamometer gave an average oil consumption of 117 grams per hour. The team was reassured because the oil consumption at the high clearance level was close to this average. Second, the verification experiment used only a single engine. The clear results for this engine may not carry over to others. Finally, the team did not have a direct connection between oil consumption in the field and in the result of the accelerated test on the dynamometer. In spite of these limitations, they proceeded, assuming that lifter clearance was the dominant cause.

Crossbar Dimension

In the crossbar dimension example, discussed in Chapter 12, the team identified barrel temperature as the dominant cause. For the verification experiment, the team chose the low level for barrel temperature as 75° and the high level as 80°. This covered close to the full range of barrel temperatures seen in the earlier investigation.

Barrel temperature could be controlled for the experiment and changed in a few minutes. As a result, the verification experiment used only two runs. The barrel temperature was set, 25 parts were made to ensure the temperature had stabilized, and the next 10 parts were selected and measured. There are two runs with 10 repeats per run and no replication.

The data from the experiment are given in the file *crossbar dimension verification* and are presented in Figure 13.3. Barrel temperature has a large effect on crossbar dimension relative to the full extent of variation (given by the dashed lines). The team concluded that barrel temperature was the dominant cause of crossbar dimension variation.

The lack of randomization was not an important limitation here since previous investigations had shown that the dominant cause acted in the time-to-time family. Over the short time frame of the verification experiment, it is unlikely that we will see the full range of variation in the output unless barrel temperature is a dominant cause. There is insufficient time for other causes in the time-to-time family to change substantially.

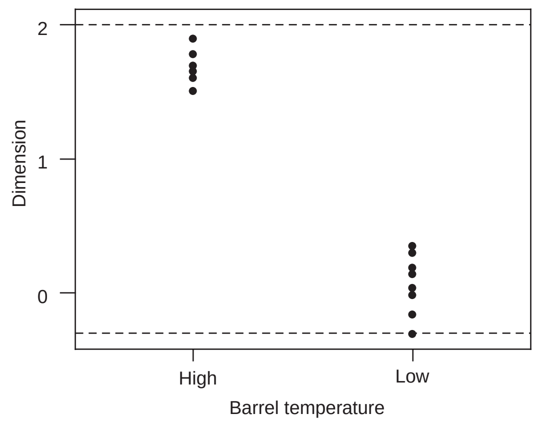


Figure 13.3 Crossbar dimension verification experiment results (dashed lines show the full extent of diameter variation).

The lack of replication may have serious consequences. If the dominant cause is an interaction between barrel temperature and some other cause in the time-to-time family, we may see little effect on crossbar dimension when we change the barrel temperature, depending on the unknown level of the second cause. To remove this uncertainty, we need to take the expensive step of replicating the whole experiment over several time periods, allowing the second cause time to change.

13.2 ISSUES WITH SINGLE SUSPECT VERIFICATION EXPERIMENTS

There are many issues in the planning and analysis of an experiment to verify a single suspect as a dominant cause.

Do We Need to Verify the Suspect?

In deciding if verification is necessary, the team should consider:

- The risk that the suspect is not a dominant cause
- The cost and difficulty of conducting the verification experiment

To assess the risk, the team should think about how they identified the suspect and the size of the remaining family. The risk is high if the team has selected the suspect from a large family of remaining possibilities based on uncertain process knowledge.

The cost of verification may be high, especially if the levels of the suspect are difficult to control. In the engine block porosity problem discussed in Chapter 10, the team noticed that severe porosity occurred when there was downtime in the operation. The team compared the porosity levels before and after breaks and identified two suspects, iron-pouring temperature and the addition of ladle wash. During work stoppages, iron that remained in the six pouring ladles cooled off since there was no external heat source. At the start of the break, ladle wash was added to the ladles to protect the refractory (surface). Since the wash was water-based, the team suspected it was a source of porosity when the process was restarted. The team could not cheaply manipulate the pouring temperature, but they could change and control the level of ladle wash. They added the full amount of wash to three ladles selected at random and half the amount to the other three for two lunch breaks. In each case, they measured the porosity of the first 30 blocks poured (five from each ladle). The data are given in the file *engine block porosity verification*, and the results are discussed in the Chapter 13 exercises. Based on the results and process knowledge, the team eliminated ladle wash as a suspect. They declared pouring temperature a dominant cause, even though they did not verify it.

Choosing the Study Population

Most verification experiments are conducted over a short time using only a few parts. In the oil consumption example, the team used the same engine for all experimental runs. They assumed that the conclusions about valve lifter clearance for this engine would apply more broadly. The team could have strengthened the conclusion by doing the same six-run experiment with three different engines. This change to the plan would have increased the cost and complexity of the experiment.

The more that we know about the nature of the output variation, the easier it is to select a study population. For example, if we know the time-to-time family contains the dominant cause, we can conduct the experiment over a short time period in which the process does not vary materially. There is a small chance of another cause in the time-to-time family changing substantially during the experiment. However, since the suspect is also in the time-to-time family, we may have difficulty manipulating its levels within this short time period.

Choosing the Levels of the Suspect

We select the two levels of the suspect for the verification experiment at the low and high end of its range of values in the regular process. If the suspect is a dominant cause, then changing from the low to high level will produce the full extent of output variation. To determine the levels, we have to know the range of variation of the suspect cause. To acquire this knowledge, we may have to carry out a small investigation on the suspect.

There is some risk in making the levels of the suspect too extreme. First we may induce a different failure mode into the process. Second, we may fool ourselves in concluding that the suspect is a dominant cause because the very extreme levels, which rarely occur in the regular process, may induce the full extent of variation in the output.

For a verification experiment, we strongly recommend using only two levels per suspect. Extra levels give little additional information about whether or not the suspect is dominant and increase the complexity of the experiment.

Randomization and Replication

In some cases we conduct a verification experiment over a short time frame without the protection provided by randomization and replication.[3] In planning the experiment—that is, defining a run, choosing the number of runs, and so on—we need to assess the risk that a dominant cause, other than the suspect, acts within the time of the experiment. The nature of the variation in the output over time is a key piece of information to help assess this risk. If the dominant cause acts over the long term, as in the crossbar dimension example, we can plan a verification experiment over a relatively short time with two runs, one at each level of the suspect. If the dominant cause acts in the part-to-part family, we can use two runs with a moderate number of repeats. That way, if the suspect is not a dominant cause, the true dominant cause (that acts in the part-to-part family) will have time to generate close to the full extent of variation in the runs at both the high and low level of the suspect.

When we do not know the nature of the variation over time, as in the oil consumption example, we need to be careful. We should use replication—that is, several runs for each level of the suspect—and randomize the order in which the runs are conducted. The key question is, "Is there some other unknown cause that might change from run to run in a way that matches the pattern of change of the suspect?" If the answer is yes, then we randomize the order of the runs with at least three replicates per level to reduce the risk.

What If the Output Is Binary?

If the output is binary, we suggest many repeats within each run of the experiment. We can then see if changing the suspect produces a large change in the proportion of defectives within each run. We can assess the importance of the change by comparing the proportions to the baseline. We need runs with many repeats to estimate the proportion of defectives at each level of the suspect. We may have difficulty holding the normally varying suspect constant for a long run.

Is the Cause Dominant?

In the analysis of a verification experiment, we always check that the observed variation in the output characteristic is a substantial proportion of the full extent of variation seen in the baseline. Otherwise the identified suspect is not a dominant cause. We do not recommend a formal hypothesis test to see if changing the suspect produces a statistically significant change in the output. Such a change may be too small to be helpful.

What If the Suspect Is Not a Dominant Cause?

What should we do if the results of the experiment show the suspect is not dominant? First, review the plan and conduct of the experiment in light of the information gathered during the search for the suspect. Some possible questions are:

- What is the family of remaining suspects? Are there other suspects in this family? Are these suspects eliminated by the experiment?
- Could we have missed an interaction with another cause that did not change during the verification experiment?

If the answers to these questions are not helpful, then we have few options. First, review the diagnostic tree. Did we rule out any families or causes during the search without sufficient evidence? If the answer is yes, we may have to reinitiate the search. Second, review the approaches (Chapter 8) and choose a working approach that is not cause-based. If none seem feasible, then we may decide that there is no dominant cause for this problem as formulated and go back to the beginning and reformulate the problem. Finally, we may abandon the problem and accept that Statistical Engineering has failed.

Care in Running an Experiment

When conducting an experiment on a production process, the team requires excellent communication with all parties involved. In practice, things can go wrong, such as:

- Lost runs or repeats within a run
- The level of the suspect may vary within the run or be held at the wrong level over the run
- Unsuspected consequences because of the intervention in the process

More advice on planning and conducting experiments is given in Hahn (1984), Coleman and Montgomery (1993), and Robinson (2000).

13.3 VERIFYING A DOMINANT CAUSE FROM A SHORT LIST OF SUSPECTS

When verifying a dominant cause from a short list of suspects, the experimental plan is more complicated, but many issues are the same as in Section 13.2. Since we plan to investigate all of the suspects simultaneously, we call a particular combination of the input levels used in a run a *treatment*. When the treatments are combinations of the levels of two or more inputs, we call the plan a *factorial experiment.* In other contexts, the inputs are called *factors,* which explains this name. In the plan for the experiment, we must:

- Choose the high and low levels for each suspect.
- Determine the treatments.
- Define a run and determine the number of repeats.
- Decide how much replication is necessary.
- Decide whether randomization of the run order is worth the cost.

We use two levels for each suspect, selected near the extremes of the normal range of variation of the input. If we have three suspects each with two levels, there are eight possible treatments, as shown in Table 13.2. If we have two suspects, there are four possible treatments; with four suspects, there are 16 possible treatments.

We recommend using all the possible treatments in the experimental plan. This is called a *full factorial experiment.* With three suspects, this means we need at least eight runs. Note that even if we do not replicate any treatments, the addition of other suspects in the verification experiment gives us some built-in replication (also called *hidden* replication in some texts). As seen in Table 13.2, half the experimental runs will be conducted with each

Table 13.2 A two-level factorial experiment with three suspects and eight treatments.

Treatment	Level of Suspect A	Level of Suspect B	Level of Suspect C
1	Low	Low	Low
2	Low	Low	High
3	Low	High	Low
4	Low	High	High
5	High	Low	Low
6	High	Low	High
7	High	High	Low
8	High	High	High

of the suspects at their low and high levels. For example, we use the low level of suspect C on four runs and the high level of suspect C on four runs.

To analyze the results of factorial experiments, we rely on comparing averages and graphical displays.

Brake Rotor Balance

In the brake rotor balance problem described in one of the case studies, rotors required rework if their balance weight was too high. Using a group comparison, the team found two suspect dominant causes, core position in the mold and core thickness. Another suspect came from the chronological link between a change to new tooling in the core-making process and the increase in balance weight rejects. The team planned a verification experiment with the three suspects.

In Table 13.3, we show the two levels of each suspect chosen to capture its full range of variation. The two core-related suspects were expected to produce good results at their nominal levels and deteriorating performance away from the nominal. The team decided to make eight rotors for each of the eight treatments. That is, there were eight runs with eight repeats. No treatment was replicated.

Table 13.3 Suspects and levels for brake rotor verification experiment.

Suspect	Low level	High level
Tooling	Old tooling (4-gang)	New tooling (6-gang)
Core position	Offset (20-thousandths of an inch)	Nominal
Thickness variation	30-thousandths of an inch	Nominal

Table 13.4 Brake rotor verification experiment results.

Treatment	Tooling	Core Position	Thickness variation	Run order	Average balance weight
1	4-gang	Offset	30-thousandths	8	0.56
2	4-gang	Offset	Nominal	1	0.17
3	4-gang	Nominal	30-thousandths	3	0.44
4	4-gang	Nominal	Nominal	7	0.08
5	6-gang	Offset	30-thousandths	2	1.52
6	6-gang	Offset	Nominal	5	0.37
7	6-gang	Nominal	30-thousandths	4	1.34
8	6-gang	Nominal	Nominal	6	0.03

To obtain the cores for the experiment, the team measured thickness variation and sorted cores until they had 16 with high thickness variation and 16 with nominal thickness variation from each set of tooling. The experiment required careful planning since the balance weight of each casting can only be determined after shipping and machining. The castings were tagged for identification after production and tracked through the subsequent process.

The team randomized the casting order of each treatment as given in Table 13.4. The 64 rotors were cast and machined as planned. The experimental plan and the measured balance weights are given in the file *brake rotor balance verification*. The average weight of the eight repeats for each run is given in Table 13.4.

To analyze the data, we plot the weights by treatment in Figure 13.4. We see there are large differences in the balance weights produced by different treatments and relatively little difference within each treatment. We also notice that we have seen roughly the full extent of variation in balance weight given by the problem baseline (the dashed line on Figure 13.4). If this were not the case, we would conclude that we have not found a dominant cause.

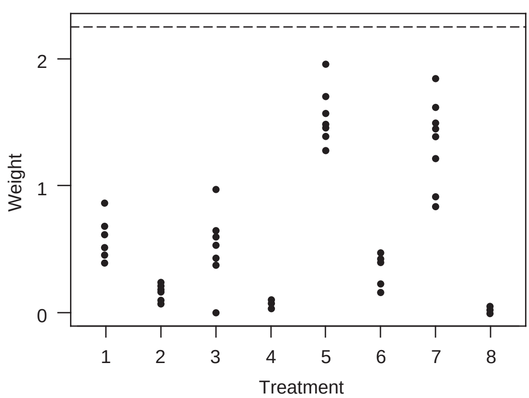


Figure 13.4 Weight by treatment for the brake rotor verification experiment (dashed line gives the full extent of weight variation).

We analyze the experimental results using the within-run average weight for each treatment. We estimate and rank the *effects* of the suspects. A *main effect* due to a particular suspect is the difference in average output (high-low) for that input. For example, using the data in Table 13.4, the main effect for tooling is:

$$\frac{1.52+0.37+1.34+0.03}{4}-\frac{0.56+0.17+0.44+0.08}{4}=0.50$$

An *interaction effect* measures the change in the main effect for one input as a second input changes. For example, the interaction effect for tooling and thickness is:

$$\frac{1}{2}\left\{\left(\frac{1.52+1.34}{2}-\frac{0.56+0.44}{2}\right)-\left(\frac{0.37+0.44}{2}-\frac{0.17+0.08}{2}\right)\right\}=0.43$$

The first term on the left side of the equation (except for an extra factor of ½) compares the effect of changing the tooling when the thickness variation is 30 thousandths. The second term measures the effect of changing the tooling when the thickness variation is nominal. Half the difference in the two effects is the interaction.

We can similarly define interactions for three inputs and more. Interactions are important if the effect on the output of changing levels in one suspect depends on the level of another suspect. In terms of finding a dominant cause, we may get extreme values of the output only if both inputs are at their high level. In this instance, we say that the dominant cause is an *interaction* between the two suspects.

In this experiment with eight treatment combinations, we can estimate seven effects. The number of effects estimable is always one less than the number of treatments. We fit a *full model* with all possible effects and construct a Pareto chart (see Appendix F for MINITAB instructions) of the unsigned effects to distinguish between those that are large and small. In Figure 13.5, we see that thickness variation, tooling, and the interaction between thickness variation and tooling are relatively large effects. The vertical dashed line in Figure 13.5 is added by MINITAB based on a test of significance. We ignore this dashed line in our interpretation.

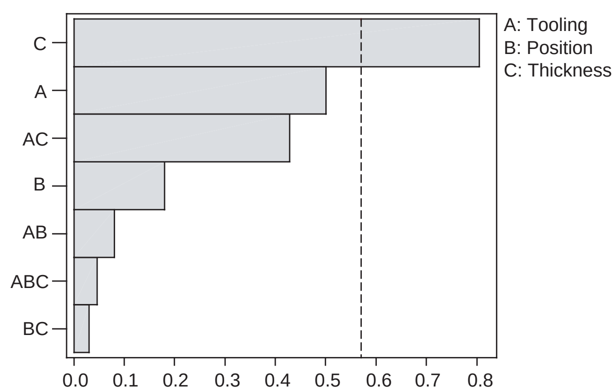


Figure 13.5 Pareto chart of the effects for brake rotor verification experiment.

Table 13.5 Balance weight averages by thickness variation and tooling.

	Thickness variation		
Tooling	**0.03**	**Nominal**	**Average**
4-gang	0.50	0.12	0.31
6-gang	1.43	0.20	0.82
Average	0.97	0.16	

The effect of core position is small, so we eliminate it as a suspect dominant cause. Since tooling and thickness variation have a relatively large interaction, we need to study the effect of these two suspects simultaneously. In Table 13.5, we calculate the average balance weight for the four combinations of tooling and thickness variation.

We can see the interaction in the table. The effect of changing thickness variation is much greater for the 6-gang tooling. We use a multivari chart, as shown in Figure 13.6, as a convenient way to display the interaction. Note here we plot the individual values as well as averages. From the chart, we see that for the 4-gang tooling, the effect of changing core thickness variation from nominal to the extreme level is moderate. However, for the 6-gang tooling the effect is very large.

In summary, the team found that core position had little effect, and the combination (that is, interaction) of the switch to the 6-gang tooling and thickness variation was a dominant cause. The conclusion that position was not important was a surprise, since that was identified as a suspect in a group comparison (see Chapter 12). This illustrates the potential danger of proceeding without verification. The team could have spent time and effort trying to reduce core position variation with little impact on imbalance.

We finish the story of the brake rotor balance problem in one of the detailed case studies on the CD-ROM.

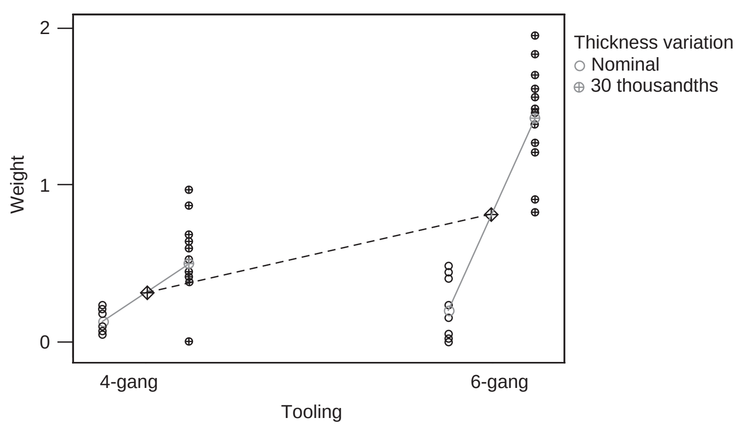


Figure 13.6 Interaction between tooling and core thickness variation.

13.4 FURTHER ISSUES AND COMMENTS

We discussed a number of issues in Section 13.2 relating to experiments to verify a single suspect. We need to consider all of these issues in any verification experiment. Here we consider some further issues relevant in the conduct of an experiment with several suspects.

How Short a List of Suspects Do We Need?

We recommend that you produce as short a list of suspects as possible before undertaking an experiment. If the list is long, the experiment becomes difficult to manage and expensive.

First, if the number of suspects is k, even with only two levels per input, the number of possible treatments is 2^k. For a full factorial design (one replicate of each treatment), the number of runs becomes unfeasible if the number of suspects is greater than about four. We can then use a *fractional factorial experiment* (see Chapter 15 and the supplement to Chapter 15) to reduce the number of runs, but we pay a price in terms of information about interactions. We look at another alternative in the chapter supplement.[4] Second, all of the inputs in the list of suspects are naturally varying and, during each run of the experiment, we need to control the values to define the treatment for the run. With many suspects, the task of simultaneously controlling the inputs becomes very difficult.

We illustrate the dangers of searching for a dominant cause among a long list of suspects using experimental plans with the following example. A team was assigned to reduce scrap due the geometric distortion of rear axle hypoid gear components during heat treatment. The pinion and gear set are critical to provide smooth and quiet transmission of power from the driveshaft to rear axle driveline. In a baseline investigation, the team found the scrap rate was 6%. Using Pareto analysis, they focused on the pinion. They decided to search for a dominant cause without using the recommended method of elimination. Based on existing process knowledge, the team identified five suspect causes: the average size of four pinion gear teeth, variation in tooth size, stem runout before heat treatment, position of the cone in basket (up or down) and the furnace track (left or right). They planned a full factorial experiment using all 32 possible treatment combinations of the five suspects at two levels each. They decided to have two pinions (repeats) per run, so they needed 64 pinions in total.

The team could easily control the last two suspects for each run. However, to find the required eight pinions for each of the eight treatment combinations defined by the first three suspects, the team needed to measure and sort pinions before heat treatment. They continued measuring and sorting for a number of days. Then disaster struck. During the night shift, there was a shortage of parts for the heat treatment process. There was great pressure to sustain volumes, so when members of the night shift found the pinions set aside for the experiment, they processed them. The next day the team decided the proposed experiment was too complicated. They decided to look for a dominant cause with simpler observational investigations.

Can We Verify the Suspects One at a Time?

To verify a dominant cause from a list of suspects, we do not recommend the use of one-at-a-time experiments, where we change only a single suspect at a time. These traditional experiments are a poor choice because they cannot find important interactions between the suspects. For more information on the dangers of one-at-a-time experiments, see Montgomery (2001).

Sometimes, we can test all the suspects simultaneously. In the window leaks example discussed in Chapter 12, the team isolated the primary seal gap and the plastisol application as suspects. To verify, the team selected eight trucks that had passed the aggravated leak test and reinstalled the rear window using different windows and seals, ensuring that the primary seal gap exceeded 1.5 millimeters and that the plastisol score was between 3 and 4. Seven of the eight trucks leaked when they were retested. The conclusions from the verification experiment and group comparison are summarized in Figure 13.7. All the experimental runs came from the upper left quadrant and the results verified that the combination of poor plastisol application (low score) and high primary seal gap resulted in leakers. The team proceeded to improve the plastisol application process by ensuring that all trucks received a plastisol score greater than 4.5.

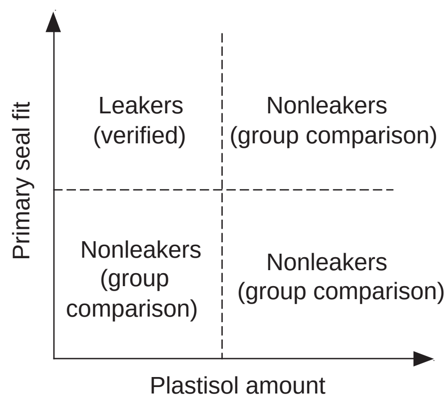


Figure 13.7 Window leaks verification experiment results.

Verification Experiment Summary

We have a list of one or more suspects.

Question

In the current process, are any of the suspects a dominant cause?

Plan

- For each suspect, choose two levels at the extremes of their normal range.
- Define the runs using available information about the time-based family that is home of the dominant cause.

- Select the number of runs. For:
  –A single suspect: use at least three replicates for each level (at least six runs)
  –Two suspects: use a full factorial design and at least two replicates (at least eight runs)
  –Three or more suspects: use a full factorial design with at least one replicate for each treatment
- Determine the number of repeats for each run.
- Randomize the order of the runs as much as possible.
- Make everyone potentially impacted aware of the plan.

Data

Carry out the experiment. Record the output, suspect levels, treatment number, and order for each run. Use a separate row for each repeat.

Analysis

- Plot the output against the treatment number. Add horizontal lines showing the full extent of variation in the output.
- For several suspects, fit a full model and construct a Pareto plot of the main and interaction effects. For large interactions, create a multivari chart with the corresponding inputs.

Conclusion

- If the output does not show the full extent of variation, then none of the suspects is a dominant cause.
- A suspect with a large effect relative to the full extent of variation is a dominant cause.
- If there is a large interaction, the dominant cause involves two (or more) suspects.

We summarize the terminology for experimental plans in Table 13.6.

Table 13.6 Designed experiments terminology.

Term	Meaning
Balance	An experimental design is balanced if there is the same number of replicates for each treatment and the same number of repeats in each run.
(Full) factorial experiment	An experimental plan where all possible combinations of the input levels are used to define the treatments.
Interaction effect	The change in the main effect for one input as a second input changes.

Continued

Table 13.6 Designed experiments terminology. *(Continued)*

Term	Meaning
Levels	The different values of an input characteristic used in an experiment.
Main effect	The change in the average output produced by a change in one input.
Randomization	Radomizing the order of the runs.
Repeat	More than one part (or measurement) made within a run.
Replication	Carrying out more than one run for each treatment.
Run	Assigning the treatment, running the process, and measuring the output.
Treatment	A particular combination of input levels.

Key Points

- We use an experimental plan to verify that one or more suspects obtained using the method of elimination is a dominant cause.
- For each suspect we choose two levels at the extremes of its normal range.
- In the verification experiment, we recommend:
  –For a single suspect: three or more replicates for each level
  –For two suspects: two or more replicates at each of the four treatments
  –For three or four suspects: a full factorial experiment
- We identify a dominant cause by examining main effects and two input interactions.
- In the analysis, we check that the full extent of variation in the output is seen over the runs of the experiment. Otherwise, none of the suspects varied in the experiment is a dominant cause.

Endnotes (see the Chapter 13 Supplement on the CD-ROM)

1. The difference between repeats and replicates is a great source of confusion in the analysis of experimental data. We explore this issue briefly in the supplement to justify the analysis methods we propose.
2. When we identify a suspect cause using the method of elimination, we know that the pattern of variation in the suspect must match that of the output. To verify that the suspect is a dominant cause, we must show that there is no other cause that also matches this pattern of variation. In the

supplement, we explain how randomization and replication can help to reduce the risk of falsely identifying a suspect as the dominant cause. We also explain how we implicitly use blocking, the third fundamental principle of experimental design, to help to avoid this risk.
3. See note 2.
4. In the supplement, we briefly discuss an alternative verification experiment based on a technique called *variables search*.

Exercises are included on the accompanying CD-ROM

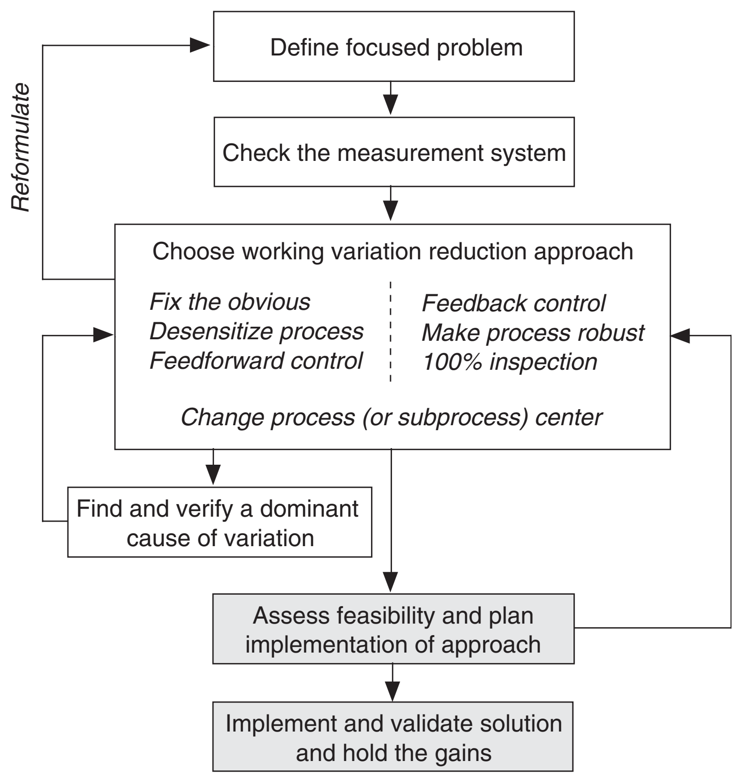
PART IV

Assessing Feasibility and Implementing a Variation Reduction Approach

Opportunity is missed by most people because it is dressed in overalls and looks like work.

—Thomas Edison, 1847–1931

In this final part of the book, we address revisiting the choice of working variation reduction approach, the issues around the assessment, implementation of each approach, and the validation of a solution. The choice of working approach may need to be reconsidered in light of the process knowledge obtained in a search for a dominant cause or in assessing feasibility of a particular approach. We provide detailed how-to directions for assessing feasibility, including consideration of costs, and for implementing each of the seven approaches. We also discuss validating a solution.



Statistical Engineering variation reduction algorithm.

14

Revisiting the Choice of Variation Reduction Approach

If at first you don't succeed, try, try again.

—Proverb

After conducting a search for a dominant cause, the team should reconsider their choice of working variation reduction approach. In this chapter, we consider the options and how to choose an approach.

If a specific dominant cause has not been found, the team must select one of the three non-cause-based approaches, resume the search for a more specific cause, or abandon the project. If a dominant cause has been found, the team can now consider the feasibility of any one of the variation reduction approaches. They will have accumulated substantial information about the behaviors of the output and the causes that can be used in selecting the approach.

There are four options that directly use the knowledge of the dominant cause:

- Fix the obvious.
- Compensate for the variation in the dominant cause (desensitization or feedforward control).
- Reformulate the problem in terms of the dominant cause and reapply the algorithm (that is, reduce variation in the cause).
- Continue the search for a more specific dominant cause. This is an informal version of reformulation.

The variation reduction approaches that do not require knowledge of a dominant cause should also be considered:

- Use feedback control on the output.
- Make the process robust to variation in the unknown dominant cause.
- Implement or tighten 100% inspection on the output.

Given the knowledge accrued during the search for the cause, the non-caused-based approach may be more or less feasible. The final option is to abandon the project and devote the resources to another problem with a greater likelihood of success.

We explore how to decide among these choices. Since there are few general rules, we present a large number of examples. We consider costs and likelihood of success. There is considerable risk and uncertainty in each choice, but we cannot proceed without making this decision.

14.1 FIXING THE OBVIOUS: IMPLEMENTING AN AVAILABLE SOLUTION

In many problems, once a dominant cause has been identified, the team has sufficient knowledge to determine an obvious solution and move to the Implement and Validate Solution stage of the algorithm. In making this decision, the team must consider possible side effects and the cost of the fix relative to the costs associated with the problem.

The definition of *obvious* depends on the process and the level of process knowledge. The team has an obvious fix if they are confident that it is feasible and there is no need to learn more about the process before it can be implemented. The key issue here is cost. There may be obvious solutions to the problem available at the start of the project (for example, buy new equipment), which are eliminated because of high cost. The team may now identify a lower-cost obvious solution that was not apparent without knowledge of the dominant cause.

There are many examples.

Engine Block Leaks (Center Leaks)

In Chapter 6, we described a project of leaking engine blocks and how the project was focused to three problems related to leaks at three different locations. The team discovered that the dominant cause of the center leakers was core breakage during the casting process. Based on their process knowledge, the team knew they could eliminate or reduce the core breakage by better supporting the core in the mold during the pouring operation. The team added several chaplets (small steel inserts) set in the mold to support the core. They knew that this process change would produce no unfavorable side effects. The change eliminated the problem of center leakers. The extra cost of the chaplets was more than justified by the reduced scrap costs.

Engine Block Leaks (Rear Intake Wall)

The team also found leaks at the rear intake wall of the casting. Using a group comparison (see Chapter 12), they discovered that the dominant cause of this type of leak was wall thickness at a specific location in the left rear wall. Thin walls led to leaks. An obvious solution was to change the mold to make the wall thicker. This process change could be made with a small one-time cost but had negative consequences in terms of the weight of the block. The team made the change and virtually eliminated rear intake wall leakers.

Truck Pull

In Chapter 10 and elsewhere, we described a project to reduce variation in a truck wheel alignment process. The team had access to alignment data that was automatically collected for every truck produced. Each truck was measured on one of four alignment machines. At one point during the project, the team looked at the right caster data, stratified by the alignment machine. The daily averages are plotted in Figure 14.1.

The team was surprised to see the persistent differences among the four gages. The trucks enter the gages haphazardly, so the observed differences must be due to differences in the alignment machines. The alignment machine was a dominant cause of variation in right caster. This cause was not acting during the baseline investigation or the initial investigation of the measurement system. The team took immediate action to recalibrate the four gages to remove the systematic differences. To prevent recurrence, they established a daily monitoring program to correct such differences. There was a small cost associated with the daily check.

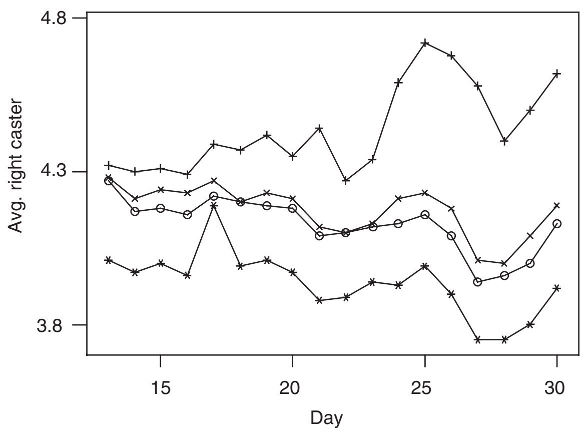


Figure 14.1 Right caster daily averages by alignment machine.

Cylinder Head Shift

In a casting operation, excess side shift variation was a problem. A multivari investigation (see Chapter 11) found the dominant cause of side shift acted in the pattern-to-pattern (cavity-to-cavity) family. The team had the process knowledge to make a one-time adjustment to the dies to shift the pattern averages and reduce the pattern-to-pattern side shift. They then monitored the process to detect the recurrence of the problem before it became material. The value of the variation reduction was greater than the costs of the one-time pattern adjustment and the monitoring.

Fascia Cratering

In the fascia cratering example discussed in Chapter 11, the team found that major cratering occurred on every tenth fascia. They concluded that the dominant cause was the mold release spray. With this knowledge, an obvious (to the team) and immediate short-term solution was to wipe the mold after the spray. This solution had the potential to introduce dirt in the mold and met with resistance from the mold operators. The obvious fix was not feasible. The team decided instead to investigate different mold sprays. That is, they adopted desensitization as the working approach.

Window Leaks

As reported in Chapter 12, a team found that the dominant cause of truck rear window upper ditch leaks was an interaction between the primary seal gap and the plastisol application to the seams. We reproduce the results in Figure 14.2.

The pattern on the plot suggested an obvious fix. The instructions were changed so that the operators where the plastisol was applied, brushed the plastisol to ensure coverage of the critical seals. Because of this change, the rear window upper ditch leaks were totally eliminated. Note that the variation in primary seal gap was not changed. The process was made less sensitive to this variation. The team also decided to pursue another solution based on changing the primary seal to reduce the variation in gap. The operating costs of this proposed solution were less than those of the obvious fix. However, management postponed implementation of this alternative solution because it required robotic application and substantial capital expenditure.

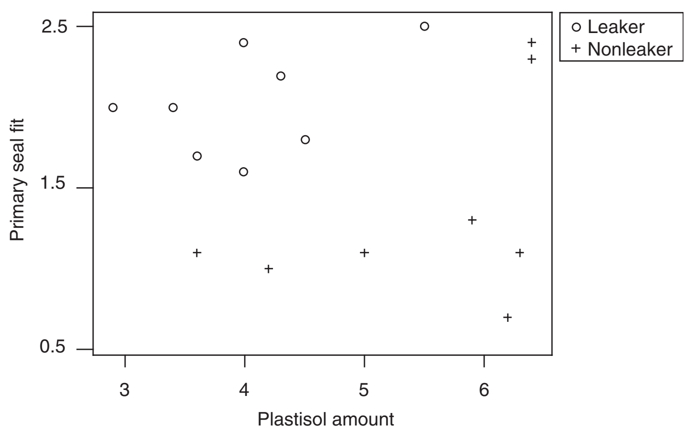


Figure 14.2 Plot of primary seal fit by plastisol amount.

Hubcap Damage

Because of customer complaints, a team was assigned a goal of reducing the incidence of wheel trim and hubcap damage. The team discovered that the dominant cause of broken retaining legs and other damage was a combination of cold weather and contact with curbs. Through comparison with competitors, they found that an obvious fix was to change the hubcap material and design. The team replaced the brittle existing ABS hubcap with a new design made of mineral-reinforced polypropylene. In addition, they increased the number of retaining legs. The increased cost was justified by the elimination of customer complaints.

Assembly Nuts Loose

In an assembly process, a power wrench driven by pneumatic pressure tightens nuts to a specified torque. The dominant cause of loose nuts was found to be occasional drops of pneumatic pressure. An air pressure meter was installed in the air line. An alarm sounded if the pressure dropped below a critical point providing constant inspection of the cause. This is an example of error proofing as defined by Shingo (1986).

Comments

The Fix the Obvious approach is a catchall category that makes use of the other variation reduction approaches. For example:

- 100% Inspection (for example, loose assembly nuts)
- Moving the output center (for example, rear intake wall leakers, cylinder head side shift, truck pull)
- Desensitizing the process (for example, window leaks)

When applying the Fix the Obvious approach, the team should ensure the problem does not recur. In the truck pull, camshaft journal diameter, and cylinder head shift examples, the fix was obvious but all three problems were likely to recur unless a new control or regular maintenance scheme was put in place.

Fix the Obvious is the preferred choice because it applies when there is a clear solution to the problem. There is little uncertainty that the fix will be effective. The main considerations are the cost of the fix relative to the gain and potential negative side effects.

14.2 COMPENSATING FOR VARIATION IN THE DOMINANT CAUSE

If there is no obvious solution that is economical, we can look for changes in the process (that is, changes in fixed inputs) that will eliminate or reduce the effect of the dominant cause. In particular, we are interested in deciding whether we can compensate for variation in the dominant cause through process desensitization (Chapter 16) or feedforward control (Chapter 17).

Refrigerator Frost Buildup

In Chapter 1, we described the problem of frost buildup in refrigerators, where the dominant causes were environmental and usage inputs outside of the control of the manufacturer. The team could not reduce variation in the causes and adopted the Desensitization approach. The team had several design changes in mind that might reduce the effect of the environmental causes of frost buildup. See Chapter 16 for further details.

Iron Ore Variation

A team was charged with reducing variation in iron ore composition as the ore was loaded into ships. This was a reformulated problem from the customer, a steel mill that wanted to reduce variation as the ore was processed. The dominant cause of the variation was the composition of the ore as it came out of the ground. To compensate for this cause, the team decided to look at new ways to create and dismantle stockpiles that were placed in the process to help blend the ore. They could not estimate the costs or benefits without further investigation.

Crossbar Dimension

In Chapter 12, we discussed a problem to reduce variation in a crossbar dimension of a molded part. From the baseline investigation, the full extent of variation was –0.3 to 2.0 thousandths of an inch. The team found that barrel temperature was a dominant cause as shown in Figure 14.3.

At first, the team considered reformulating the problem. Reducing variation in barrel temperature would result in reduced variation in crossbar dimension. Looking closely at Figure 14.3, the team realized there was evidence of a nonlinear relationship between barrel temperature and crossbar dimension. We have added a quadratic fit to the scatter plot to make this conclusion clearer. The variation in crossbar dimension is greater as the barrel temperature varies from 74° to 77° than it is if the barrel temperature ranges from 77° to 80°. The team decided to try to desensitize the process to barrel temperature variation by increasing the average barrel temperature.

The team could not predict the benefits of increasing the average barrel temperature from the data shown in Figure 14.3 since they needed to extrapolate beyond the normal range of barrel temperature. The team could assess the direct cost of the change but they were unsure if the process would gracefully tolerate barrel temperatures much above 80°. That is, there might be negative side effects. Further investigation was required.

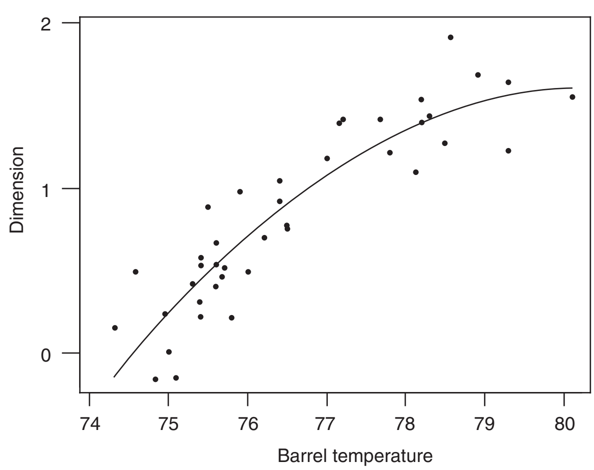


Figure 14.3 Scatter plot of crossbar dimension versus barrel temperature.

Comments

There is uncertainty about finding an effective solution if the team decides to compensate for the effects of a dominant cause. In the refrigerator frost buildup example, the team did not know if they could find affordable changes to the design that would desensitize the refrigerator to changes in the usage and environmental inputs. They could not quantify the benefits until they had investigated the changes. They needed to carry out further process investigations with no certainty of an efficient and effective resolution. Fortunately, they were able to find good design changes. Otherwise, they would have had to absorb the high costs of investigation and reconsider the other variation reduction approaches.

14.3 REFORMULATING THE PROBLEM IN TERMS OF A DOMINANT CAUSE

We can reduce output variation by reducing the variation in a dominant cause. In reformulation, we restart the algorithm with a new problem defined in terms of the dominant cause. We specify the goal for the new problem so that we meet the goal of the original problem. Reformulating the problem is the classic route to variation reduction. Rephrased, we are told to find the "root cause" of the problem and then somehow deal with that cause, hopefully with an obvious fix.

Reformulation is attractive because the problem is moved upstream. This may reduce the cost and complexity of the solution. Reformulation is unattractive because we are replacing one problem with another and we may be no closer to a solution. Note that we may adopt any of the seven variation reduction approaches to solve the reformulated problem. Also, we may reformulate several times before adopting one of the variation reduction approaches.

Sunroof Flushness

Customers were dissatisfied if the sunroof was not flush with the roof of the car. In the sunroof installation process, there was a 90% rework rate due to flushness variation. A team set out to reduce the rework costs.

Flushness was the difference in height between the sunroof seal and the metal roof. It was measured using digital calipers at six points, three at the front and three at the back of the sunroof. A baseline investigation showed that flushness variation was largest at the two front corners. Front corner flushness ranged between –3.5 and 4 millimeters, with an estimated standard deviation of 1.25 millimeters. The team established a goal of reducing the front corner flushness standard deviation to less than 0.5.

Based on engineering knowledge, the team suspected the dominant cause of flushness variation was either crown height or attachment pad height. When the roof panel was adapted to allow installation of a sunroof, six attachment pads were added. The team carried out an investigation using two sets of six vehicles with extreme flushness. For each vehicle, they removed the sunroof module and measured the attachment pad heights and roof crown height at the front and back. The data are given in the file *sunroof flushness input-output*. Here we report the results for left front flushness only. The team used regression analysis

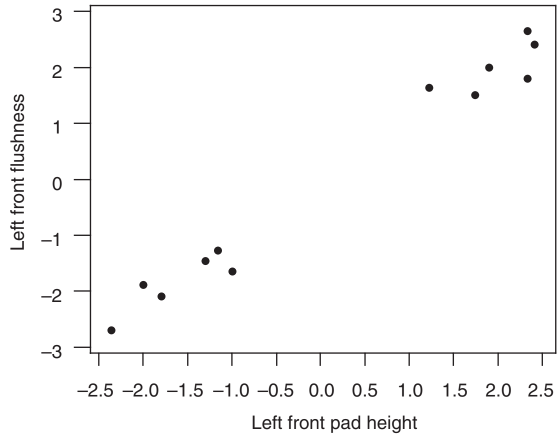


Figure 14.4 Scatter plot of flushness versus left front pad height.

and the scatter plot shown in Figure 14.4 to demonstrate a clear connection between the corner flushness and pad height closest to the corner. Roof crown height was eliminated as a suspect.

The team assumed that pad height was the dominant cause of the variation in flushness without verifying this conclusion. They decided to reformulate the problem in terms of pad height. They carried out a baseline investigation and found the standard deviation in pad height was 1.18 millimeters.

To determine a goal for the reformulated problem, they used the fitted equation from the regression analysis:

```
The regression equation is
left front flushness = -0.126 + 1.05 left front pad height

Predictor        Coef     SE Coef         T        P
Constant     -0.12631     0.09352     -1.35    0.207
left fro      1.04811     0.05015     20.90    0.000

S = 0.3222      R-Sq = 97.8%     R-Sq(adj) = 97.5%

Analysis of Variance

Source           DF          SS          MS        F        P
Regression        1      45.348      45.348   436.87    0.000
Residual Error   10       1.038       0.104
Total            11      46.386
```

The regression analysis corresponds to modeling the output/cause relationship as

$$flushness = a + b^{*}\ pad\ height + noise$$

where $a + b^*$ pad height represents the effect of the dominant cause and *noise* the effects of all other causes. The model constants a (intercept) and b (slope) describe the approximate linear relationship between the output flushness and the dominant cause pad height. Based on the model (see Chapter 2), and assuming that the dominant cause varies independently of the other causes, we have

$$\mathrm{sd}(\textit{flushness}) = \sqrt{b^2 \mathrm{sd}(\textit{pad height})^2 + \mathrm{sd}(\textit{noise})^2}$$

From the regression analysis, we estimate b as 1.05. From the two baseline investigations, we have estimates of the standard deviations of flushness and pad height. Substituting the three estimates into the given equation, we get an estimate for the standard deviation of the noise to be 0.118 millimeters.

Now, we can use the equation to translate the goal of reducing the flushness standard deviation variation to 0.5 into a goal in terms of pad height variation. We use the estimate for the slope b from the regression analysis and the estimated standard deviation of *noise* from the given calculation. Substituting the estimated values for b, the standard deviation of *noise,* and the goal for flushness standard deviation, we get

$$0.5 = \sqrt{1.05^2 \,\mathrm{stdev}(\textit{pad height})^2 + 0.165^2}$$

Solving gives 0.45 as the goal for pad height standard deviation. To meet the original goal for flushness, in the reformulated problem we need to reduce pad height variation by over 60%.

Battery Seal Failure

Management assigned a team to reduce the frequency of battery leaks. Using a group comparison, the team determined that the dominant cause of leaking was seal strength between the top and bottom of the battery casing. See the results in Figure 14.5. They concluded that stronger seals would eliminate the leak problem. The problem was reformulated to increase the average seal strength. Based on Figure 14.5, the team set the goal to increase seal strength to a minimum of 320 pounds. They did not try to quantify the relationship between the continuous cause, seal strength, and the binary output and hence, they could not quantify the potential gain of increasing the average seal strength.

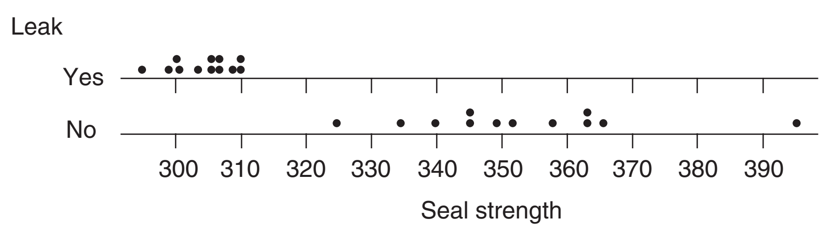


Figure 14.5 Dot plot of seal strength for leakers and nonleakers.

In this example, reformulation is especially useful since the original output is binary and the dominant cause is continuous. It should be easier to find a solution to the reformulated problem than the original since, in any investigation, more information is available from each battery when we measure a continuous as compared to a binary output. Since seal strength is a destructive measurement, this gain is offset by increased costs.

Roof Panel Updings

In a stamping operation, there was a problem with updings (small outward dents in the metal surface) on a roof panel. There was 100% inspection to sort the panels. Those with updings were reworked. The team established the baseline and also found that the measurement system was adequate. Next, using a multivari investigation (see the exercises in Chapter 11), they found that the dominant cause acted in the pallet-to-pallet family. Further investigations showed that updings were caused by particles on the metal blanks.

The team decided to reformulate the problem in terms of the amount of particulate on the steel blanks. Since the dominant cause acted from pallet to pallet, the team was able to use leverage (see Section 9.3) to roughly define the baseline for the amount of particulate. They determined the full extent of variation for particulate amount by counting the total number of particulates on blanks selected from two pallets that had produced stamping with extreme low and high upding counts. This was an inexpensive way to establish the baseline variation compared with the recommended investigation for a new problem as given in Chapter 6.

Comments

We choose to reformulate only after we have considered the other options. Reformulation of a problem is not a solution. That is, we cannot continue to reformulate indefinitely. Eventually one of the variation reduction approaches must be applied. If we can repeatedly find a dominant cause, we can reformulate the problem a number of times. However, the potential benefit arising from solving the reformulated problem is reduced in each iteration by the variation not explained by the identified dominant cause(s). Also, the final cause may be outside local control.

With reformulation, the dominant cause (an input) becomes the output. We then establish a baseline for the new output. We set the goal for the reformulated problem by exploiting what we have learned about the relationship between the cause and the original output. We can set this goal formally, as in the sunroof flushness example, or informally, as in the battery seal failure example.

There is little value in reformulating a problem in terms of a cause that is not dominant. Suppose, as in the sunroof flushness example, we found a cause that is linearly related to the output so that we have the equation

$$\text{stdev}(\mathit{output}) = \sqrt{b^2\text{stdev}(\mathit{cause})^2 + \text{stdev}(\mathit{noise})^2}$$

If the cause is not dominant, then the standard deviation of the noise, the variation due to all other causes, is large fraction of the standard deviation of the output. Even if we completely eliminate the effect of the cause, we do not reduce the output standard deviation substantially.

14.4 CONTINUING THE SEARCH FOR A MORE SPECIFIC DOMINANT CAUSE

Continuing the search for a more specific dominant cause is similar to reformulation, but we proceed less formally. The key is the baseline we use in the search. With reformulation, we determine a new baseline and goal in terms of the identified dominant cause. If we continue the search for a more specific cause, we use the original baseline and continue to measure the original output.

We select this option when the identified dominant cause is vague and there is no obvious solution. In the truck pull example described earlier, the dominant cause of variation was systematic differences among the four gages. This was not a specific cause since we did not identify what was different among the four gages. However, an obvious fix was available.

Here are some examples when the team decided to look for a more specific cause.

Engine Oil Consumption

In the oil consumption example, discussed in Chapter 13, the supplier of the valve lifters was a dominant cause of variation. The two engine assembly plants had different valve lifter suppliers. An immediate fix to reduce customer complaints was to eliminate the valve lifter supplier linked to the excess oil consumption. However, moving to a single supplier was not practical for volume and contractual reasons. Since the identified cause was a large family of inputs including all the specific differences between the two suppliers, the team decided to look for a more specific dominant cause. In particular, the team explored the variation within the valve lifters from the problem supplier to determine a more specific cause of oil consumption.

Camshaft Lobe BC Runout

In the camshaft lobe runout example discussed in Chapter 10 and elsewhere, the problem was excess base circle runout. Through a number of investigations, the team found that a dominant cause of runout variation was heat treatment spindle as illustrated in Figure 14.6. The data for Lobe 12 are given in the file *camshaft lobe runout variation transmission*.

Based on this dominant cause, the team could not see how to apply one of the variation reduction approaches. To keep up with volume requirements, they had to continue using spindle 3. One possible solution was to reduce the runout center for spindle 3. However, the team did not know of an adjuster, and since runout is a measure of variation itself, finding

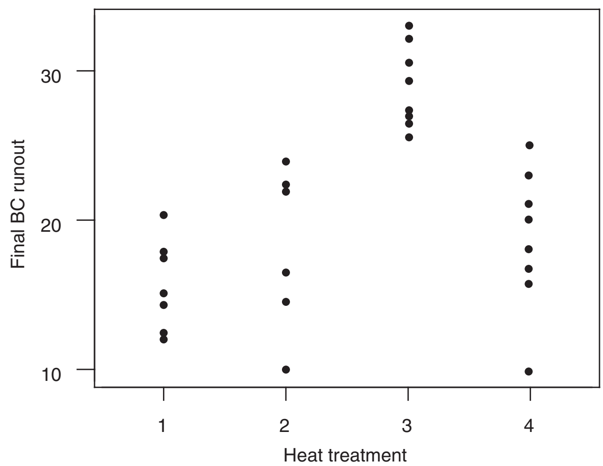


Figure 14.6 Plot of final base circle BC runout by heat treatment spindle.

an adjuster seemed difficult. The team decided to look for a more specific cause. They proceeded by comparing the heat treatment spindles, looking for reasons why they were different. In the end they were unable to find a specific cause. Out of desperation they arranged for overdue maintenance on the heat treatment operation. This somehow eliminated the spindle-to-spindle difference. Next, the team had to worry about how to prevent recurrence of the problem. They decided to monitor the runout after heat treatment on two parts per day from each spindle. Any large spindle-to-spindle differences triggered maintenance of the heat treatment operation.

Crankshaft Main Diameter

In the crankshaft main diameter problem considered in one of the case studies, the dominant cause of final diameter variation was found to be the diameter at an intermediate processing step. There was no obvious fix based on this knowledge. Also, there was no way to desensitize the final diameter variation to diameter variation at the intermediate step. The only processing step between the intermediate and final gages that changed the diameter was lapping to polish the surface. The team opted to look for a more specific upstream cause and continued to use the final diameter baseline.

Comments

We distinguish between reformulation and continuing the search for a more specific cause based on whether or not we determine a new baseline. We are more likely to reformulate in cases where the cause is a different continuous characteristic than the output. If the dominant cause is a discrete characteristic, like machine number or process stream, we typically continue the search for a more specific dominant cause without reformulating.

14.5 DEALING WITH DOMINANT CAUSES THAT INVOLVE TWO INPUTS

When a dominant cause is not a single input, there are a variety of ways to proceed. The choice of the best variation reduction approach depends on the nature of the dominant causes and the problem context. Sometimes we can address such problems by addressing only one of the causes. In other cases, we need to address both causes either simultaneously, or separately, possibly with different variation reduction approaches.

We illustrate using three examples.

Window Leaks

In the truck window leaks problem, the dominant cause of upper ditch leaks was the combination of a large primary seal gap and poor plastisol application to the seams. The proposed and implemented solution was a process change that required operators to brush the plastisol to ensure coverage of the critical seals. With this solution, the team addressed only one of the inputs involved in the dominant cause.

Door Closing Effort

The dominant cause of door closing effort, as discussed in Chapter 10, involved both the component and assembly families. Here there were two large causes acting in separate families. The team proceeded to address each cause (family) separately. Since each cause on its own was not dominant, they could not meet the project goal without addressing both.

Steering Wheel Vibration

In a problem to reduce steering wheel vibration, the dominant causes were found to be imbalances in two transmission components. Both components contributed equally to the problem. The proposed solution was to desensitize the process to both causes by vectoring the two components during the installation so that their imbalances tended to cancel out. Alternately, the team could have reformulated into two problems to address the imbalance in each component separately.

14.6 SUMMARY

The thought process in choosing how to proceed after finding a dominant cause is summarized in Figure 14.7.

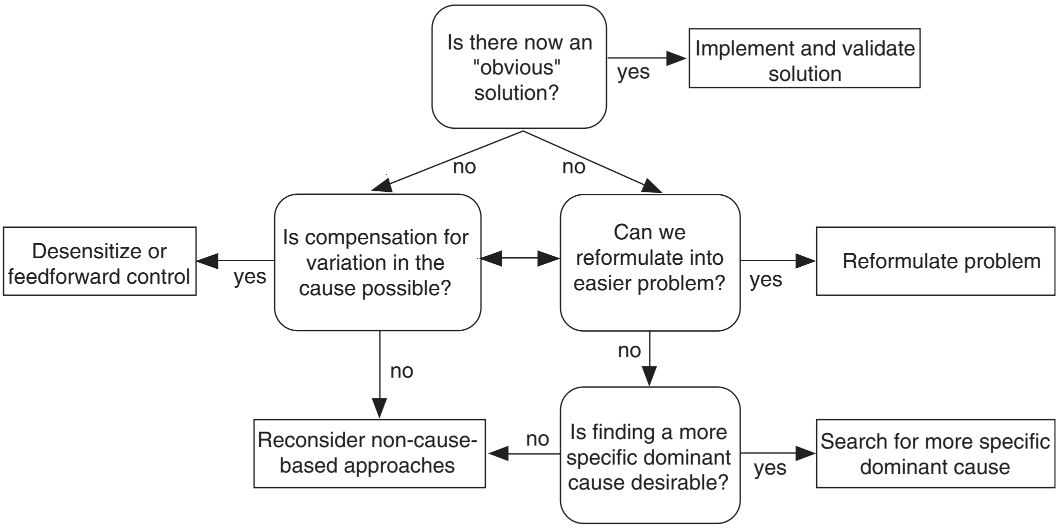


Figure 14.7 Flowchart to help decide how to proceed after finding a dominant cause.

Key Points

- There are three options once a dominant cause has been found:
  - –Fix the obvious.
  - –Compensate for the cause using desensitization or feedforward control.
  - –Reformulate the problem in terms of the dominant cause and reapply the algorithm (that is, reduce variation in the cause).
  - –Continue searching for a more specific dominant cause.
- Subject to cost constraints and concerns over side effects, implementing an obvious fix is the most desirable option.
- There is a risk of choosing an approach that will be neither effective nor efficient because of the lack of knowledge. However, we cannot proceed with making a choice.

Exercises are included on the accompanying CD-ROM

15

Moving the Process Center

This time, like all times, is a very good one, if we but know what to do with it.

—Ralph Waldo Emerson, 1803–1882

Moving the process center is the only approach that directly addresses an off-target process center. The goal of the approach is to find a way to move the process output center either closer to the target or in the desirable direction if higher or lower is better. An example for a higher is better output is shown in Figure 15.1. We do not need to identify a dominant cause to apply this approach.

The only requirement for moving the process center is an *adjuster;* that is, a fixed process input that can be changed to move the process output center.

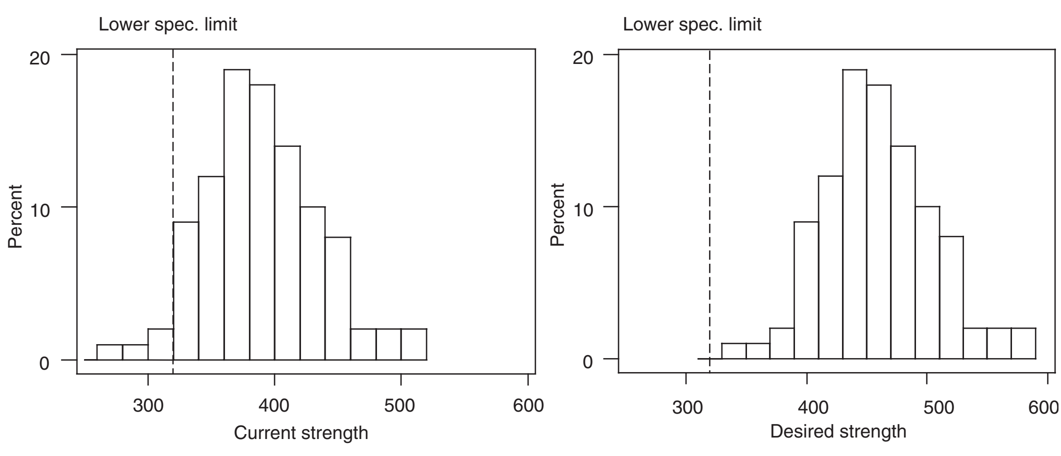


Figure 15.1 Changing the process center.

In many processes, the team will have an available adjuster and thus an obvious fix to the problem. To find an adjuster, we recommend an experimental plan. We use engineering knowledge and experience with the process to choose fixed inputs, called *candidates,* to vary in the experiment. There is a risk that an adjuster will not be found.

The potential costs of moving the process center include the costs of:

- An experiment to find an adjuster, if necessary
- A one-time change to the adjuster
- The ongoing operation of the process at the new setting for the adjuster

To assess the potential benefit of moving the process center, we imagine shifting the baseline process center to a new value.

15.1 EXAMPLES OF MOVING THE PROCESS CENTER

To illustrate, we discuss three examples where the team had to first find an adjuster before they could apply the Move the Process Center approach.

Battery Seal Failure

In the production of automotive batteries, a sealing process joins the top of the batteries to the base. Due to field complaints about leaky seals and high rework costs in the plant, management assigned a team to address the problem. Looking back at several months' production records, the team estimated that the baseline within-plant leak rate was 2.3%. The goal was to reduce this rate to less than 0.5% with the assumption that this would eliminate most of the field failures.

Note that the output is binary. Seals either leak or do not leak. The team suspected that a dominant cause of leaks was low tensile strength of the seal. To assess this idea, the team compared the tensile strengths (in pounds) of 12 leaky seals and 12 good seals. From Figure 15.2, we see that leaky seals are strongly related to low tensile strengths. Based on this conclusion, they

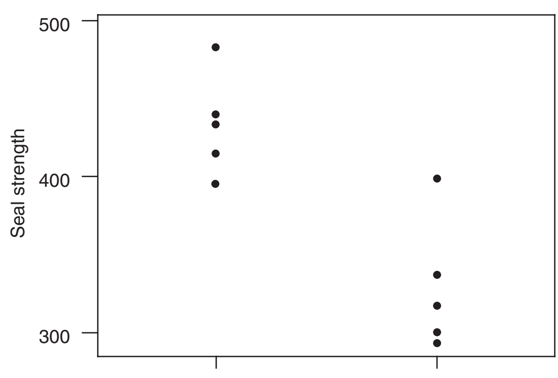


Figure 15.2 Battery seal group comparison results.

reformulated the problem in terms of tensile strength. The tensile strength was measured on a sample of 100 batteries selected from one week's production. The data are stored in the file *battery seal strength baseline* and the results are shown in the histogram in the left panel of Figure 15.1. From the group comparison, the baseline investigation, and engineering knowledge, the team set the minimum acceptable seal strength at 320 pounds, which corresponds to increasing the center of the process output by about 60 pounds. The desired process histogram is shown in the right panel of Figure 15.1. At this point, the team did not attempt to determine if this shift in the process center would meet the initial project goal of reducing the leak rate to less than 0.5%.

The team set out to find an adjuster. Based on engineering and process knowledge, they chose three candidates (fixed inputs) at two levels each. The further apart the levels, the more likely the experiment will detect an effect and the greater the risk of negative side effects. They planned a factorial experiment with eight runs. The candidates and levels are given in Table 15.1.

Table 15.1 Heat seal experiment candidates and levels (existing process levels are indicated by *).

Candidate	Low level	High level
Melt temperature	750°F	800°F*
Melt time	2.5 seconds*	3.1 seconds
Elevator speed	low	high*

In each of the eight experimental runs, five batteries were produced. The runs were conducted in random order. Treatment 7 corresponds to the existing process. The data are given in the file *battery seal strength move center* and in Table 15.2. There were no leaking seals in the 40 batteries produced in the experiment.

Table 15.2 Treatments and seal strength for battery seal experiment.

Treatment	Order	Melt temperature	Melt time	Elevator speed	Seal strength
1	7	Low	Low	Low	413, 505, 489, 452, 465
2	6	Low	High	Low	468, 493, 484, 393, 423
3	8	High	Low	Low	383, 368, 280, 377, 370
4	4	High	High	Low	383, 365, 353, 389, 353
5	1	Low	Low	High	440, 415, 483, 395, 433
6	3	Low	High	High	466, 387, 505, 393, 456
7*	5	High	Low	High	399, 294, 317, 300, 337
8	2	High	High	High	373, 379, 383, 385, 345

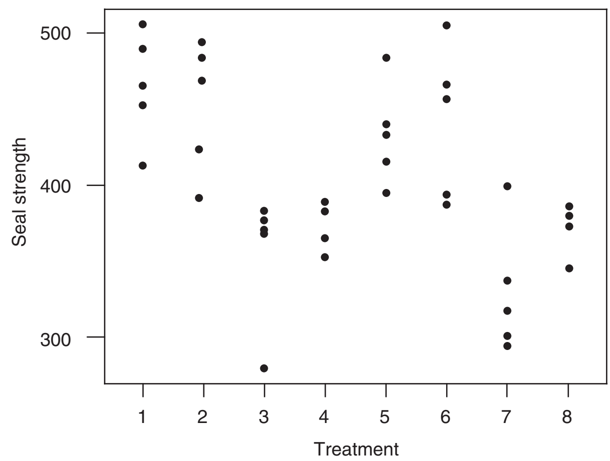


Figure 15.3 Seal strength by treatment.

In the analysis, we first plot the data by treatment as in Figure 15.3. We see, encouragingly, that many treatments have average strength greater than treatment 7, the current operating condition, and many meet the desired minimum of 320 pounds.

To isolate possible adjusters, we use MINITAB to fit a *full model* with all possible main and interaction effects. Then we look at the Pareto plot of the effects in Figure 15.4. Only the effect for melt temperature is large. Also, because none of the interaction effects is large, we conclude that melt temperature is an adjuster.

We can use the main effect plot for melt temperature, given in Figure 15.5, to assess quantitatively the effect of changing melt temperature. Decreasing the melt temperature by 50°F increases the average seal strength by about 90 pounds.

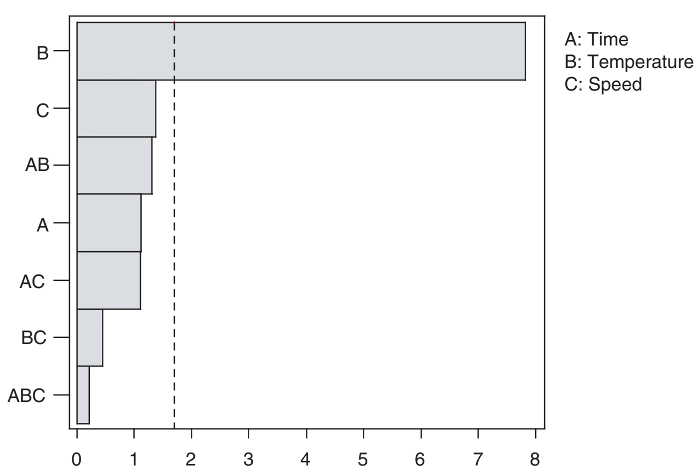


Figure 15.4 Pareto chart of effects for battery seal experiment.

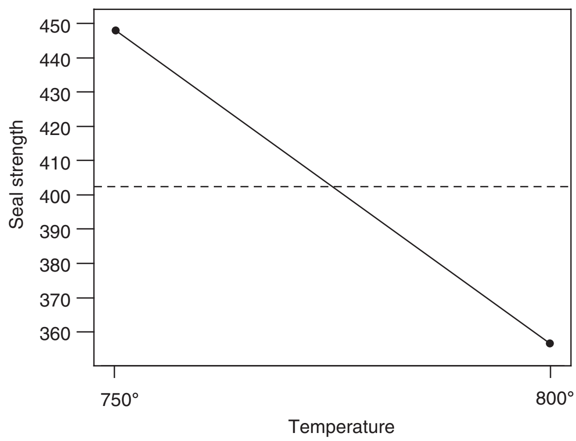


Figure 15.5 Main effect plot for melt temperature.

We do not know from this two-level experiment if the effect of the adjuster is linear. The team decided to use a melt temperature of 770ºF, a little over halfway between the low and high levels in the experiment. They did not want to decrease the melt temperature too much for fear of creating other problems related to material flow. To validate the solution, they changed the melt temperature to 770˚ for one shift and carefully monitored the process for leaks and negative side effects. The results were very promising. Since the cost of the change was negligible, the team implemented the new melt temperature setting and monitored the process for two months. The leak rate was reduced to 0.6%.

Engine Block Leaks (Cylinder Bore)

In Chapter 6, we introduced a problem of leaking engine blocks with three distinct failure modes. The team found that the dominant cause of cylinder bore leaks was the occurrence of *dip bumps* on the barrel cores used to create the bores in the block when it was cast. Each sand core is dipped in core wash, a paintlike substance, to improve the surface finish of the casting. Dip bumps occurred when the core wash ran before drying. The dip bump score for each core was the total number of bumps divided by four, the number of barrels per core. The team reformulated the problem to reduce the average dip bump score and established a new baseline. They did not precisely quantify the relationship between the occurrences of a cylinder bore leak and the dip bump score. The team assumed if they could reduce the average dip bump score to close to zero, then they would eliminate cylinder bore leakers.

The team used a single operator to define and count dip bumps. They did not assess this simple measurement system but assumed that it would add little variation. The team decided not to search for a dominant cause of dip bumps. They felt strongly that dip bumps were inevitable with the current core wash. They decided to experiment with five core washes from different suppliers. They hoped to find a core wash that would reduce the average dip bump score to close to zero.

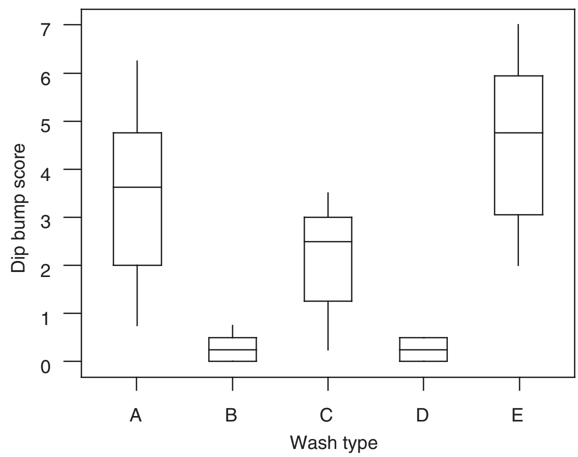


Figure 15.6 Box plots of dip bump score by core wash solution.

In the experiment, for each core wash, 50 cores were processed on each of two days. There were 10 runs, two replicates of each treatment, and 50 repeats for each run. Within each day, the order of treatments was randomized. The dip bump score for each of 500 cores was recorded. The data are given in the file *engine block leaks move center*. The cores from the experiment were recycled. Core wash *A* is the current wash.

The experimental results are summarized in the Figure 15.6. There were virtually no dip bumps with core washes *B* and *D*. The team recommended changing the process to use core wash *B* since it was cheaper than *D*.

During the validation of core wash *B*, the foundry scrapped a half shift worth of engine blocks and shut down the engine assembly line. This disastrous event occurred because, over the longer time frame used in the validation investigation, the new water-based core wash separated and failed to effectively coat the cores. The experiment to assess the different core washes was conducted over such a short time that the separation problem had not occurred. The team had not anticipated that continuous mixing was required with the new wash. Rather than changing the process to incorporate the necessary mixing, the team decided to look again at core wash *D* (very carefully this time), since it was similar in formulation to the original core wash. In the end, the foundry switched to core wash *D* and cylinder bore leaks were virtually eliminated.

Differential Carrier Shrink Defect

In the manufacture of a cast iron differential carrier casting, there was a problem with a shrink defect at a boss, a thick area of the casting. The defect was discovered at the drilling operation. The baseline defect rate was about 5%. We first discussed this problem in Chapter 3.

For the purposes of process improvement, a single operator assessed the shrink defect on a scale of 0 to 4, where 0 was the best. Scores of 3 or 4 correspond to a defective casting.

The problem goal was to reduce the average defect score. To establish a baseline and to simultaneously assess the measurement system, the team selected a sample of 200 castings from one day's production. The operator scored each casting twice. The data are in the file *differential carrier shrink defect baseline* and are summarized in Table 15.3.

Table 15.3 Boss shrink defect scores.

		Second measurement				
		0	**1**	**2**	**3**	**4**
	0	105	5	0	0	0
First measurement	**1**	3	40	1	0	0
	2	0	3	30	1	0
	3	0	0	0	10	0
	4	0	0	0	0	2

The operator was very consistent in scoring the castings, and the team judged the measurement system to be acceptable. The average score (of the first measurements) in this investigation was 0.75, and 12 out of the 200 castings were defective. The problem goal was to reduce the average score to close to 0, since then there would be almost no defective castings.

The team took the risky decision not to search for a dominant cause. Instead, they adopted the Move the Process Center approach and planned an experiment. The team selected nine candidates as potential adjusters. These candidates included three iron chemistry levels (denoted *A, B*, and *C*), iron temperature *(D)*, pouring time *(E)*, concentration of in-mold alloy *(F)*, two molding inputs from changing the dies that make the sand molds *(G* and *H*), and squeeze pressure *(J)*. The team selected the low and high levels of each candidate using engineering judgment. We code the two levels of each candidate –1 and +1.

To define a run of the experiment, the team planned to set the candidate levels and operate the casting process for 40 minutes. From past experience with the process, they expected to see castings with the shrink defect within that time period. Within each run, they planned to select 20 castings and score them for shrink. They also planned to randomize the order of the runs.

There are $2^9 = 512$ possible candidate combinations or treatments. It was not feasible to carry out an experiment with this many treatments so the team decided to use a *fractional factorial design* with only 16 runs (see the chapter suppliment[1]). The 16 treatment combinations must be carefully selected from the 512 possibilities. We use MINITAB to make the best selection (see Appendix F). We give the experimental design and results in Table 15.4. The team conducted the experiment without incident and stored the data in the file *differential carrier shrink defect move center*.

Table 15.4 Experimental design and results for differential carrier experiment.

Treatment	Order	Candidates									Shrink defect score freuency					Average Score
		A	B	C	D	E	F	G	H	J	0	1	2	3	4	
1	3	+1	–1	–1	+1	+1	+1	–1	–1	+1	7	5	6	2	0	1.15
2	4	–1	+1	+1	+1	–1	+1	–1	–1	–1	5	7	6	2	0	1.25
3	2	–1	+1	–1	–1	+1	+1	–1	+1	–1	1	7	8	3	1	1.80
4	5	–1	–1	+1	–1	+1	+1	+1	–1	–1	6	10	4	0	0	0.90
5	6	–1	–1	–1	–1	–1	–1	–1	–1	+1	4	7	7	2	0	1.35
6	10	–1	–1	+1	+1	+1	–1	–1	+1	+1	8	11	1	0	0	0.65
7	15	+1	+1	+1	–1	+1	–1	–1	–1	–1	0	5	6	9	0	2.20
8	16	+1	+1	+1	+1	+1	+1	+1	+1	+1	4	8	7	1	0	1.25
9	12	+1	+1	–1	+1	–1	–1	–1	+1	–1	1	9	7	3	0	1.60
10	7	–1	+1	–1	+1	+1	–1	+1	–1	+1	7	10	3	0	0	0.80
11	8	–1	–1	–1	+1	–1	+1	+1	+1	–1	12	8	0	0	0	0.40
12	9	–1	+1	+1	–1	–1	–1	+1	+1	+1	4	7	7	2	0	1.35
13	14	+1	–1	–1	–1	+1	–1	+1	+1	–1	4	7	5	3	1	1.50
14	13	+1	+1	–1	–1	–1	+1	+1	–1	+1	1	8	6	5	0	1.75
15	1	+1	–1	+1	–1	–1	+1	–1	+1	+1	1	8	6	4	1	1.80
16	11	+1	–1	+1	+1	–1	–1	+1	–1	–1	6	9	5	0	0	0.95

In Table 15.4, we give the frequency of scores and the average score for the 20 castings in each run. We are looking for a treatment that gives a low average score. Here, due to the discreteness of the scores, a plot of individual casting scores versus treatment is not very informative. Instead, we look at the *performance measure* defined as the average score over each run. In general, a performance measure is a statistic calculated over all the repeats within each run to assess the performance of the process for that treatment. There are promising treatments, such as number 11, where the average score is small.

From Table 15.4, we can see one important property of the selected design. For each candidate, 8 of the 16 runs have the level –1 and the other 8 have the level +1. Due to this balance, we can assess the main effect of any candidate by comparing the averages of these two sets of runs, as we did in Chapter 13.

The price we pay for using a fractional design (rather than a full factorial design) is that we lose information about interactions. We say that certain effects are *confounded*, because we cannot estimate them separately. MINITAB produces a list of effects, called the alias structure, that are confounded when we use MINITAB to generate the design. For the differential carrier experiment, the main effects and two input interactions that are confounded are:

```
Alias Structure (up to order 2)

I
A + F*J
B + G*J
C + H*J
D + E*J
E + D*J
F + A*J
G + B*J
J + A*F + B*G + C*H + D*E
A*B + C*E + D*H + F*G
A*C + B*E + D*G + F*H
A*D + B*H + C*G + E*F
A*E + B*C + D*F + G*H
A*G + B*F + C*D + E*H
A*H + B*D + C*F + E*G
```

In the list, the main effects are labeled by a single letter and the interactions by pairs such as $A*B$. To determine the confounding, we see from the second line, for example, that the main effect of input A is confounded with the two-input interaction between F and J. As a consequence, if the average difference between the eight runs with A at level –1 and A at level +1 is large, the difference may be due to simultaneously changing the levels of inputs F and J or to changing the levels of input A. We cannot separate these two possibilities with the data from the experiment. Because of this confounding, we need to confirm any promising adjuster found in the experiment. Third- and higher-order interactions are also confounded with the given effects. MINITAB will provide the complete confounding (aliasing) structure if desired. We call this experimental plan a *resolution* III design because some main effects are confounded with two-input interactions but not with other main effects. In a resolution IV design, some main effects are confounded with three-input interactions but no two-input interactions; in a resolution V design, some main effects are confounded with four-input interactions but no three-input interactions. Since we assume that the higher the order of an interaction, the more likely it is to be negligible, we want a design with as high a resolution as possible. For a given number of candidates, MINITAB selects, by default, one of the fractional factorial designs with the highest possible resolution.

The Pareto chart of the effects for the full model is given in Figure 15.7. We use the first effect in the list of confounded effects as the label for the string of confounded effects so that, for example, *D* corresponds to the combined effects of input *D*, the two-input interaction *E**J*, and other three-input and higher interactions not shown in the confounding list. We see in Figure 15.7 that effects for inputs *A, B, D,* and *G* are large relative to the others. If we assume that the two-input and higher-order interactions are negligible, then inputs *A, B, D,* and *G* are adjusters.

We give the main effects plots for inputs *A, B, D,* and *G* in Figure 15.8. Since a lower average score is better, the desirable levels of the four inputs are –1 for inputs *A* and *B* and +1 for inputs *D* and *G*. The team decided to investigate further the consequences of changing the levels of inputs *A* and *B*, the two iron chemistries, to the low level used in the experiment.

To confirm the findings, the team set inputs *A* and *B* to their new levels, kept all other inputs at their current levels, and operated the process for four hours. Of the 3108 castings produced, only 12 were scrapped for shrink defect at machining. This scrap rate of 0.38% was a substantial improvement over the original rate of 5%. There were no side effects, so the team adopted the proposed process change.

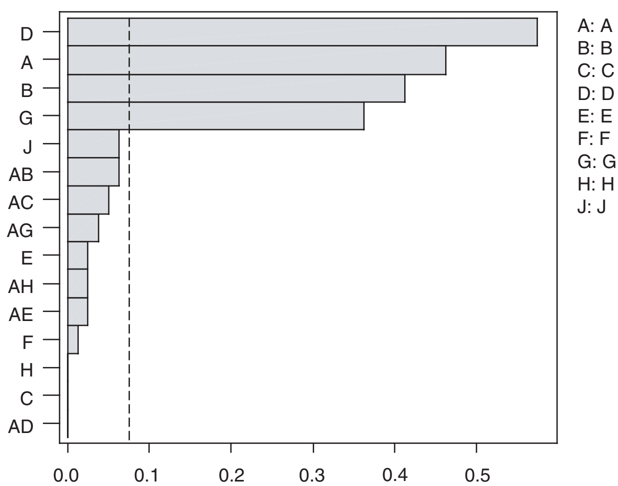


Figure 15.7 Pareto chart of the effects for piston shrink defect experiment.

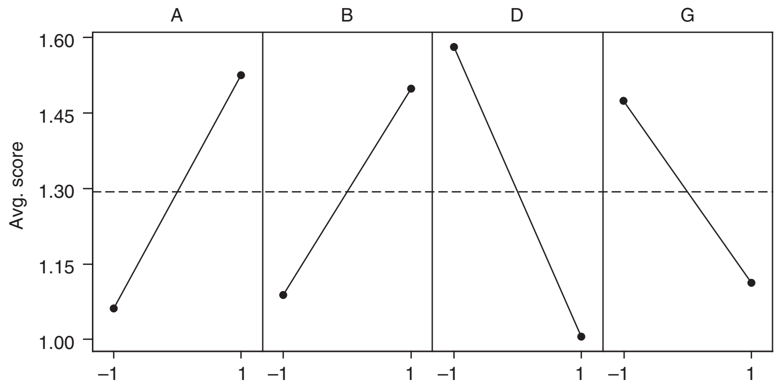


Figure 15.8 Main effects of significant inputs for shrink defect experiment.

Move Process Center Experiment Summary

Use engineering judgment to select candidates as potential adjusters.

Question

Does changing any of the candidates move the process center substantially?

Plan

- Select a study population where we expect to see the full extent of variation.
- Choose two levels for each candidate as extreme as is feasible.
- Define a run for the experiment. If the output is binary, a run must contain many repeats.
- Choose a design with at least eight runs to specify the treatments. For:
  –Three or fewer candidates, use a full factorial design
  –Four or more candidates, select a fractional factorial design with at least resolution III
- Spread the runs across the study population. Randomize the order as much as is feasible.
- Make everyone potentially impacted aware of the plan.

Data

- Carry out the experiment. For a:
  –Continuous output, record the output value, the levels of the candidates, treatment number, and run order, one row for each repeat
  –Binary output, record the proportion of defectives, the levels of the candidates, treatment number, and run order for each run

Analysis

- Plot the output values for each treatment. Look for promising treatments.
- If there are two or more candidates, use a full model and create a Pareto plot of the effects.
- Check for possible confounding in the important effects if the design is a fractional factorial.

Conclusion

- A candidate with a large effect is an adjuster.
- For fractional factorial designs, verify an adjuster, if necessary, to break confusion due to confounded effects.

15.2 ASSESSING AND PLANNING A PROCESS CENTER ADJUSTMENT

To successfully implement Move the Process Center we must:

- Identify one or more adjusters, fixed inputs that we can change to move the process center.
- Identify the best level(s) for the adjuster(s).
- Check that the adjustment does not produce substantive negative side effects.
- Estimate the costs of changing the adjuster and ongoing operating costs.
- Estimate the benefit of the adjustment.

If we can accomplish all of these tasks and the benefits outweigh the costs, we adopt the Move the Process Center approach and proceed to the validation stage of the algorithm.

There is some risk that no adjuster will be found. Experiments, such as that used in the differential carrier shrink defect example, can be complex and expensive. In planning such experiments, the candidates and their levels should be selected with great care. There is little value in including an adjuster that is very expensive to change or, if changed, increases operating costs substantially.

Once an adjuster has been identified, further experimentation may be required to determine the appropriate level of the adjuster to move the process center to the desired value. In the battery seal failure experiment, melt temperature was found to be an adjuster. The average seal strength was 356 pounds at 800° and 447 pounds at 750°. A change of 50° in melt temperature produces a change of about 90 pounds in average seal strength. Since the goal was to increase average seal strength by about 60 pounds, the team selected 770° as the new set point for the melt temperature. They assumed a roughly linear relationship between the melt temperature and the seal strength. Alternately, the team could have conducted a second experiment with several levels of melt temperature between 750° and 800° in order to determine the relationship more precisely. Precisely quantifying the adjuster/output relationship is more important if we make repeated adjustments of different sizes, as in feedforward or feedback control.

When changing the process center we need to watch carefully for negative side effects. In the core strength example introduced in Chapter 1, the team found they could eliminate core breakage by increasing core strength. They discovered an adjuster, the amount of resin in the sand mix, which they used to increase the average strength. However, the stronger cores led to more casting defects and the approach was abandoned. We may be able to avoid undesirable side effects by using two or more adjusters simultaneously rather than one adjuster as in the battery seal failure and engine block leaks examples. This is a good reason for searching for several adjusters simultaneously.

For off-target process center problems, we can see the expected benefit of changing the process center using the histogram from the baseline investigation. Based on the problem goal and baseline, we know how far we want to move the process center and the expected benefits of such a move. We also need to assess the costs of changing the adjuster and the operating costs at the new level.

In some problems the two approaches, Move the Process Center and Make the Process Robust, are identical. The differential carrier shrink defect problem is a good example. Whatever the cause of the shrink defect, the change in the fixed chemistry inputs made the process more robust to the cause.

In multistream processes, if we find that the process centers vary from stream to stream, we can reduce the overall variation by moving the centers of each stream to a common target, as shown in Figure 15.9. In the left panel, we show box plots of the output by stream and overall. In the right panel, we show the same plots after we (roughly) align the substream centers. Aligning the substream centers results in a significant reduction in the overall output variation.

In some circumstances, an output center may drift over time. In that case, a one-time adjustment of the process center will not solve the problem over the long term. We may need to use another variation reduction approach, such as feedback control.

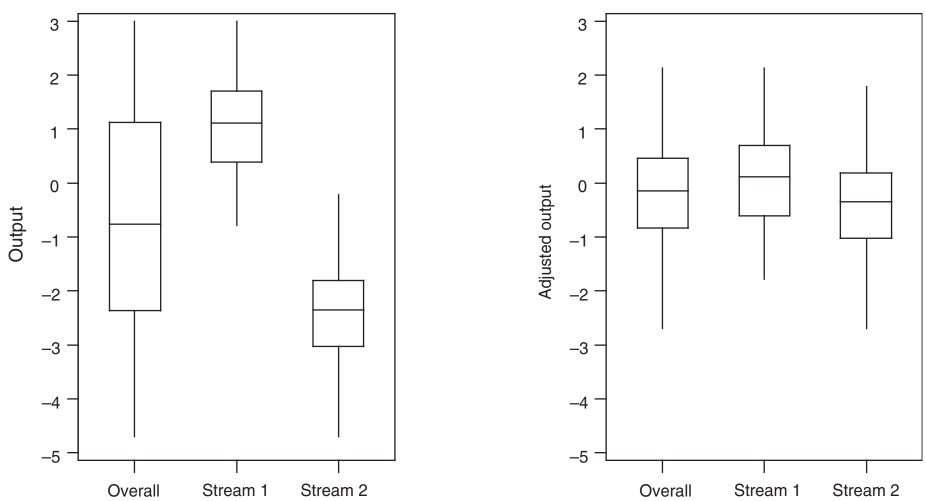


Figure 15.9 The effect of aligning substream centers on output variation.

Key Points

- To move the process center, we need to find one or more adjusters, normally fixed inputs. We change the settings of the adjusters to move the output center.

- We use an experimental plan, a full or fractional factorial design, to investigate one or more fixed inputs (candidates) as possible adjusters.
- For processes with multiple streams, we can align the centers of the streams with separate adjustments and hence reduce the overall variation in the process.

Endnote (see the Chapter 15 Supplement on the CD-ROM)

1. Fractional factorial designs are powerful experimental methods for investigating the effects of a large number of inputs simultaneously. In the supplement, we provide an introduction to these designs and references to further work. We will use these designs to determine the knowledge required to implement several of the variation reduction approaches.

Exercises are included on the accompanying CD-ROM

16

Desensitizing a Process to Variation in a Dominant Cause

Efficiency is doing things right; effectiveness is doing right things.

—Peter F. Drucker

The goal of desensitization is to find and change fixed inputs that flatten the relationship between the output characteristic and the dominant cause (see Figure 16.1). That is, we find an interaction that we exploit to desensitize the process output to changes in the dominant cause.[1] To explore desensitization, we choose a number of fixed inputs to investigate, based on knowledge of the dominant cause and the process. We use an experimental plan to determine if these *candidates* and their new levels will make the process less sensitive to variation in the dominant cause.

With this approach we do not address the dominant cause directly. We continue to live with its variation. By changing the relationship between the cause and the output, we reduce the effect of the cause.

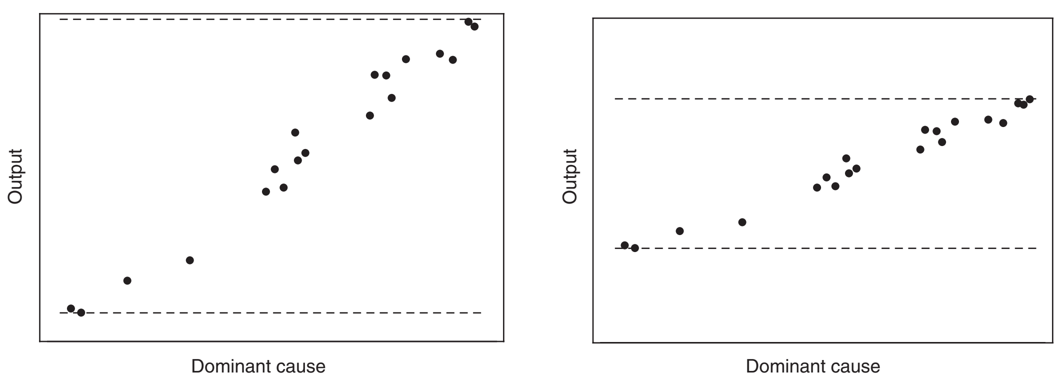


Figure 16.1 Original (left) and new (right) relationship between the dominant cause and output characteristic.

The only requirement for desensitization is to find new settings of the candidates that make the output less sensitive to variation in the dominant cause.

The costs of desensitization include:

- An experiment to find the new process settings that make the process less sensitive to variation in the dominant cause
- A one-time change to the process settings
- The ongoing operation of the process with the new settings

There is no information about whether this approach will be feasible until the experimental investigation is complete. This is a drawback, since the cost of the experimentation may be high and the returns uncertain. The benefit can be assessed using the relationship between the cause and the output. If we could totally eliminate the effect of the dominant cause, the maximum benefit is given by the residual variation in the output due to all other causes.

16.1 EXAMPLES OF DESENSITIZATION

We illustrate the approach with a series of examples in which the complexity of the experimental plan increases.

Engine Block Porosity

We discussed the problem formulation and the search for the dominant cause of excess porosity in a cast-iron engine block in chapters 6, 10, and 13. The team identified low pouring temperature as the dominant cause. During planned and unplanned work stoppages, iron that remained in the pouring ladle cooled since there was no external heat source. In normal production, this was not a problem because the pouring ladle was frequently replenished with fresh hot iron. The team could not address this cause directly. Eliminating or staggering breaks was not possible due to employee resistance, and adding temperature control to the pouring ladles was too costly. They decided to explore desensitization as a possible solution.

The first step was to choose the candidates, fixed inputs that can be changed and might desensitize the process to variation in pouring temperature. Porosity occurs when gases produced during pouring are trapped in the casting. If the pouring temperature is too low, the iron may harden before all the gas can escape. The team selected core wash and core sand composition as candidates. Cores, set in the mold to create spaces in the casting, are dipped in core wash to produce a better surface finish. The cores are made from resins and sand. Both the core wash and the resins generate gas during the pouring process.

The team decided to investigate one new core wash and a single reformulation of the core sand mixture in a factorial experiment. Each candidate had two levels, regular and alternative. There were four treatments as shown in Table 16.1.

To assess the sensitivity of the process to pouring temperature for each treatment, the team used the knowledge that the pouring temperature varied over its full range around the lunch break. For each treatment, three engine blocks were cast just before lunch (high pouring

Table 16.1 Treatments for the engine block porosity desensitization experiment.

Treatment	Sand	Core wash
1	Regular	Regular
2	Regular	Alternative
3	Alternative	Regular
4	Alternative	Alternative

temperature) and another three blocks just after lunch (low pouring temperature). The experiment was replicated over three days so that a total of 72 blocks were measured for porosity. The 12 blocks before and after lunch were divided into three groups of four. Within each group, the team randomized the order of the four treatments. Since temperature increased steadily after lunch, each treatment saw roughly the same range of temperatures. The team managed the randomization easily because the treatments only affected the cores. The data are given in the file *engine block porosity desensitization*.

We start the analysis with a plot of porosity versus treatment number with different plotting symbols for the high and low temperatures, as shown in Figure 16.2. For any treatment, we can see if there is substantially less porosity than in the current process, here given by Treatment 1. All three new treatments are a large improvement over the existing process with smaller average porosity and less variation as the pouring temperature changes.

The team decided to investigate the new core wash further to ensure that there were no side effects such as an increase in other casting defects. There was no cost to this change. The alternative core wash was eventually adopted and the scrap rate due to porosity decreased from about 4% to less than 1%.

In this example, the team designed a full factorial experiment and drew their conclusions from a simple plot. We give three more examples to illustrate some complications that can arise in a desensitization experiment.

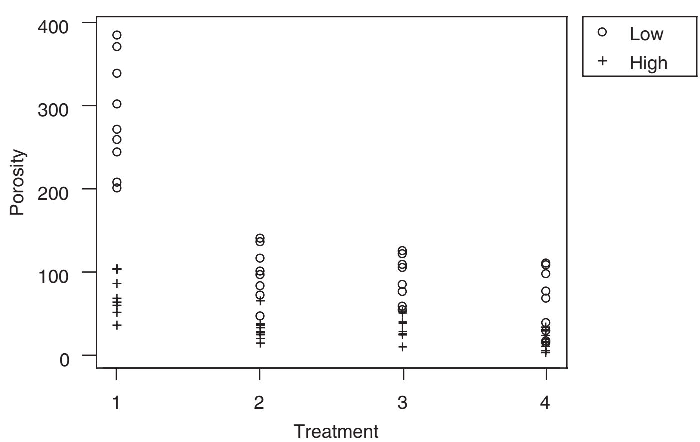


Figure 16.2 Porosity by treatment for high and low pouring temperatures.

Oil Pan Scrap

The baseline oil pan scrap rate was 8%. Applying the method of elimination, the team discovered that a dominant cause of scrap was the amount of lubricant applied to the blanks prior to the draw press. A roll coater with two steel cylinders applied the lubricant. There were no controls on the roll coater and no ongoing measurement of the lubricant amount.

During a verification experiment, the team used a visual five-point score for the amount of oil applied. They specially prepared 80 blanks with lubricant amounts at each of the five levels. The team could control the dominant cause during the experiment but not in normal production. To verify the cause, the team stamped the 80 blanks for each oil level. We see from Figure 16.3 that the scrap rate increased markedly as the lubricant score increased.

There was no obvious cost-effective way to control the amount of lubricant applied to the blanks, so the team decided to try to desensitize the process. There was a lot of uncertainty about whether this approach would be effective, but they decided that it was worth the cost of the investigation in the hopes of avoiding a solution with high capital cost.

Based on knowledge of the cause, the team chose three candidates for the desensitization experiment: lubricant supplier, die temperature, and binder force. After consultation, three lubricant suppliers (the current supplier is labeled *A* and two other suppliers *B* and *C*) provided the lubricant they judged the best suited for the application. Supplier *A* suggested the lubricant in current use. The team selected the two levels for the other candidates using engineering judgment. They decided to use only the two extreme levels 1 and 5 for the amount of oil on the blanks, the dominant cause.

The team defined a run to be the consecutive stamping of 80 blanks as in the verification experiment. Since the output was binary, each run required sufficient repeats so that some defectives were produced. The team used the scrap rate for each run as a performance measure. The experiment had 24 runs. There were 12 treatments defined by the levels of the three candidates and two runs per treatment, one for each level of the lubricant score. With this plan, each treatment would see the full range of variation in the dominant cause.

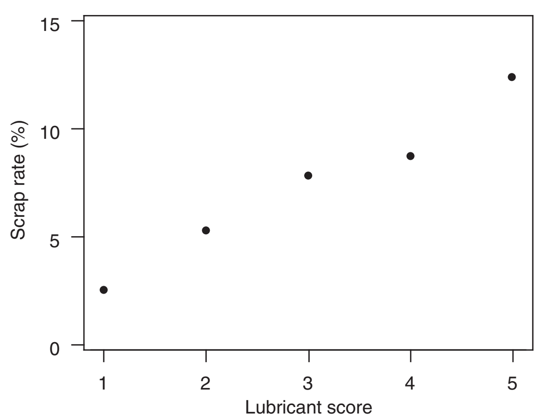


Figure 16.3 Scrap rate by lubricant score.

Table 16.2 Candidate levels and scrap rates for low and high level of lubricant amount.

					Scrap rate for	
Treatment	**Run order**	**Lubricant supplier**	**Die temperature**	**Binder force**	**Lubricant Score 1**	**Lubricant Score 5**
1	3	A	Low	Low	0.07	0.15
2	5	A	Low	High	0.14	0.16
3	10	A	High	Low	0.06	0.11
4	7	A	High	High	0.15	0.24
5	4	B	Low	Low	0.04	0.01
6	1	B	Low	High	0.10	0.11
7	8	B	High	Low	0.03	0.05
8	12	B	High	High	0.09	0.14
9	6	C	Low	Low	0.03	0.05
10	2	C	Low	High	0.04	0.08
11	11	C	High	Low	0.04	0.04
12	9	C	High	High	0.03	0.07

Since die temperature was difficult to change, all runs with the low temperature were conducted first. Within the low-temperature runs, the six treatments were applied in random order. Once they set the levels of the candidates, the team stamped the two sets of 80 blanks, one set with oil level 5 and one with oil level 1.

We give the experimental plan, including the run order, and the results in Table 16.2 and in the file *oil pan scrap desensitization*.

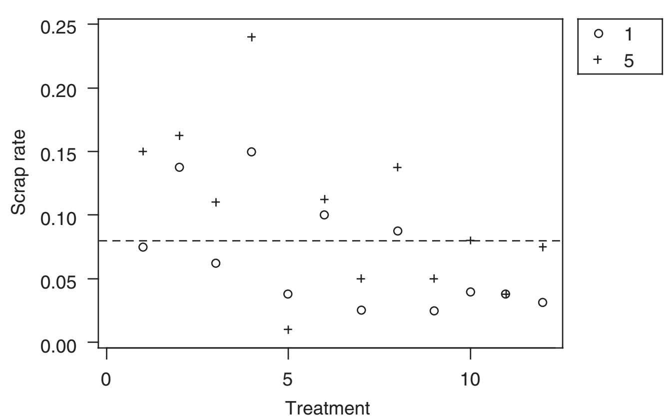


Figure 16.4 Scrap rate for low and high level of lubricant amount versus treatment (dashed horizontal line gives the baseline scrap rate).

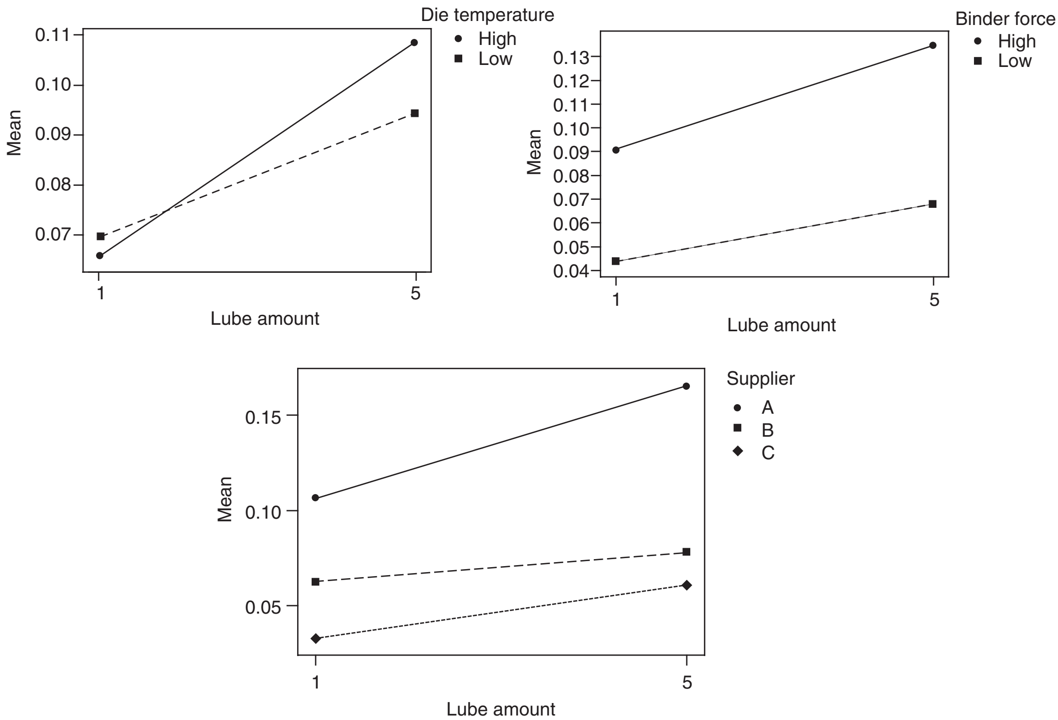


Figure 16.5 Interaction plots for oil pan scrap desensitization experiment.

We start the analysis by plotting the scrap rate by treatment with different plotting symbols for the two lubricant scores. In Figure 16.4, we see that there are promising treatments with low scrap rates for both levels of the dominant cause.

We can look at the effects in more detail using the interaction plots between the candidates and the lubricant score. In Figure 16.5, we see that the scrap rate is:

- Less sensitive to changes in the lubricant amount if the die temperature is low
- Substantially lower with binder force at the low level for both lubricant scores
- Much lower and less sensitive to changes in lubricant amount for lubricant *C*

Since lubricant *C* looks so promising, we next consider the interactions between binder force and die temperature by lubricant amount for supplier *C* only. We subset the data set in MINITAB (see Appendix A). From the interaction plots in Figure 16.6, we conclude that low binder force is the preferred setting for lubricant *C* to reduce the scrap rate. Die temperature has little effect.

During the investigation, the team measured other important outputs to make sure that changing the process settings would not result in other problems. They found that lubricant *C* was also better than the original lubricant in terms of steel flow at the deep end of the pan. The team decided to change to the lubricant from supplier C. After the change, the long run scrap rate was reduced to about 3%. There was a small increase in cost that was far outweighed by the savings in scrap costs.

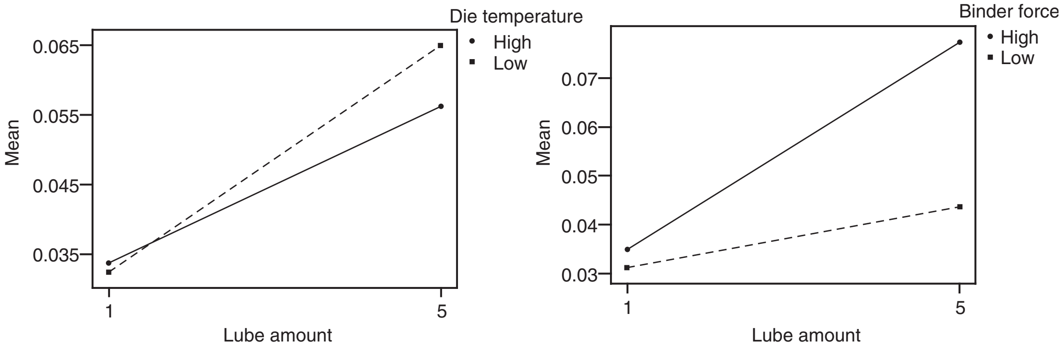


Figure 16.6 Interaction plots of die temperature and binder force by lubricant amount (supplier *C* only).

Refrigerator Frost Buildup

The manufacturer of a frost-free refrigerator, designed for temperate climates, expanded its market to several tropical countries. Almost immediately, there were many customer complaints about the buildup of frost inside the refrigerator. These complaints had been nonexistent in the traditional market. Management feared that the opportunity for market expansion would be lost. They formed a team to reduce the number of complaints. We discussed this example previously in chapters 1, 3, and 8.

As a first step, the team replaced field measurements by the temperature of the refrigerator cooling plate. They verified that high temperature on the cooling plate led to a buildup of frost under certain conditions. The team set a goal to keep the cooling plate temperature less than 1°C. There was no baseline investigation for either the original problem or for the problem as reformulated in terms of cooling plate temperature. Either investigation would have been very expensive.

Because the problem did not occur in temperate climates, the team knew that the dominant cause must relate to differences in usage or ambient conditions between the tropical and the traditional market. Through further discussion and interviews with field representatives, they decided that the dominant cause of temperature variation on the cooling plate was ambient temperature and relative humidity in combination with the frequency of door openings.

The team could not address the dominant cause directly since it was not under the control of the manufacturer. They decided to pursue desensitization. The goal was to find a refrigerator design that would maintain a consistent low temperature on the cooling plate under varying conditions defined by ambient temperature, relative humidity, and number of door openings per hour.

The team considered four fixed inputs, D1 to D4, related to the refrigerator design. They selected two levels, *original* and *new*, for each candidate, based on engineering knowledge. There were 16 possible prototypes (treatments) using the levels of the four candidates. Due to the high cost of prototypes, the team decided to use a fractional factorial design and build only 8 of the 16 possible combinations. The team used MINITAB to select the design. See the supplement to Chapter 15 for more detail on fractional factorial designs. We show the selected treatments in Table 16.3.

Table 16.3 Eight-run fractional factorial refrigerator desensitization experiment.

Treatment	D1	D2	D3	D4
1	New	New	New	New
2	New	New	Original	Original
3	New	Original	New	Original
4	New	Original	Original	New
5	Original	New	New	Original
6	Original	New	Original	New
7	Original	Original	New	New
8	Original	Original	Original	Original

In this resolution III design, the main effect of input D4 is confounded with the three-input interaction of D1, D2, and D3. Thus pairs of two input interactions such as D1*D2 and D3*D4 are also confounded.

The team set up the testing laboratory with two sets of environmental conditions:

Normal — Conditions to mimic use in temperate climate

Extreme — Conditions to match the worst anticipated usage conditions with high temperature, high humidity, and frequent door openings

For each prototype listed in Table 16.3, the team planned two runs in the laboratory, one at normal and one at extreme environmental conditions. We call this a *crossed design* since we use both levels of the dominant cause for each of the treatments. There are 16 combinations of the 5 inputs (D1 to D4 and environment) out of the 32 possible treatments. We use MINITAB to determine the confounding structure for this fractional design.[2] We assume third- and higher-order interactions are negligible, so we attribute any large differences to main effects and two-input interactions. We reproduce the confounding structure with any effects involving three or more of the fixed inputs (candidates) erased.

```
I
D1
D2
D3
D4
environment
D1*D2 + D3*D4
D1*D3 + D2*D4
D1*D4 + D2*D3
```

```
D1*environment
D2*environment
D3*environment
D4*environment
D1*D2*environment + D3*D4*environment
D1*D3*environment + D2*D4*environment
D1*D4*environment + D2*D3*environment
```

The purpose of the desensitization experiment is to isolate useful interactions between the dominant cause (environment) and one of the candidates (D1 to D4). From the list of confounded effects, we see that the interactions between the four candidates D1 to D4 and the environment are in separate lines and, hence, can be estimated separately. These interactions are the key to desensitizing the cooling plate temperature to changes in the usage conditions. On the other hand, we cannot separate the joint effects D1*D2 and D3*D4 with the environment.

After 30 minutes of cycling, the temperature of the cooling plate was measured. Randomization of the run order was not feasible for time and cost reasons. All eight refrigerators were placed in a special testing room and subjected to one of the environmental conditions simultaneously. This lack of randomization was not a major risk here since all refrigerators are subject to the same environmental conditions. The complete experimental plan and the data are given in Table 16.4. The data are stored in the file *refrigerator frost buildup desensitization* with one row for each temperature measurement and columns indicating the values of D1–D4 and the environmental conditions.

We start the analysis by plotting the cooling plate temperature versus treatment with a separate plotting symbol for the two levels of environment. From Figure 16.7, we see that there are several promising treatments, especially treatment 6.

Table 16.4 Plan and data for refrigerator frost buildup desensitization experiment.

					Environmental conditions	
Treatment	**D1**	**D2**	**D3**	**D4**	**Low (normal)**	**High (extreme)**
1	New	New	New	New	0.7	2.1
2	New	New	Original	Original	2.9	4.8
3	New	Original	New	Original	2.4	9.6
4	New	Original	Original	New	3.8	5.9
5	Original	New	New	Original	1.9	4.0
6	Original	New	Original	New	–0.2	0.1
7	Original	Original	New	New	–0.1	3.5
8	Original	Original	Original	Original	0.2	7.2

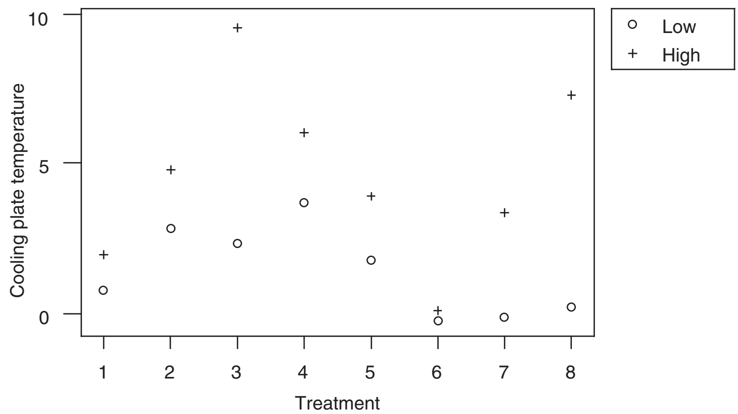


Figure 16.7 Temperature on cooling plate by treatment (different symbols represent output values for different environmental conditions).

In Figure 16.8, we look at the interaction plots for each candidate versus the environmental cause. We hope to find an interaction that gives consistently low cooling plate temperatures for both levels of the environment, as seen in treatment 6.

Using Figure 16.8, the team concluded that they should change D2 and D4 to their new levels. However, changing to the new level of D4 added significant cost. As a result, the

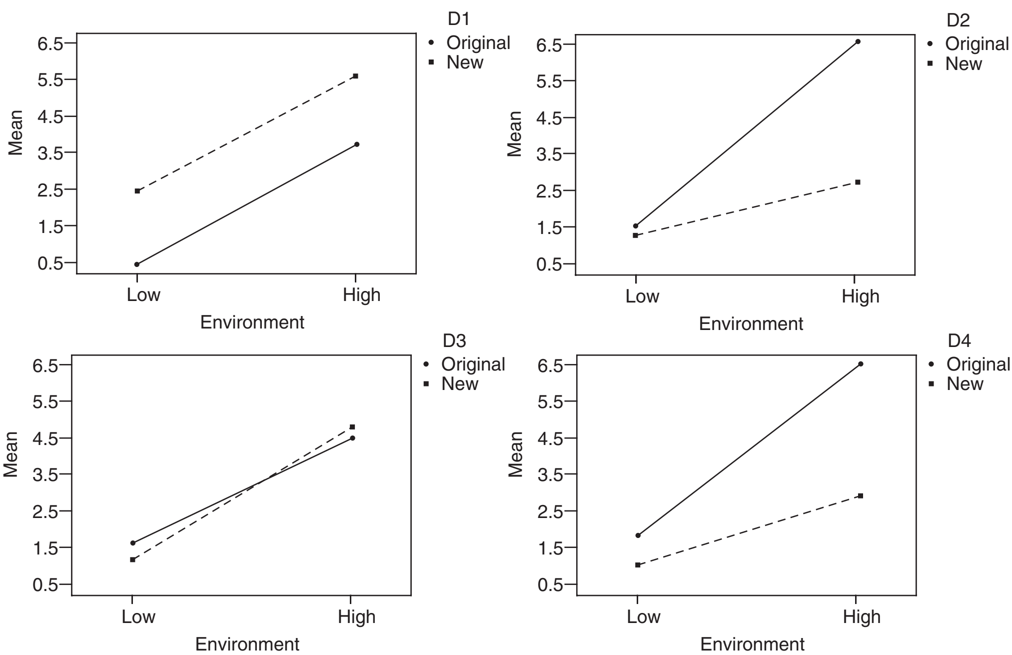


Figure 16.8 Interaction plots of candidates and the environment cause output in cooling plate temperature.

team decided to investigate how much the cooling plate temperature variation could be reduced if only D2 was changed to its new level. This treatment was not included in the original fractional design. When they produced a prototype with only D2 changed and subjected it to both levels of the environmental cause, they found that changing this input alone did not sufficiently reduce the cooling plate temperature and its variation as the environmental input was changed.

Despite the added cost, the team recommended adopting a refrigerator design with both D2 and D4 changed to their new levels. After the change, the frequency of complaints about frost buildup from the tropical market was substantially reduced.

Eddy Current

In a process that produced castings later machined into brake rotors, there was 100% inspection of hardness using a measurement system based on eddy currents. This system was fast and nondestructive. Despite the 100% inspection, there were frequent complaints from the customer, a machining operation, about castings out-of-specification with respect to hardness. The customer used the Brinell method to measure hardness, a standard procedure that is partially destructive and time consuming, not suitable for 100% inspection.

There was a high reject rate at the 100% eddy current inspection station. The operators measured all rejected castings a second time. The plant shipped castings that passed at least one test and scrapped those that failed both tests.

Management assigned a team to reduce scrap costs and customer complaints by reducing casting hardness variation. To establish a baseline, each day for one week, the team collected a haphazard sample of 100 first-time eddy current measurements for a total of 500 hardness measurements. The data are available in the file *eddy current baseline* and are plotted by day in Figure 16.9. The hardness specification is 4.3 to 4.7. The baseline standard

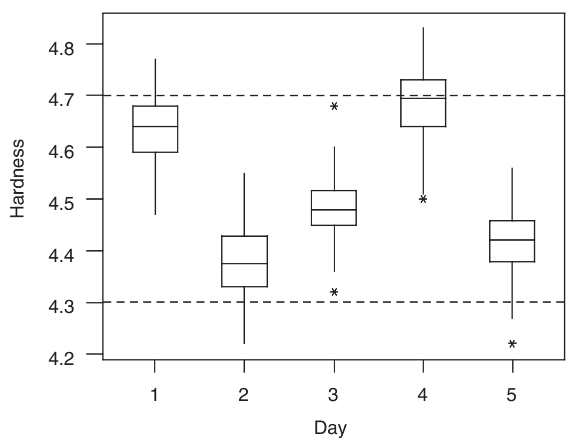


Figure 16.9 Box plots of eddy current hardness measurements by day (dashed horizontal lines give the specification limits).

deviation is 0.1365 and the full extent of hardness variation is 4.52 ± 3*0.14 (average ±3 standard deviation), or about 4.1 to 4.9.

The team assessed the measurement system using a standard gage R&R procedure as described in the supplement to Chapter 7. The measurement system standard deviation was 0.021, so it appeared the measurement system was adequate to proceed to the next stage of the algorithm.

Because of the past complaints from their customer, the team also carried out an investigation to compare the eddy current and Brinell hardness measurement systems. They selected 30 castings with widely varying eddy current measurements, then tagged and shipped these castings to the customer. The customer measured the Brinell hardness for each casting. The data are plotted in Figure 16.10, and given in the file *eddy current Brinell measurement*.

The real problem is now apparent. There is poor correlation between the two measurement systems. Although the earlier gage R&R investigation showed that the eddy current system was repeatable, for any set of castings with the same Brinell hardness, the eddy current system gave widely varying values. The team reformulated the problem to improve the correlation between the two measurement systems. Said in another way, the team decided to reduce the variation in eddy current measurements among castings with the same Brinell hardness.

The team discovered that day-to-day fluctuations in iron chemistry and the level of dirt on the castings were dominant causes of the variation of the eddy current measurement of castings with the same Brinell hardness. They did not look for a more specific cause in the chemistry family. Castings were cleaned by shot blasting. The level of cleaning varied since the shot blast machine did not run using a first-in, first-out protocol. The cleaning times ranged from 5 to 19 minutes.

In regular production, iron chemistry was expensive to control, and it was difficult to remove all the dirt from the casting before measurement. As a result, the team decided to look for a way to run the eddy current measurement system that was less sensitive to varying levels of dirt and iron chemistry.

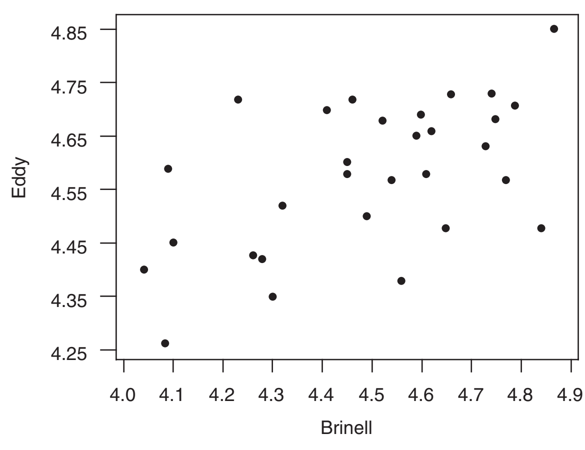


Figure 16.10 Scatter plot of eddy current versus Brinell hardness measurements.

Table 16.5 Candidate levels for eddy current measurement experiment.

Candidate input	Low level	High level
Frequency	200	350
Probe temperature	35	65
Gain	30	40

For the desensitization experiment, the team chose three candidates corresponding to settings on the eddy current system. For each candidate they selected high and low levels on either side of the current settings. The team planned a full factorial with eight treatments using the levels given in Table 16.5.

To ensure that each treatment was assessed under the extreme levels of the two dominant causes, the team selected 16 castings from each of two days' production. Each casting was cleaned for 5 minutes, measured with the eddy current system, cleaned for another 14 minutes, and then measured again. The team measured every casting using all eight treatments at both points in the cleaning cycle. Finally they measured the Brinell hardness of each casting at two different positions on the casting. The team knew that the Brinell hardness measurements had little measurement variation relative to the variation in the eddy current machine. In the analysis, they used the average of the two Brinell values as the true hardness of the casting.

The team hoped that the two days would represent the day-to-day variation in iron chemistry. This was a risky decision. With only two days, the full day-to-day variation in chemistry might not be captured. If the team had found the specific cause within the chemistry family, they could have more easily ensured that each treatment was subject to the full range of the dominant cause. We give the data from the experiment in the file *eddy current desensitization*. There are 16 eddy current measurements and the average Brinell reading for each of 32 castings.

The analysis of this experiment is complicated. The goal in the analysis is to find a combination of levels for the candidates that yields a strong relationship between the Brinell hardness values and the values given by the eddy current system, consistently over the four combinations of the levels of the dominant causes. We do not worry about measurement bias, since if the bias is consistent it can be easily removed.

In the analysis, for each of the eight treatments we plot the hardness as measured by the eddy current system versus the (average) Brinell hardness for all four combinations of the dominant causes. For example, in Figure 16.11, we show the plot for the treatment frequency = 200, temperature = 35, and gain = 30. This plot is typical of all the eight treatments. We see from the Figure 16.11 that there is a weak relationship between the two methods of measurement. For a given Brinell hardness, there is large variation in the eddy current measurements.

The experiment failed to find settings of the candidates that make the eddy current measurement system less sensitive to changes in chemistry and dirt. In all cases, there was no relationship that could be used to predict the Brinell hardness from the eddy current measurements. Had one or more of the treatments shown promise, the team could have conducted a more formal analysis as discussed in the chapter supplement.[3]

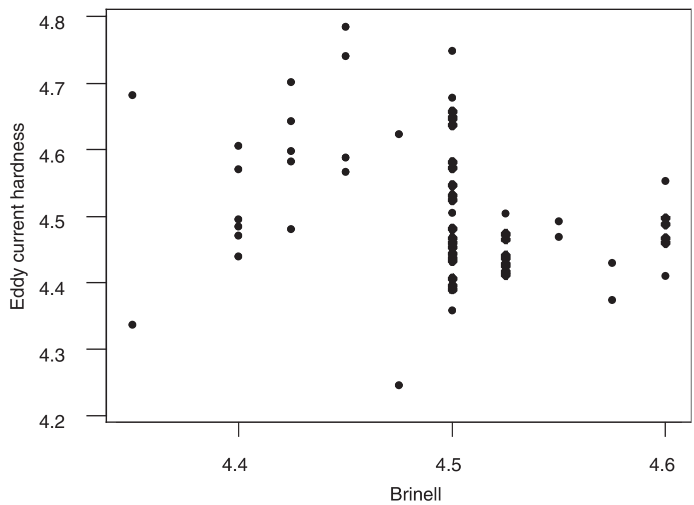


Figure 16.11 Plot of eddy current versus Brinell hardness for freq. = 200, temp. = 35, and gain = 30.

In the end, the team concluded that the experiment was not a failure. They had shown that the eddy current system did not work well and could not be easily improved. There was little value in the 100% inspection. By removing the eddy current system, the foundry was forced to concentrate on reducing variation in the process instead of relying on imperfect 100% inspection.

Desensitization Experiment Summary

Select candidates and levels based on knowledge of the process and the dominant cause.

Question

Which combination, if any, of the candidate levels results in a process less sensitive to variation in the dominant cause?

Plan

- Select a design for the candidates. For:
  –Three or fewer candidates, use a full factorial design
  –Four or more candidates, use a fractional factorial design with resolution III or higher
- Define a treatment number for each combination of the candidates.
- Select two levels for the dominant cause at the extremes of its normal range.
- Use a crossed design where, for each treatment, there are runs for both levels of the dominant cause.
- Define a run of the experiment including treatment assignment and the number of repeats.

- Randomize the order of the runs as much as is feasible.
- Make everyone potentially impacted aware of the plan.

Data

Carry out the experiment. Record the output value, the levels for candidates, the level of the dominant cause, treatment number, and run order, one row for each repeat.

Analysis

- Plot the output for each treatment number. Use a different plotting symbol for the two levels of the dominant cause. Look for promising treatments.
- Construct all cause by candidate interaction plots.

Conclusion

Examine the plots to determine the levels of the candidates (process settings) that make the output less sensitive to variation in the dominant cause.

16.2 ASSESSING AND PLANNING PROCESS DESENSITIZATION

To successfully implement process desensitization, we must:

- Identify new process settings (that is, fixed inputs and their levels) to make the process less sensitive to changes in the dominant cause.
- Move the new process center if not close to the target.
- Check that the change of settings does not produce substantive negative side effects.
- Estimate the costs of changing the settings and the change in ongoing operating costs.
- Estimate the benefit of the process change.

If we can accomplish all these tasks, and the benefits outweigh the costs, we proceed to the validation stage of the algorithm.

In most applications, there is a degree of uncertainty in the choice of candidates. If the experiment fails to find settings that reduce the sensitivity, we may consider other fixed inputs or different levels of the current candidates before we abandon the approach.

Sometimes we may have only partial knowledge of the dominant cause, as in the refrigerator frost buildup example. There, the team did not isolate the dominant cause among ambient temperature, humidity, and usage conditions. In other cases, there may be more than one dominant cause. Then, for the desensitization experiment, we recommend defining a composite cause with two levels that corresponds to the extremes of the identified causes.

The dominant cause varies in normal operation of the process. During a desensitization experiment, however, we need to hold the cause fixed at its low or high level for a run of the experiment. If this is not possible, we may need to resort to a robustness experiment as described in Chapter 19.

We strongly recommend the crossed design to ensure that all interactions between the individual candidates and the dominant cause are separately estimable, even if a fractional design is used for the candidates.

In many instances, we choose not to completely randomize the order of the runs. In the refrigerator example, eight prototypes were built and then simultaneously tested under the two levels of the (composite) dominant cause. Because of high cost, it was not feasible to randomize the order in which each refrigerator was exposed to each level of the cause.

There are some special considerations when planning a desensitization experiment for binary output. There must be enough repeats in each run so that some defectives occur on at least half the runs in the experiment. Otherwise the experiment will provide little useful information. See tables in Bisgaard and Fuller (1995a, 1995b, and 1996) for some guidance concerning the sample sizes necessary. The analysis of an experiment with a binary output can be based on the proportion defect within each run, as in the oil pan scrap example.

We start the analysis of a desensitization experiment with a plot of the individual output values by treatment number, as in Figure 16.2. We use a different plotting symbol for each level of the dominant cause. For complex problems such as the eddy current measurement system, we use scatter plots or other graphical summaries to visually characterize the performance of the process for each treatment. From these summaries, we can assess if desensitization is feasible. In a full factorial experiment, if none of the treatments are promising, we cannot desensitize the process using the selected levels of the candidates. If we have a fractional design, or if one or more treatments appear promising, we look at interaction plots between the individual candidates and the cause as in Figure 16.5.

We may find process settings that are less sensitive to variation in the dominant cause, but that result in an undesirable shift of the process center or some other side effect. In this case, we may look for an adjuster to move the center (see Chapter 15) or we may formulate a new problem to deal with the side effect. In the crossbar dimension example discussed in Chapter 12, barrel temperature was identified as a dominant cause. The team found that increasing the set point for barrel temperature (but not controlling the variation) increased the average size but substantially reduced the variation in the crossbar dimension. They then needed to move the process center and also solve a new problem called *burns*, a visual defect that occurred on some parts molded with the higher barrel temperature. See the exercises for chapters 16 and 18.

We recommend experiments with only two levels for each candidate. If we are successful in finding a candidate that desensitizes the process, we may optimize using a follow-up experiment with the identified candidate at several levels and the dominant cause at two levels. If a candidate is categorical, such as the supplier in the oil pan scrap example, we may use more than two levels in the desensitization experiment.

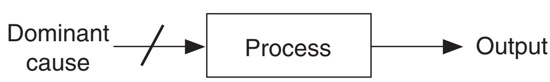
We have seen many teams proceed directly to the desensitization approach without first identifying a dominant cause. They then conduct an experiment where they change both candidates and suspects. In this way, they hope to identify a dominant cause and desensitize the process at the same time. This is a poor strategy. The experiment will be large and complex since there is little knowledge about which varying input, if any, is a dominant cause. There is also little information to help select the fixed inputs as candidates. It is more effective, both in terms of cost and the likelihood of finding a solution, to search first for a dominant cause (or at least clues about the dominant cause family) using the method of elimination with observational investigations, and then to consider a desensitization experiment.

We sometimes analyze the results of desensitization experiments using statistical models. Model building is an advanced topic that we do not cover here. See experimental design references such as Montgomery (2001) and Wu and Hamada (2000). Additional analysis using models may be necessary if we wish to predict the performance for candidate levels not used in the experiment.

The idea of desensitizing a process to a dominant cause was popularized by Taguchi (1985, 1986). He calls a suspect a *noise factor*. Taguchi (1985) cites several examples, including the famous Ina tile case. In other examples, Taguchi proceeds without knowledge of a dominant cause. This corresponds to the robustness approach described in Chapter 19. We provide a more detailed discussion of the designs and analysis suggested by Taguchi in the supplement to Chapter 19.

Key Points

- A desensitized process is one that is less sensitive (in terms of output variation) to changes in an identified dominant cause.



- To find process settings that desensitize the process, we conduct an experiment with one or more candidates. We expose each combination of candidates in the experiment to the full range of the known dominant cause to identify good treatments.
- Within each run of the desensitization experiment, we must be able to control the level of the dominant cause that normally varies.
- Desensitizing a process requires a one-time change to the product design, process design, or the process control plan.

Endnotes (see the Chapter 16 Supplement on the CD-ROM)

1. In the supplement we show how process desensitization exploits a special type of interaction between the dominant cause and one or more of the candidates.
2. The supplement provides more details on crossed fractional factorial experiments used in desensitization experiments and how to determine the confounding structure.
3. In the eddy current example, we concluded from a series of plots that none of the treatments was helpful in making the measurement process less sensitive to dirt and chemistry variation. In the supplement, we consider a formal analysis, useful if one or more of the treatments had shown promise.

Exercises are included on the accompanying CD-ROM

17

Feedforward Control Based on a Dominant Cause

To improve is to change, to succeed is to change often.

—Winston Churchill, 1874–1965

We use a feedforward controller to reduce the effect of an identified dominant cause of variation. The basic idea is to measure the cause and then predict the output value. If the predicted output is not close to the target value, we make an adjustment to the process to compensate for the predicted deviation. Figure 17.1 is a schematic of the implementation.

For feedforward control to be effective, the requirements are:

- A known relationship between the output and the dominant cause
- A reliable system to measure the dominant cause
- A timely way to adjust the process center

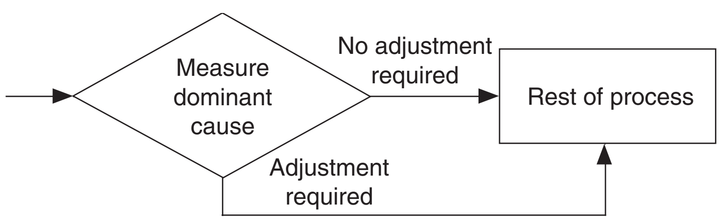


Figure 17.1 Feedforward control schematic.

The potential costs of feedforward control include:

- An investigation to quantify the relationship between the output and the dominant cause
- An experiment to find an adjuster, if not already known
- The ongoing costs of measuring the cause and making the compensating adjustment

We can assess the potential benefit of using feedforward control with the knowledge of the dominant cause and the relationship between the cause and the output. The maximum potential improvement is given by the residual variation in the output due to all other causes.

17.1 EXAMPLES OF FEEDFORWARD CONTROL

We illustrate the use of the feedforward approach with three examples.

Potato Chip Spots

When making potato chips, dark spots on the chips were undesirable to the customers. In the existing process, the dark spot problem was an uncommon but significant concern. A team was assigned to reduce the occurrence of dark spots. Dark spots were measured on a scale from 1 to 10 for each lot of chips. The rating was subjective, but by using the same assessor, the measurement system added little variation. The team used existing production records to quantify the baseline. The lot average score was 1.83 and the full extent of variation was a range of scores from 1 to 8. They also discovered that there was strong time-to-time variation in the occurrence of dark spots.

Stratifying by potato batch and investigating further, they suspected that the sugar concentration in the incoming potatoes was the dominant cause. To verify the suspicion, the team produced chips using five different batches of potatoes chosen to have a wide range of sugar concentrations. From each batch of potatoes, three lots of chips were produced. The resulting data are given in Table 17.1 and plotted in Figure 17.2.

Table 17.1 Potato chip spots data.

Batch	Sugar concentration	Dark spot scores
1	0.3%	1, 1, 2
2	0.4%	2, 1, 3
3	0.5%	4, 5, 5
4	0.6%	5, 4, 4
5	0.7%	5, 6, 7

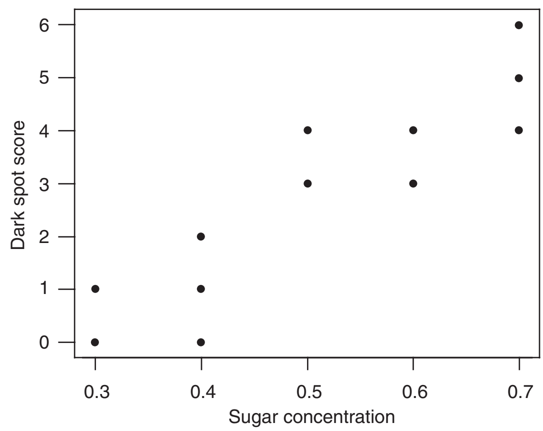


Figure 17.2 Scatter plot of dark spot score versus sugar concentration (%).

The team decided to implement a feedforward control scheme. They noticed that a significant number of dark spots occurred when the sugar concentration was 0.5% or greater. They knew they could reduce the average sugar concentration in a batch of potatoes by storing them for several weeks at temperatures above 13°C. Any batch of potatoes with initial sugar concentration above 0.4% was stored until the sugar concentration decreased. Otherwise the potatoes were processed immediately. The dark spot problem was virtually eliminated (Nikolaou, 1996). There were added storage and logistic costs but no other reported negative side effects due to the storage.

Steering Wheel Vibration

An automobile manufacturer incurred major warranty costs because of customer complaints about steering wheel vibration. A Six Sigma team found that the dominant cause of vibration was imbalance in the transmission. Reformulating the problem in terms of transmission imbalance, the team established a baseline using 200 transmissions selected from two days' production. The data are given in the file *steering wheel vibration baseline*. The distance between the center of gravity and the axis of rotation of the transmission quantifies imbalance, so lower is better. We show the histogram of the baseline data in Figure 17.3. The goal of the project was to reduce the imbalance to less than 2.0 on all transmissions since this would eliminate the vibration problem. In the baseline data, 20.5% of the transmissions exceeded this limit.

The team proceeded to look for a dominant cause without checking the measurement system since they were confident in their ability to measure imbalance. Using a component swap investigation (chapters 11 and 12), the team found that the dominant cause of the imbalance was an interaction between two mating components based on their relative orientation as set in the assembly process. Scaling for relative mass, the imbalance in the two components contributed roughly the same amount to the overall variation. The team also observed in the component swap investigation that the imbalance could be substantially reduced during reassembly by aligning the two components so their individual imbalances were offset.

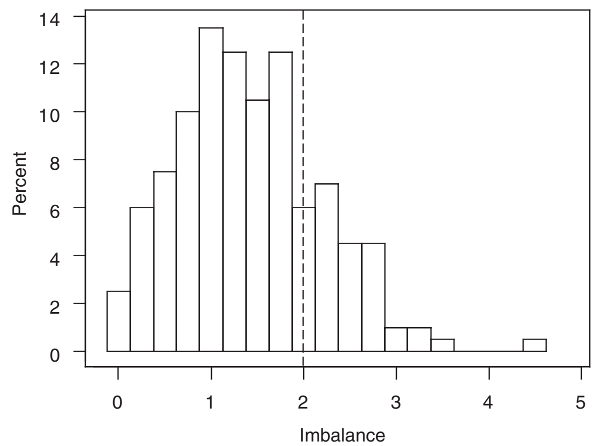


Figure 17.3 Baseline histogram of imbalance.

The team had several choices of approach. They could reformulate the problem again and try to reduce the imbalance in each of the two components. They could implement 100% inspection and rework transmissions with imbalance greater than 2.0, a very expensive option. Instead, they decided to assess the feasibility of feedforward control. In the assembly process, they planned to measure the center of gravity of each component and then assemble the two parts so that there was a maximum offset (that is, the two centers of gravity were set 180° apart). Assembling in this way is called *vectoring*. See Figure 17.4.

To estimate the possible improvement from vectoring, the team measured the center of gravity for 100 pairs of the two components selected over two days to match the baseline investigation. We give the data in the file *steering wheel vibration feedforward*.

The team calculated the imbalance for each sampled pair if vectoring had been used in the assembly. The histogram for these simulated data is shown in Figure 17.5. For these data, 3.5% of the transmissions had imbalance greater than 2.0, a marked improvement over the baseline.

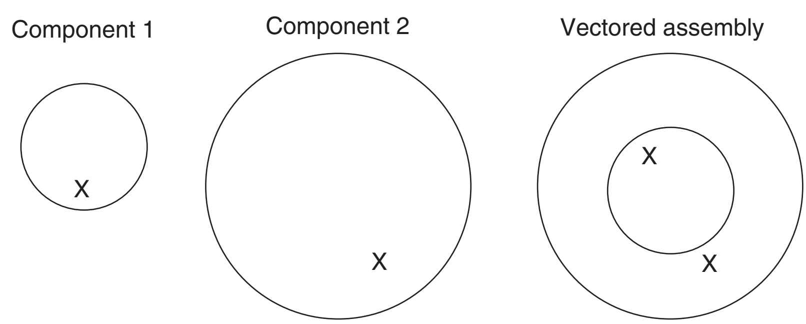


Figure 17.4 Vectoring to reduce imbalance (the *X* shows the location of the center of gravity for each component).

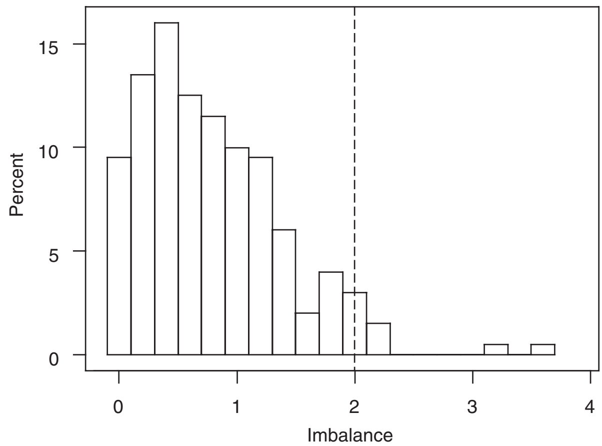


Figure 17.5 Histogram of simulated imbalances for vectored assemblies.

Since vectoring alone cannot meet the goal, the team considered *selective fitting* followed by vectoring, a more complicated and expensive option. The idea was to select a pair of components with similar imbalances and then vector them. There are two steps:

- Sort the second component into bins based on its measured imbalance.
- Measure the first component and select a matching second component from the appropriate bin.

The team needed to determine how best to sort the components (that is, determine the number of bins and their boundaries) and whether sorting was worth the trouble.

To quantify the possible benefits of selective fitting with different bin structures, the team used the data from the feedforward investigation to simulate several scenarios. They explored the effect of sorting by creating two or three bins for the second component. With two bins, the components were divided based on whether their imbalance was less than or greater than 0.75. This is roughly the middle of the range for both components in the sample. With three bins, the boundaries were 0.33 and 0.90. The simulation results are presented in Table 17.2. The benefits of selective fitting are clear. The goal can be reached by using a combination of sorting components and then vectoring. We give a more detailed explanation of how this simulation was done in the supplement.[1]

The team decided to implement feedforward control using selective fitting with two bins and vectoring. They arranged for the off-line measurement of the center of gravity (imbalance and angle) of the second component. The angle was marked on each component. The components were then sorted into two bins depending on whether the imbalance exceeded 0.75. In the assembly process, the center of gravity for the first component was measured. This component was then matched with a mating second component from the appropriate bin and the two selected components were vectored.

Table 17.2 Results of simulating selective fitting.

Assembly method	Average imbalance	Standard deviation of the imbalance	Percentage greater than 2.0
Baseline	1.31	0.83	21%
Vectoring only	0.80	0.61	4.8%
2 bins and vectoring	0.46	0.41	0%
3 bins and vectoring	0.36	0.38	0%

In determining this implementation plan, the team considered a number of logistical issues. They determined that it was feasible to:

- Keep the necessary bin of the second component on the assembly lines.
- Measure the imbalance and direction for the first component as part of the assembly process.

Originally, they planned to have the component suppliers measure, mark, and bin both components. Then, in the assembly process, matching components would have been selected from the appropriate bins. The suppliers resisted and the idea could not be implemented.

The team validated their solution by repeating the baseline investigation. The average overall imbalance of the transmission was 0.59, and fewer than 1% of the assemblies had imbalance exceeding 2.0, not quite reaching the goal but a substantial improvement.

Selective fitting and vectoring added substantial logistic and measurement costs. Occasionally, there were problems because one of the bins for the second component was empty. This is one reason the team had opted for two rather than three bins. The extra improvement from three bins did not seem worthwhile considering the additional cost and complexity.

Truck Pull

In Chapter 1, we described a project to reduce variation in pull, an alignment characteristic of trucks. At an earlier stage of the project than that described in Chapter 1, the team found that the dominant cause of pull variation was the truck frame geometry. They selected feedforward control as a working approach. As part of the search for the dominant cause, 100 frames had been selected from regular production over two weeks. Working with the supplier, the geometry was measured for each frame and reduced to four summaries (left front, right front, left rear, and right rear) thought to affect caster and camber, the measured alignment characteristics that determine pull. The team had the frames built into vehicles using the normal process and measured the caster and camber for each truck. The data are given in the file *truck pull feedforward*. We focus here on left caster.

To establish a relationship between caster and the frame geometry, we fit a regression model to predict caster as a function of the truck geometry. See the supplement to Chapter 11 for more details on regression models. The MINITAB results are:

```
The regression equation is
left caster =  -18.6 + 1.24 left front + 0.677 right front + 0.140 left rear +
0.156 right rear

Predictor         Coef     SE Coef          T        P
Constant      -18.6097      0.6100     -30.51    0.000
left front     1.24243     0.04018      30.92    0.000
right front    0.67669     0.03585      18.88    0.000
left rear      0.13953     0.02763       5.05    0.000
right rear     0.15624     0.04162       3.75    0.000

S = 0.1760      R-Sq = 96.4%     R-Sq(adj) = 96.2%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         4      77.827      19.457    627.82    0.000
Residual Error    95       2.944       0.031
Total             99      80.771
```

Truck frame geometry is a good predictor of the left caster value. We make the prediction by substituting the known frame geometry summaries into the equation:

$$\text{left caster} = -18.6 + 1.24\ \text{left front} + 0.677\ \text{right front} + 0.140\ \text{left rear} + 0.156\ \text{right rear}$$

The standard deviation of the left caster for the 100 vehicles in the investigation is 0.903. The residual standard deviation is 0.176. If we could eliminate the effect of the variation in the frame geometry, we could reduce the variation in left caster by a factor of five. There was a similar pattern for the other alignment characteristics.

In reality, a mathematical prediction model was built using knowledge of the geometry of the suspension. The regression model mimicked the mathematical model well.

Since the potential benefit was large, the team recommended implementing feedforward control. For each frame, the supplier measured and bar-coded the frame geometry summaries. Then at the truck assembly plant they:

- Read the geometry summaries from the bar code and used the values to predict caster and camber.
- Built positioning components based on the prediction to adjust the caster and camber.
- Used the custom components to position the suspension components on the frame during the assembly.

After implementing feedforward control, the standard deviation in left caster values, given in the file *truck pull validation*, was 0.25. The full reduction in left caster variation predicted by the regression model was not achieved because there was some error in building and using the positioning components and there were also small errors in the mathematical model.

There was a large benefit for the extra cost in terms of reduced warranty claims. The truck pull variation was reduced by roughly 70% through the use of the feedforward controller.

17.2 ASSESSING AND PLANNING FEEDFORWARD CONTROL

To successfully implement feedforward control, we must:

- Identify an adjuster that can be used repeatedly to quickly move the process center.
- Define a scheme to sample and measure the dominant cause.
- Determine a method to predict the output on the corresponding part.
- Specify an adjustment rule, that is, determine when and how much to adjust the process.
- Assess the sampling, measuring, and adjusting costs.
- Estimate the benefit of the process change.

If we can accomplish all these tasks, and the benefits outweigh the costs, we proceed to the validation stage of the algorithm.

The use of an adjuster for feedforward control is different from its use to move the process center as described in Chapter 15. We expect to make frequent adjustments of different sizes over time to compensate for changes in the dominant cause. Adjusters suitable for a one-time shift of the process may not be useful in feedforward control. In implementing feedforward control, perfect adjustment is typically not possible. For example, when using selective fitting, finding perfectly matched components may be too much effort.

We must be able to measure the dominant cause with an effective and timely measurement system. If there are large errors in measuring the cause, these errors will affect the prediction of the output and hence the adjustment procedure. The measured value of the cause must be available in time to make the adjustment.

In both the potato chip spots and steering wheel vibration examples, we used informal prediction of the output using the measured value of the dominant cause. The simplest formal model is a linear relationship between the cause and the output characteristic. That is

$$\text{predicted output} = \text{intercept} + \text{slope} * \text{dominant cause}$$

Predictions should be accurate for the full range of values normally seen for the dominant cause. Informal methods and simple models may give poor predictions and more complicated models are sometimes warranted.[2]

Feedforward control is effective only if based on a dominant cause. By definition, we cannot predict the output well with a cause of variation that is not dominant. If we base an adjustment on a poor prediction, we may increase rather than decrease the output variation.

If we measure the dominant cause for every unit, we must be able to apply the adjustment to each unit. This can be a difficult task in a complex process, especially if the adjustment takes place far downstream from the measurement of the cause. In the truck pull example, the team solved this difficulty by bar-coding the frames.

The frequency of adjustment depends on how the dominant cause varies. If the dominant cause acts in the part-to-part family, we may need to make an adjustment for every part. If the dominant cause acts in the time-to-time family, we can make adjustments less often, as in the potato chip spots example. Since frequent adjustment adds to the cost and complexity of the process, the feasibility of feedforward control depends on the nature of the variation of the dominant cause. If adjustment costs are large, we may decide to only adjust when the difference between the predicted output and the target is large.

In the truck pull example, we use the residual standard deviation after fitting a regression model relating the output and the dominant cause to estimate the benefits of feedforward control. This standard deviation underestimates the remaining variation because there are likely measurement, prediction and adjustment errors. In cases where the dominant cause acts batch-to-batch, such as in the potato chip spots example, we use the within-batch variation as an optimistic estimate of the benefit. We can sometimes use available process data to simulate the performance of a proposed feedforward controller to assess its benefit, as in the steering wheel vibration example.

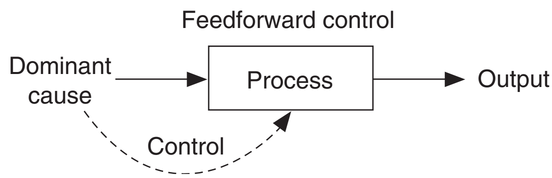
Feedforward control is related to feedback control, described in Chapter 18. In both approaches we reduce variation by adjusting the process based on a prediction of the output. The fundamental difference is that, with feedforward control, we predict the output using the dominant cause, while with feedback control, we predict the future output using previously observed output values.

Feedforward control can also be thought of as a way of desensitizing the process to variation in a dominant cause, as discussed in Chapter 16. In the desensitization approach, we searched for process settings that permanently made the process less sensitive to variation in a dominant cause. With a feedforward controller, the desensitization requires repeated adjustments to compensate for the different values of the dominant cause.

There is little published work on feedforward controllers available in the statistical literature. The most comprehensive work is Box and Jenkins (1976). See also Jenkins (1983).

Key Points

- A feedforward controller compensates for the value of a dominant cause by predicting the output from the measured value of the cause and then adjusting the process based on the deviation between the prediction and the target.



- To implement feedforward control, we require the value of a dominant cause, knowledge of the relationship between the cause and the output in order to make the prediction, and a timely and cost-effective adjuster.
- If the dominant cause acts in the time-to-time family, we need to adjust once, at most, once time period.

Endnotes (see the Chapter 17 Supplement on the CD-ROM)

1. In the supplement, we explain how to assess the potential benefits of vectoring and selective fitting in the steering wheel vibration example using simulation and the available data.
2. In some applications more complicated prediction methods or models are warranted. In the supplement we give references for these methods.

Exercises are included on the accompanying CD-ROM

18

Feedback Control

Quality is never an accident, it is always the result of intelligent effort.

—John Ruskin, 1819–1900

Feedback control is used to compensate for a predictable pattern in the output characteristic due to known or unknown causes. We reduce variation by predicting the next output value using previously observed output values. Then we adjust the process center to compensate for the deviation between the predicted output and the target. We do not need to identify a dominant cause of variation with this approach. We illustrate how to implement a feedback control scheme in Figure 18.1.

For feedback control to be effective, the requirements are:

- The dominant cause must act in the time-to-time family.
- There must be a timely way to adjust the process center.

The potential costs of feedback control include:

- An investigation to explore the time pattern in the output, if not known
- An experiment to find an adjuster, if one is not known
- The ongoing costs of measuring the output and making the adjustment

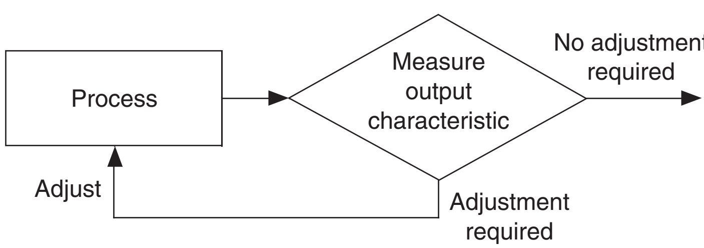


Figure 18.1 Feedback control schematic.

We can assess the potential benefits of feedback control once we know the time pattern of variation in the output. With feedback control we can at best hope to eliminate the time-to-time component of the output variation.

18.1 EXAMPLES OF FEEDBACK CONTROL

We illustrate the use of feedback control with five examples.

Parking Brake Tightness

In a vehicle assembly process, there was a problem with excess variation in parking brake tightness that led to considerable rework. To measure tightness the team determined the number of clicks until the brake locked. The specification was 5 to 8 clicks. At final inspection, the plant found parking brakes that were both too tight and too loose.

In the baseline investigation, the tightness ranged between 3 and 9 and there were runs of high and low values. The dominant cause acted in the time-to-time family. Based on this clue and process knowledge, the team speculated that the dominant cause of tightness variation was the length of either the front or axle cables. The two assemblies that include these cables were delivered in batches from a supplier. To search for the dominant cause, the team conducted an investigation where parking brake tightness was measured for three consecutive vehicles for 12 combinations of batches of the two cable assemblies.

The data are given in the file *parking brake tightness multivari*. From Figure 18.2, we see that the dominant cause acts in the front cable assembly and differs from batch to batch. The variation within batches is substantially smaller than the batch-to-batch variation. The baseline tightness values ranged over seven possible values; within a batch of front cable assemblies, the range covered only three values. In the multivari investigation there were six different tightness values, which roughly matches the full extent of variation.

The team did not further explore front cable length as a suspect. They could not measure the length of the cable in the plant since it arrives as part of an assembly. Because the dominant cause was in the batch-to-batch family, they decided to consider feedback control.

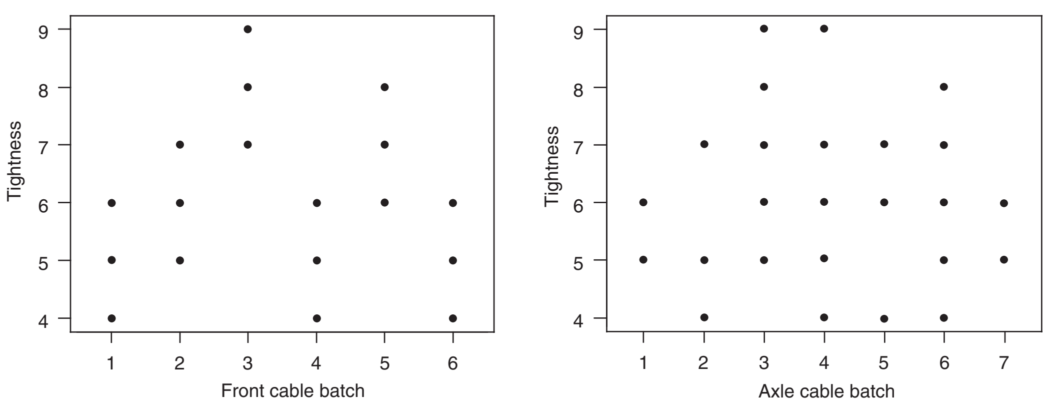


Figure 18.2 Box plot of tightness by front and axle cable batch.

To implement the approach, they needed to find an adjuster. The team knew that changing the depth of an adjustment nut would change parking brake tightness. To calibrate the adjuster, they selected five vehicles of varying tightness to cover the full extent of parking brake tightness variation. The parking brake system on each vehicle had been originally installed with the adjustment nut at a depth of 24 millimeters. The team reinstalled the parking brake system on each vehicle with the nut at the four depths 22, 23, 25, and 26 millimeters and measured the corresponding tightness. The data are given in Table 18.1 and in the file *parking brake tightness adjuster calibration*.

We use MINITAB to fit a regression model to relate the average tightness to the depth with results:

```
The regression equation is
average = -15.0 + 0.880 depth (mm)

Predictor        Coef    SE Coef        T       P
Constant      -15.000      1.666    -9.01   0.003
depth (mm)    0.88000    0.06928    12.70   0.001

S = 0.2191      R-Sq = 98.2%     R-Sq(adj) = 97.6%

Analysis of Variance

Source           DF         SS         MS        F        P
Regression        1     7.7440     7.7440   161.33    0.001
Residual Error    3     0.1440     0.0480
Total             4     7.8880
```

The estimated slope is 0.88, so changing the adjustment nut depth by one millimeter will change the tightness, on average, by about 0.88 clicks.

There was little cost to measuring tightness and to changing the depth of the adjustment nut. Since the dominant cause was in the batch-to-batch family, the team concluded that feedback control was feasible.

To implement the approach, the team decided to measure the tightness of the first five vehicles assembled for each new batch of front cable assemblies. These measurements were made as soon as possible once the parking brake was installed. They made an adjustment

Table 18.1 Parking brake tightness adjuster experiment results.

Vehicle	22 mm depth	23 mm depth	24 mm depth	25 mm depth	26 mm depth
1	2	2	3	4	5
2	2	3	4	5	6
3	4	5	6	7	7
4	7	7	8	9	10
5	8	8	9	10	11
Average	4.6	5.0	6.0	7.0	8.0

if any of these measurements was outside the specification limits (5 to 8). The amount of the adjustment was based on the difference between the target tightness (6.5) and the average tightness for the five vehicles. To increase the average tightness by one click, they increased the depth of the adjustment nut by roughly 1/0.88 = 1.14 millimeters. For ongoing protection, and to ensure that the feedback control was effective, they changed the control plan so that within a batch of front cables, the parking brake tightness was measured on every tenth vehicle. If an out-of-specification vehicle was found, then all vehicles for that batch were inspected and reworked if necessary. After implementation, there was a marked reduction in the amount of rework required because of parking brake tightness variation.

V6 Piston Diameter

We discussed the problem of excess diameter variation in the production of V6 aluminum pistons in chapters 2, 5, 9, and 11. From the problem baseline investigation, the team learned that the process was centered on target with standard deviation 3.32 microns. The problem goal was to reduce the standard deviation to less than 2.0 microns. The team determined that the measurement system was adequate and, using a variation transmission investigation (see Chapter 11), that the home of the dominant cause was the intermediate Operation 270. They reformulated the problem to reduce the variation in the diameter as measured at Operation 270.

At Operation 270, there were two parallel grinders. To better understand the performance of the process, the team measured one piston a minute from each grinder for 200 minutes. They ensured that no adjustments were made to the grinders while the data were collected. The data for stream 1 are given in the file *V6 piston diameter 270*. Diameter is recorded in microns as the deviation from 87 millimeters. Thus, a measured diameter of 87.595 millimeters is recorded as 595. We show the run chart for stream 1 in Figure 18.3.

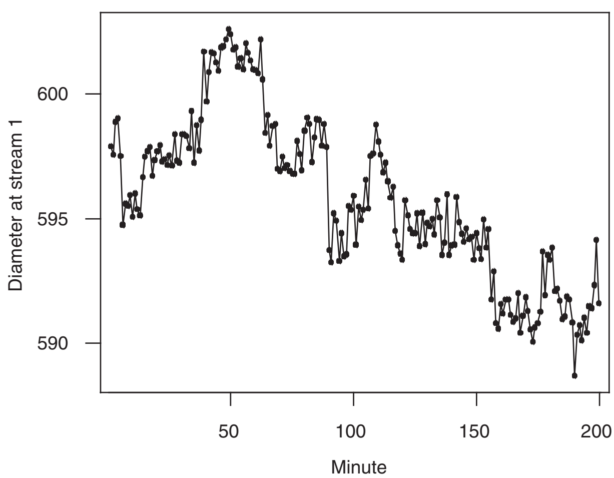


Figure 18.3 Diameter at stream 1 at Operation 270 by minute.

The plot for the second stream is similar. Once the data were stratified by stream, the team saw that the short-term variation in the diameter was small relative to the time-to-time variation.

Based on the run chart, the team decided to look at the feasibility of feedback control separately for each stream. They found this approach attractive because they knew that the operator could make an adjustment at Operation 270 within a short time and with immediate effect.

The team selected an informal feedback scheme similar to precontrol.[1] The rules for each stream were:

- Every 15 minutes, measure the diameter of two consecutive pistons after machining.
- Compare the average diameter to the adjustment limits 592.7 and 600.7 microns.
- If the average falls outside the limits, adjust the process to the target 596.7 microns.

The size of the adjustment, if any, is the difference between the observed average diameter and the target.

The team derived the adjustment limits by working backwards using information about the relationship between the diameter at Operation 270 and the final diameter. To meet the overall goal, they needed the final diameter to be well within the specification limits of 591 ±10 microns.

To assess the potential gain from this approach, the team used the data in the file *V6 piston diameter 270* to simulate the effect of feedback control.[2] They assumed the adjustment was perfect in the simulation, which would not be true in practice. We see the results of the simulation in Figure 18.4.

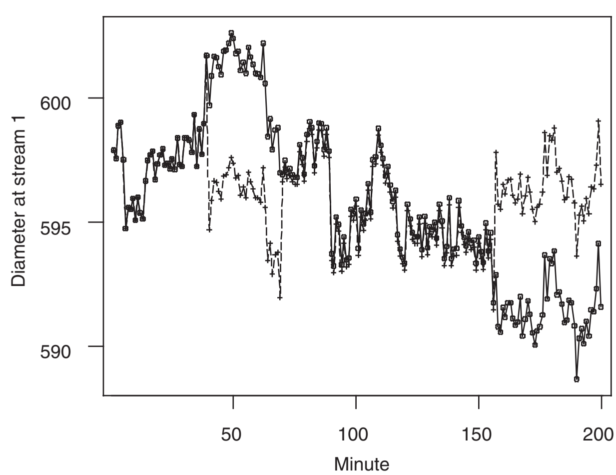


Figure 18.4 Diameter at stream 1 of Operation 270 by minute (solid line gives the original series, dashed line gives the adjusted series).

We estimate the benefit of the proposed feedback scheme by comparing the adjusted and unadjusted series. Using MINITAB, we get:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
original	200	5.808	5.548	5.770	3.269	0.231
adjusted	200	6.024	6.078	6.027	1.694	0.120

Variable	Minimum	Maximum	Q1	Q3
original	-1.300	12.586	3.552	7.916
adjusted	1.490	11.703	4.753	7.231

Based on the simulation, we expect the feedback controller to reduce the standard deviation of the diameter at Operation 270 from 3.3 to about 1.7. We combine this information with the results from the variation transmission investigation (see the supplement to Chapter 11) to predict that after implementing the feedback controller, the standard deviation in the final diameter would be reduced to

$$\sqrt{0.88^2\left(1.7\right)^2+1.224^2}=1.93 \text{ microns}$$

The team changed the control plan at Operation 270 and trained the operators to use the new control scheme. They also arranged for a periodic process audit to ensure the new scheme was being used.

After implementation, the team repeated the baseline investigation to validate the improvement. The standard deviation of the finished diameter was reduced from 3.32 to 2.20 microns.

V8 Piston Measurement System

In the production of V8 pistons, a 100% inspection gage checked many outputs and rejected pistons not meeting specifications. An auditor retrieved a piston from shipped inventory and remeasured it on the inspection gage. The piston was badly out of specification with respect to one of its diameters, measured at a fixed height on the skirt. The measured diameter was 13.2 microns (measured from the nominal dimension) compared to the specification ±9.0 microns.

Management assigned a team to prevent the shipment of out-of-specification pistons. The team reviewed the calibration procedure for the gage, which was carried out once per day. They conducted a short-term measurement system investigation, measuring 10 pistons three times each in haphazard order within 15 minutes. The data are given in the file *V8 piston diameter short-term measurement*. The results of a one-way analysis of variance from MINITAB are:

Analysis of Variance for diameter

Source	DF	SS	MS	F	P
part	9	524.587	58.287	328.07	0.000

```
Error      20     3.553     0.178
Total      29   528.140

Pooled StDev =   0.4215
```

The estimate of measurement variation was 0.42 microns. If the diameter was within specification at the original measurement, we cannot explain a measured value 13.2 by the short-term measurement system variation.

The team next decided to assess the stability of the measurement system by measuring the same two pistons every 15 minutes for 12 hours, starting immediately after calibration. The data are given in the file *V8 piston diameter measurement stability* and plotted in Figure 18.5.

Over the 12-hour period, the variation was much larger than expected based on the results from the short-term measurement investigation. The range of diameter values for each piston was around 8 microns. This drift could easily be responsible for the out-of-specification piston found in the audit.

The team immediately increased the frequency of calibration to every two hours. They decided not to look for the cause of the drift but instead implemented a feedback control scheme. Immediately after calibration and then every 15 minutes, the operator measured the same piston and recorded its diameter. If the change from the initial measurement was more than 1.8 microns (in either direction), the process was stopped and the gage was recalibrated. Note that the short-term standard deviation of the difference in two measurements on the same piston is $0.42\sqrt{2} = 0.59$, so a difference of 1.8 microns indicates that the measurement system has drifted.

The changes added to the operating cost of the inspection gage. However, no further out-of-specification pistons were found in audits. There was a positive side effect. The scrap rate due to out-of-specification diameters decreased by about 50%, which more than made up for the loss of cycle time of the gage. Before the change, the gage rejected many

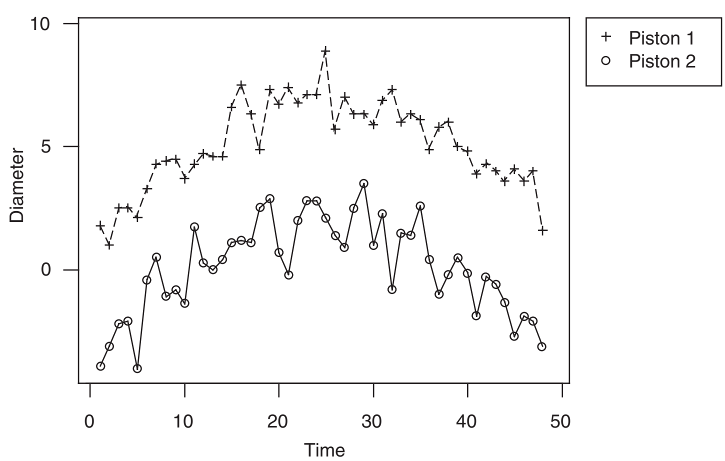


Figure 18.5 Piston diameters versus time.

good pistons. As well, there were fewer adjustments to the upstream machining process that had been driven by the drift in the inspection gage.

Fascia Film Build

We introduced a problem of excess variation in the film build (paint thickness) of fascias in Chapter 3. After finding the dominant cause, the team reformulated the problem in terms of flow rate. To establish a baseline for the new problem, the team recorded the flow rate every minute for three hours. The data are given in the file *fascia film build baseline* and plotted in Figure 18.6. Feedback control is a possibility here since eliminating the drift over time would reduce the variation in flow rate substantially. The range of flow rate values over the short term is less than half the long-term range.

Since the team knew they could adjust the flow rate quickly using a valve, they decided to investigate the possibility of using feedback control to reduce flow rate variation.

This team took a more formal approach. To predict the flow rate at the next minute, they used an *exponentially weighted moving average* (EWMA).[3] To explain this method, let y_t be the measured output at time *t* (measured in minutes here). Then we predict the next value of the output, denoted $\hat{y}_{t+1}$, using a weighted average of previous values as given in Equation (18.1).

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \alpha(1-\alpha)^3 y_{t-3} + \ldots , \qquad (18.1)$$

where α (the Greek letter *alpha*) is a constant and $0 < \alpha \le 1$. In the prediction, the most recent value y_t gets the highest weight α, the next most recent point, y_{t-1}, gets the second highest weight a (1 – a*)*, and so on. You can use a bit of probably forgotten high school algebra to show that the weights add to one.

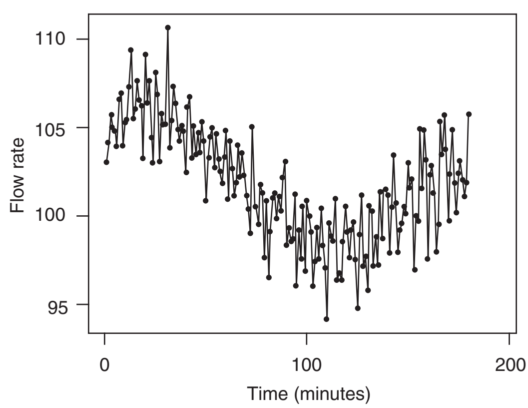


Figure 18.6 Run chart of flow rate.

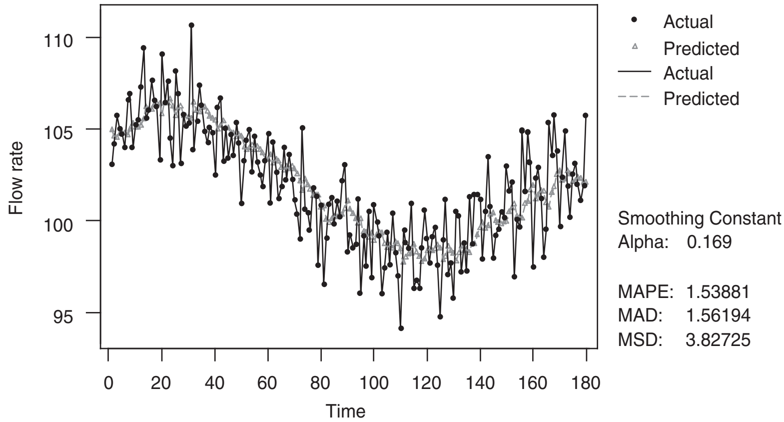


Figure 18.7 EWMA smoothing of paint flow rate.

We use MINITAB (see Appendix C) to estimate the parameter α from the data collected over a period of time when there were no adjustments. For the flow rate data, we get the results shown in Figure 18.7. The estimated value for α is 0.169. We also show the series of predicted values in Figure 18.7 to demonstrate how the EWMA prediction captures the long-term pattern in the series and smoothes the short-term variation.

Once the controller is operational, we start to make adjustments and see the series that includes the effects of earlier adjustments. Let y_t^* be the observed output with the controller in operation. After some algebra,[4] we can show that the appropriate adjustment at time t is $\alpha(y_t^* - \text{Target})$ in the direction of the target. In other words, if y_t^* is greater than the target value, we move the process center down by $\alpha(y_t^* - \text{Target})$. Here we assume adjustments are made after every measurement.

If we make perfect adjustments, we get an estimate of the process standard deviation when the controller is in use from the MINITAB results. For this example, the estimated standard deviation is 1.96 ($\sqrt{MSD}$ in Figure 18.7), a substantial reduction from 3.23, the standard deviation of the unadjusted series, which serves as a baseline here.

To implement this feedback control scheme, the team calibrated the valve and automated the adjustment so that it could be carried out each minute. The standard deviation of the film build was markedly reduced.

Truck Pull

In Chapter 1 and elsewhere, we discussed a problem with excess pull variation in trucks. Pull, an alignment characteristic, is a linear function of left and right camber and caster. The baseline data *truck pull baseline* include the values for left and right caster and camber on more than 28,000 trucks produced over a two-month period. As a result of the problem focusing effort discussed in Chapter 6, the team determined that reducing caster variation was the highest priority. They searched unsuccessfully for a dominant cause of this variation.

The team then decided to investigate feedback control as an approach to reduce the caster variation. We illustrate the assessement of feasibility using right caster. There were similar results for left caster. The target for right caster is 4.5°. A summary of the baseline data is:

```
Descriptive Statistics: r-caster

Variable          N      Mean    Median    TrMean     StDev   SE Mean
r-caster      28258    4.5188    4.5210    4.5192    0.2427    0.0014

Variable    Minimum   Maximum        Q1        Q3
r-caster     3.0440    5.9380    4.3600    4.6780
```

The process is well centered with standard deviation 0.243°. We see in the plot of right caster angle over time (see Figure 18.8) that there is some drift in the process near the middle of the series.

The team knew they could adjust caster using the same custom components that were built for the feedforward control scheme discussed in Chapter 17. Since these components were manufactured and assembled approximately two hours before caster was measured, feedback control based on adjustments after each truck was not feasible. The team decided to investigate the effects of making an adjustment once per shift. The process is run with three eight-hour shifts per day.

As a first step, they looked at a one-way analysis of variance as described in the supplement to Chapter 11. The idea was to separate the baseline variation into two components, variation within shifts and variation from shift to shift. The partial results from MINITAB are:

```
One-way ANOVA: r-caster versus shift

Analysis of Variance for r-caster
Source     DF        SS        MS        F        P
shift  131   456.378     3.484    81.09    0.000
Error    28126  1208.334     0.043
Total    28257  1664.712
```

The estimate of the within-shift standard deviation, labeled MS(Error) in the MINITAB results, is $\sqrt{0.043} = 0.207$. If we could use feedback control to make all the shift averages equal, we would expect a reduction in the baseline standard deviation from 0.243 to 0.207, a modest improvement.

Summarizing the caster shift averages gives:

```
Variable          N      Mean    Median    TrMean     StDev   SE Mean
right caster    132    4.5271    4.5135    4.5269    0.1255    0.0109
shift average

Variable        Minimum   Maximum        Q1        Q3
right caster     4.1909    4.9721    4.4473    4.6154
shift average
```

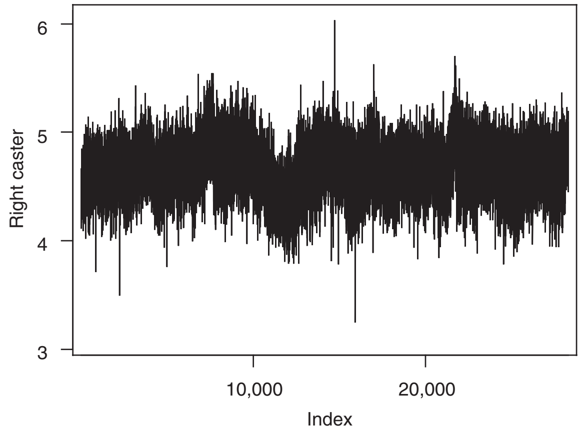


Figure 18.8 Right caster angle over time.

To obtain the shift average summary from the available data, we first store the caster averages by shift (see Appendix B). Across the 132 shifts, the standard deviation of the shift averages is 0.126. We plot the shift averages and the predicted averages from the EWMA in Figure 18.9. The smoothing constant is $\alpha = 0.944$, so the most recent shift is given very high weight in the prediction of the subsequent shift average. If we use feedback based on the EWMA to adjust the process at the start of the shift, then we estimate that the standard deviation of the shift averages will be reduced to 0.077 $\left(\sqrt{0.00598}\right)$, less than half the baseline shift-to-shift variation, 0.126. However, since the within-shift variation is so large, this reduction has little impact on the overall variation.

At this point, the team decided that the potential gain of feedback control would not be worth the cost. Given the two-hour time lag and the imperfect adjuster, they knew they would not be able to achieve the small gain predicted using the baseline data.

The team abandoned the problem because no other approach was feasible.

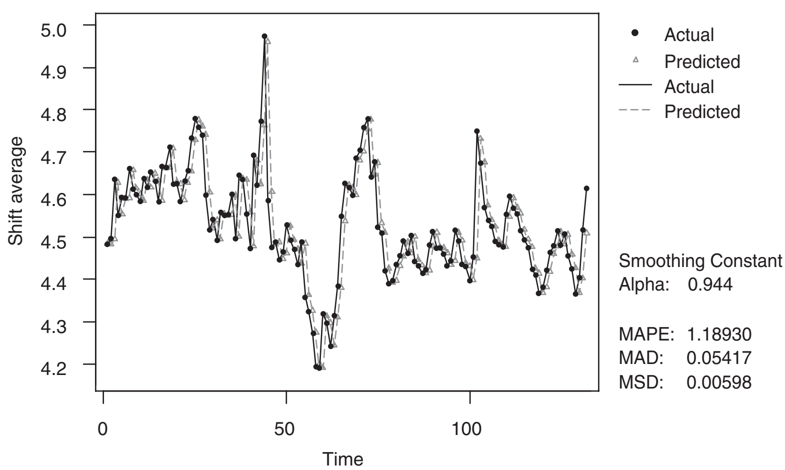


Figure 18.9 Right caster average by shift.

18.2 ASSESSING AND PLANNING FEEDBACK CONTROL

To successfully implement feedback control, we must:

- Identify an adjuster that can be used repeatedly to quickly change the process center.
- Define a scheme to sample and measure the output.
- Determine a method to predict future output values from the current and past output values.
- Specify an adjustment rule, that is, determine when and how much to adjust the process center.
- Assess the sampling, measuring, adjusting costs, and possible side effects.
- Estimate the benefit of the process change.

If we can accomplish all these tasks, and the benefits outweigh the costs, we proceed to the validation stage of the algorithm.

The pattern of variation in the output can be a gradual drift, as in the fascia film build example, or a sudden persistent shift, as shown for machining data in Figure 18.10. Sudden shifts can occur if the output center depends on the setup procedure, the batch of raw materials, tooling changes, and so on. See the chapter exercises for more details on the machining process example.

We observe a pattern of variation with respect to a particular sampling scheme. For example, in the fascia film build example, the team measured the flow rate every minute. The observed pattern would have been different had they measured the flow rate every 15 minutes or every 0.1 seconds. The sampling protocol in the baseline investigation may be sufficient to reveal the time pattern of variation in the output. If not, we suggest a simple multivari investigation with two families, time-to-time and part-to-part. Select the time-to-time family (for example, every 15 minutes, every hour, every shift, and so on) based on the time required to measure the output and to adjust the process.

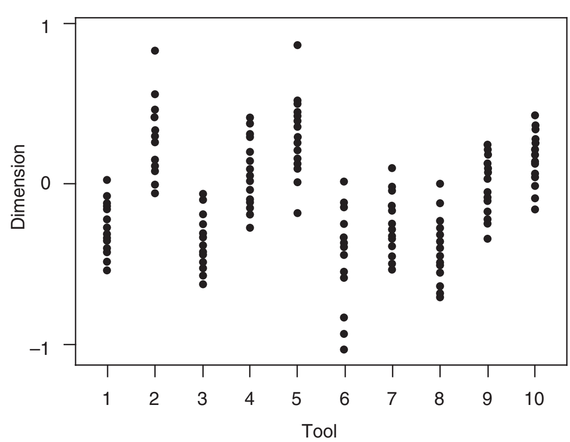


Figure 18.10 A machining process with persistent shifts due to tooling.

We must be able to measure the output quickly compared to the time period that defines the family containing the dominant cause of variation. In an iron foundry, there was a one-hour delay between the time the iron was sampled and the time the chemistry measurements became available. Since the within-hour variation in the process chemistry was relatively large, feedback control was not feasible. In the truck pull example, there was a lag of two hours between the point of adjustment and the measurement of caster. Hence the team considered feedback control schemes that would adjust caster at most every two hours .

We can make similar comments about the action of the adjuster. If the time for the adjustment to take effect exceeds the time period that defines the family containing the dominant cause, we cannot use feedback control effectively.

We recommend determining the pattern of variation before looking for an adjuster since the cost of finding an adjuster may be substantial. If needed, we can search for an adjuster using an experimental plan as described in Chapter 15. Sometimes we need to calibrate a known adjuster, as in the parking brake example, so that we can make adjustments of different sizes to the process center as required.

To specify a feedback control scheme, we need:

- A sampling plan to determine when and how to measure the output
- A rule to predict future output values and to decide what adjustment, if any, is required
- A procedure to make the adjustment

The design of the sampling plan involves a trade-off between the cost of measurement and the ability to predict the future output values. In the V6 piston diameter example, the team decided to measure two pistons every 15 minutes. In the fascia film build example, the team measured flow rate every minute.

To predict future output values, we can use an exponentially weighted moving average or a simple method such as the value of the most recent observation. In the supplement, we discuss the choice of prediction method more fully.[5]

Adjusting after every observation may be undesirable due to the added process complexity and adjustment costs. A simple alternative is to use a dead band or bounded feedback adjustment (Box and Luceno, 1997). With dead bands, we do not adjust the process if the predicted deviation from the target is small. Compared to adjusting after every observation, a dead band scheme results in a smaller decrease in variation and a large reduction in the number of adjustments.

The size of an adjustment was the difference between predicted output and the target in the examples discussed here. However, in cases where a process is subject to regular drift either upward or downward, say due to tool wear, the adjustment may be to the opposite side of the specification limit rather than to the target. This idea is illustrated in the discussion of feedback control in Chapter 3.

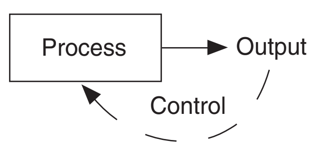
Once we have designed the feedback scheme, we can assess the potential benefits in several ways if we have a series of historical data that matches the sampling protocol of the scheme. We can simulate the effect by applying the adjustments to the historical series. If we base the control scheme on an EWMA, we can use MINITAB to estimate the standard deviation of the adjusted process. If we have multivari data with short-term and time-to-time families, we can use a one-way ANOVA to estimate the variation within the short-term family. This is the best we could hope to achieve with a feedback control scheme designed to eliminate the time-to-time variation. In all cases, we assume adjustments are made without error and so, if this is not the case, we overestimate the benefit of a feedback control scheme.

A comprehensive treatment of feedback control from a statistical perspective is given in Box and Luceno (1997) and Del Castillo (2002). There are many variations of feedback control. See Tucker et al. (1993), Box and Jenkins (1994), and Box and Luceno (1997) for further details. Specific examples include acceptance control charts (Duncan, 1986) and precontrol (Shainin and Shainin, 1989; Juran, Gryna, and Bingham, 1979).[6] There is also an extensive engineering literature on feedback control.[7]

If a dominant cause acts part to part, using feedback control will result in increased variation. This is called tampering by Deming (1992, pp. 327–328). The effect of tampering is shown by Deming's famous funnel experiment.

Key Points

- Feedback control is based on predicting future output values using the current and past output values and adjusting if appropriate.
- Feedback control may be effective if the (unknown) dominant cause acts in the time-to-time family.



- Feedback control requires an adjuster that moves the process center quickly, has low cost to change, and has small effect on other outputs.
- We may need to calibrate the adjuster with an experiment because we need to make repeated adjustments of different sizes.
- To implement feedback control, we require a sampling plan and a rule to decide when and how large an adjustment to make.

Endnotes (see the Chapter 18 Supplement on the CD-ROM)

1. There are many other informal prediction and adjustment rules. We briefly describe Grubbs' rule, precontrol, and statistical process control charts used for feedback control.
2. We describe how to simulate the benefits of a feedback control scheme using a historical series.

3. In the supplement, we compare EWMA to some other formal prediction methods. We also show that using partial compensation of the current deviation from target to determine the adjustment is appropriate under some assumptions.
4. See note 3.
5. See note 3.
6. See note 1.
7. We describe the connection between the simple feedback control schemes discussed here and engineering proportional-integral-differential (PID) controllers.

Exercises are included on the accompanying CD-ROM

19

Making a Process Robust

It is a capital mistake to theorize in advance of the facts.

—Sir Arthur Conan Doyle (as Sherlock Holmes), 1859–1930

To make a process robust, we must find changes to fixed inputs that make the process output less sensitive to variation in the *unknown* dominant cause. The robustness approach is similar to desensitization, described in Chapter 16. However, now we do not have knowledge of a specific dominant cause.

The requirements for robustness to be effective are:

- The unknown dominant cause acts in the short-term family of variation.
- New settings of fixed inputs that reduce the effect of the unknown dominant cause resulting in less output variation.

To assess the feasibility of robustness, we choose a number of fixed inputs (candidates) and an experimental plan to determine if changing the levels of these candidates will make the process robust. The first requirement is important because we must define a run to be long enough to see the full extent of variation of the output under the current candidate settings. As a consequence, the dominant cause will vary over its full range within each run. Hence, for every run, we will be able to see if the candidate settings make the process robust to the variation in the unknown dominant cause. If the dominant cause acts in the time-to-time family, it will likely not be feasible to conduct such an experiment since the runs would need to be too long.

The costs of the robustness approach include:

- An experiment to search for the new settings of the candidates
- A one-time change to the process settings
- The ongoing operation with the new process settings

There is no information about whether this approach will be feasible until the experimental investigation is complete. There is a risk of running a high-cost experiment with no return. We cannot assess the benefits of this approach until we find the new settings for the candidates.

19.1 EXAMPLES OF PROCESS ROBUSTNESS

We present three examples of the robustness approach. We selected the third example to demonstrate what can go wrong.

Crossbar Dimension

We discussed the problem of reducing variation in a crossbar dimension in Chapter 12 and the exercises for Chapter 16. The team raised the barrel temperature set point to make the process less sensitive to variation in barrel temperature, the dominant cause. When validating the solution, they showed the variation in the crossbar dimension was substantially reduced but, with the new setting, there was an increase in the frequency of a mold defect called *burn.* They decided to address the burn defect as a new problem. Using a multivari investigation, they showed that the dominant cause of burn acted in the part-to-part family. but the specific dominant cause was not found. They suspected that the defect occurred when the mold cavity filled too fast. In any case, since the suspect dominant cause could not easily be controlled, the team decided to try the robustness approach.

The team planned an experiment with four fixed inputs (candidates): injection speed, injection pressure, back pressure, and screw speed. These candidates were selected because of their influence on fill time and other potential dominant causes in the part-to-part family. They selected two levels for each candidate as given in Table 19.1.

Table 19.1 Candidates and levels for burn robustness experiment (level in current process given by *).

Candidates	Label	Low level	High level
Injection speed	A	slow*	fast
Injection pressure	B	1000*	1200
Back pressure	C	75	100*
Screw rpm	D	0.3	0.6*

The team decided to define a run as five consecutive parts. Since they knew the dominant cause acted in the part-to-part family, they expected it to act within each run. Each run was carried out once the process stabilized after changing the values of the candidates.

The team selected a fractional factorial experiment with the eight runs given in Table 19.2. Since there was no proper baseline investigation, the team assigned the letters to the candidates so that one of the treatments (treatment 5) corresponded to the current process settings (see the Chapter 15 supplement for details). The confounding pattern of the chosen design is given as follows. In the resolution IV design, pairs of two input interactions are confounded.

```
Alias Structure
A + BCD
B + ACD
C + ABD
```

```
D + ABC
AB + CD
AC + BD
AD + BC
```

The burn on each part was classified into one of four categories of increasing severity. Levels 1 and 2 were acceptable, while levels 3 and 4 resulted in scrap. The order of the runs was randomized. The experimental results are given in Table 19.2 and the file *crossbar dimension robustness*.

Treatments 2 and 3 are promising relative to the current process as given by treatment 5. We plot the burn scores against treatment in Figure 19.1. Because the data are discrete, we add jitter in the vertical direction (see Appendix C). We use MINITAB (see Appendix F) to stack the treatment and burn score columns to create one row for each repeat in the experiment.

Table 19.2 Experimental plan and data for burn robustness experiment.

Treatment	Run order	Injection speed	Injection pressure	Back pressure	Screw speed	Burn scores	Average burn
1	4	Slow	1000	75	0.3	1, 2, 1, 1, 1	1.2
2	8	Fast	1000	75	0.6	1, 1, 1, 1, 1	1.0
3	2	Slow	1200	75	0.6	1, 1, 1, 1, 1	1.0
4	3	Fast	1200	75	0.3	1, 2, 2, 2, 2	1.6
5	5	Slow	1000	100	0.6	1, 3, 2, 2, 1	2.2
6	7	Fast	1000	100	0.3	3, 3, 2, 2, 4	3.4
7	1	Slow	1200	100	0.3	1, 1, 1, 2, 2	2.0
8	6	Fast	1200	100	0.6	2, 2, 4, 3, 2	3.2

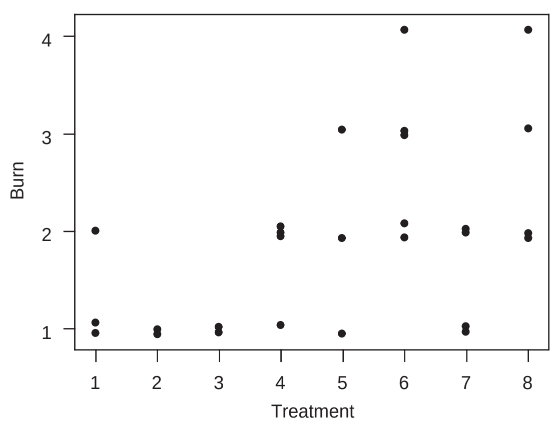


Figure 19.1 Burn by treatment with added vertical jitter.

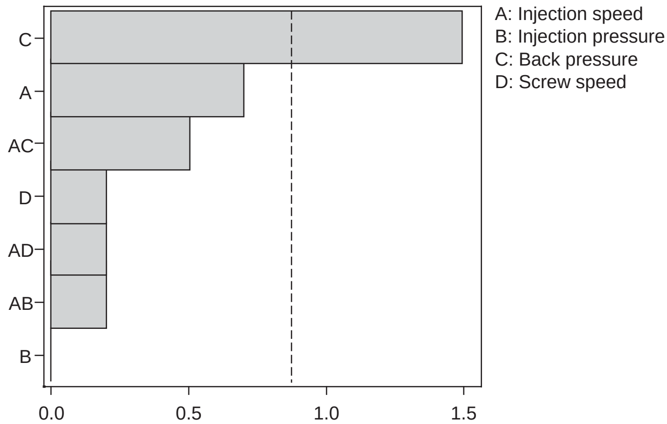


Figure 19.2 Pareto plot of effects on average burn score.

We use the average burn as the performance measure (that is, the *response* in MINITAB) for the analysis. We look for candidate settings that make the performance measure as small as possible.

Fitting a *full model* with all possible effects, we get the Pareto plot of the effects for the average burn score in Figure 19.2. We see that only factor *C* (back pressure) has a large effect. The team assumed the three-input interaction (*ABD*) confounded with *C* was negligible.

We show the main effect plot for factor *C* in Figure 19.3. The low level of back pressure gives lower burn scores on average.

The team decided to reduce the back pressure to 75 and leave the other fixed inputs at their original values. We finish this story in Chapter 21.

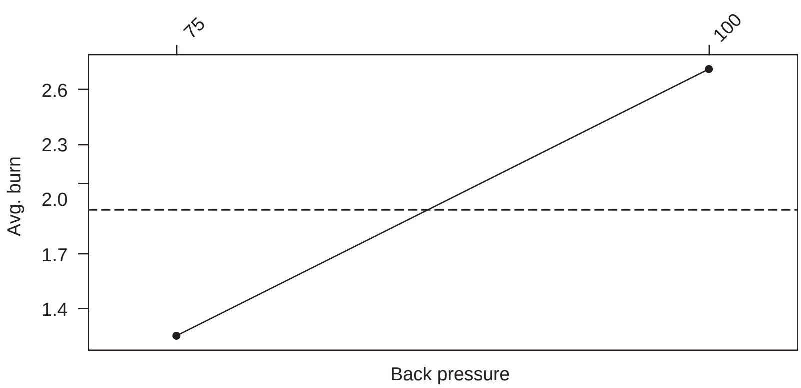


Figure 19.3 Main effect plot for back pressure.

Iron Silicon Concentration

In a project to reduce variation in the silicon concentration of molten cast iron, the team discovered that the existing measurement system was not acceptable. The discrimination ratio was 1.5 and the estimated *stdev(measurement)* was 0.33. The cause of the measurement variation was unknown but acted over the short term.

The measurement process has three major steps:

- Sample the molten iron and pour coins.
- Machine and polish the coins.
- Use a spectrometer to determine the concentration of silicon in the coins.

In the measurement system investigation, the team saw most of the measurement variation in the silicon concentration of coins poured consecutively from the same stream of iron, then prepared and measured together. Rather than search for the dominant cause of the measurement variation, the team decided to try to make the measurement process robust to the unknown cause. They knew that the spectrometer was highly repeatable and hence chose candidates from the other steps of the measurement process. Table 19.3 gives the five candidates and their levels.

Table 19.3 Candidates and levels for iron silicon concentration robustness experiment (level in current process given by *).

Candidates	Label	Low level	High level
Mold temperature	A	300°F	400°F*
Mold design	B	old*	new
Cut depth	C	skin	1/16 inch*
Surface finish	D	lathe*	polish
Sample temperature	E	70°F*	120°F

The definition of a run is a key stage in the planning of a robustness experiment. Since the coin-to-coin variation was dominant, the team decided to sample iron and prepare five coins as quickly as possible for each run. They expected the dominant cause of measurement variation would act within each run.

The team used the resolution III fractional factorial experiment with eight runs and treatments given in Table 19.4. The confounding pattern is:

```
Alias Structure (up to order 3)

A + BD + CE
B + AD + CDE
C + AE + BDE
D + AB + BCE
E + AC + BCD
BC + DE + ABE + ACD
BE + CD + ABC + ADE
```

Table 19.4 Treatments and results for iron silicon concentration robustness experiment.

Treatment	Run order	Mold temperature	Mold design	Cut depth	Surface finish	Sample temperature	Silicon measurements	Log(s)
1	1	300	Old	Skin	Polish	120	2.4, 2.5, 2.3, 2.2, 1.9	–1.55
2	6	400	Old	Skin	Lathe	70	2.8, 2.0, 2.1, 2.6, 2.2	–1.09
3	5	300	New	Skin	Lathe	120	2.2, 2.4, 2.3, 2.2, 2.6	–1.73
4	3	400	New	Skin	Polish	70	2.3, 2.0, 2.3, 2.3, 2.4	–2.03
5	8	300	Old	1/16	Polish	70	2.4, 2.1, 1.8, 2.2, 2.3	–1.51
6	7	400	Old	1/16	Lathe	120	1.9, 2.7, 1.9, 1.9, 2.4	–1.03
7	2	300	New	1/16	Lathe	70	2.5, 2.2, 2.7, 2.5, 2.7	–1.65
8	4	400	New	1/16	Polish	120	2.2, 2.4, 2.1, 2.5, 2.2	–1.79

A resolution III design can estimate main effects, assuming that two input interactions are negligible.

To conduct the experiment, the team manufactured two new molds and selected two old molds. They set the mold temperature to 300° for one of the new molds and to 400° for the second, and similarly for the two old molds. They then quickly sampled iron and poured 40 coins, 10 for each combination of mold design and temperature. Since all coins were produced from essentially the same iron, they assumed that the true concentration of silicon in each coin was the same. Next, the team prepared the 40 coins according to the experimental plan. The data are given in the file *iron silicon concentration robustness* and Table 19.4.

We are looking for a treatment combination that has little variation. As a first step in the analysis, we plot the measured silicon concentration by treatment as shown in Figure 19.4. We see that treatments 4 and 8 look promising since they have relatively little variation.

We summarize the performance of the measurement system for each run using log(s), where s is the standard deviation of the five measurements within each run. We use the log transformation to better meet the assumptions underlying the model for the data. The smaller the within-run variation, the smaller the performance measure. We show the values of the performance measure for each treatment in Table 19.4. As expected, treatments 4 and 8 have the smallest values. From the original investigation of the measurement system, the baseline performance is log (.33) = –1.11.

Figure 19.5 shows the Pareto chart of the effects when fitting a full model. The largest effects are the main effects of candidates B (mold design) and D (surface finish).

To draw conclusions, we summarize the results using the main effects plots given in Figure 19.6. With the performance measure log(s), smaller is better, so we see that switching to the new mold design and polishing the samples is beneficial.

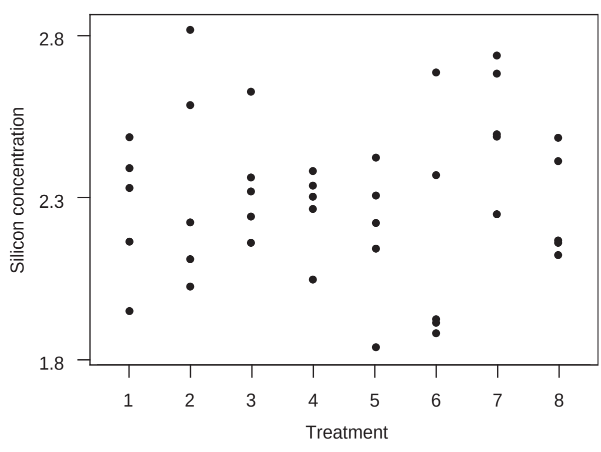


Figure 19.4 Silicon concentration by treatment.

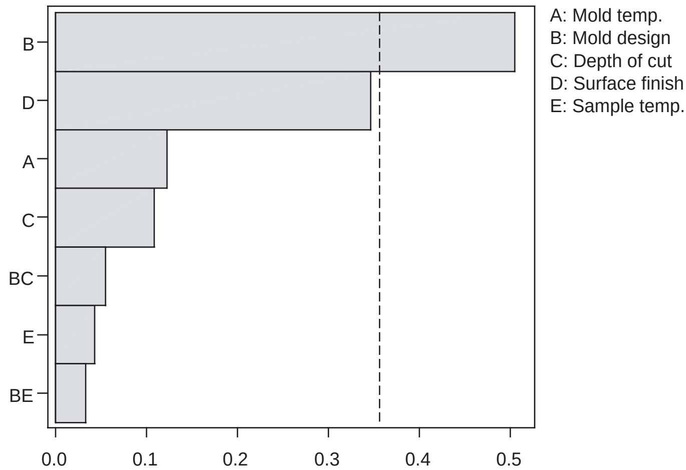


Figure 19.5 Pareto chart of the effects on log(*s*) for iron silicon experiment.

To validate that the performance of the measurement system would be improved with the new mold design and polished surface finish, the team carried out a simple investigation. They sampled 30 coins from the same stream of iron and measured the silicon concentration using the new levels for the mold design and surface finish and the current levels for the other candidates. Notice that they had not used this treatment in the experiment. The standard deviation of the 30 measurements was 0.15, substantially less then the baseline 0.33. With the new setting we expect the measurement discrimination ratio to increase to 3.3.

The team made a risky decision to use a single batch of iron in the robustness experiment. By using only one silicon concentration, there was a danger that the conclusions from the experiment would not generalize to the range of silicon concentrations seen in the process.

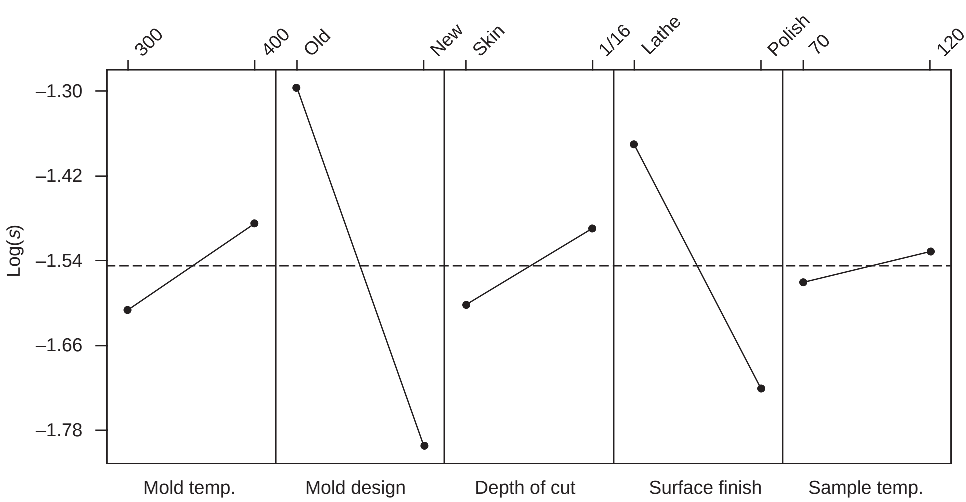


Figure 19.6 Main effects for iron silicon concentration robustness experiment.

Electroplating Pinskip Defect

In a chrome plating process, racks of plastic grills were dipped in a series of chemical baths. The baseline rate of defective grills was 7.3%, measured over several weeks of production. Using Pareto analysis, a team charged with reducing scrap and rework costs found over half the defective grills were due to pinskips, small voids in the plated surface. They decided to try to eliminate the pinskip defectives. The team did not employ the Statistical Engineering algorithm. Instead, they proceeded immediately to the process robustness approach.

They started by brainstorming a list of possible candidates. They selected the six candidates, given in Table 19.5.

Table 19.5 Candidates selected for the pinskip robustness experiment.

Candidates	Name
A	Tank 1 dip time
B	Tank 2 dip time
C	Tank 1 concentration
D	Tank 2 concentration
E	Tank 2 temperature
F	Tank 3 temperature

There was some confusion about the appropriate levels for the candidates, since they all varied somewhat in the normal process. For each candidate, the team decided to use the set point as prescribed in the control plan as one level (–1) and a new set point outside the normal range of variation as the second level (+1). They used their judgment to choose the direction of the change.

For convenience, they defined a run as a rack of 48 grills, all processed simultaneously. The performance measure for each run was the number of defective grills in the rack.

In the past, without much success, the organization had made various attempts to reduce the pinskip defect rate by changing the levels of one fixed input at a time. Now, the team hoped to find a helpful interaction among the candidates. Accordingly, they planned a resolution V experiment with 32 treatments that would allow for the separate estimation of all main effects and two input interactions. This was not possible with six candidates in a 16-run experiment. They planned to conduct the experiment over four days, giving the process time to settle down after changing the candidate levels. The candidates *C, D, E,* and *F* (the tank concentrations and temperatures) take the longest time to change, so they ordered the runs to minimize the number of changes of these candidates.

The experimental plan and the number of defective grills for each run are shown in Table 19.6 and are given in the file *electroplating pinskip defect robustness*.

Table 19.6 Electroplating pinskip experimental plan and data.

Treatment	A	B	C	D	E	F	Defectives	Order
1	–1	–1	–1	–1	–1	–1	0	1
2	+1	+1	–1	–1	–1	–1	0	2
3	–1	+1	–1	–1	–1	+1	1	3
4	+1	–1	–1	–1	–1	+1	1	4
5	–1	+1	–1	–1	+1	–1	0	5
6	+1	–1	–1	–1	+1	–1	1	6
7	+1	+1	–1	–1	+1	+1	1	7
8	–1	–1	–1	–1	+1	+1	0	8
9	+1	–1	–1	+1	–1	–1	0	9
10	–1	+1	–1	+1	–1	–1	2	10
11	–1	–1	–1	+1	–1	+1	1	11
12	+1	+1	–1	+1	–1	+1	2	12
13	+1	+1	–1	+1	+1	–1	3	13
14	–1	–1	–1	+1	+1	–1	2	14
15	+1	–1	–1	+1	+1	+1	7	15
16	–1	+1	–1	+1	+1	+1	12	16
17	–1	+1	+1	–1	–1	–1	7	17
18	+1	–1	+1	–1	–1	–1	1	18
19	+1	+1	+1	–1	–1	+1	3	19
20	–1	–1	+1	–1	–1	+1	1	20
21	–1	–1	+1	–1	+1	–1	0	21
22	+1	+1	+1	–1	+1	–1	0	22
23	+1	–1	+1	–1	+1	+1	0	23
24	–1	+1	+1	–1	+1	+1	2	24
25	+1	+1	+1	+1	–1	–1	1	25
26	–1	–1	+1	+1	–1	–1	1	26
27	–1	+1	+1	+1	–1	+1	2	27
28	+1	–1	+1	+1	–1	+1	0	28
29	–1	+1	+1	+1	+1	–1	0	29
30	+1	–1	+1	+1	+1	–1	1	30
31	+1	+1	+1	+1	+1	+1	1	31
32	–1	–1	+1	+1	+1	+1	1	32

The experiment was more time-consuming than expected. The team could not complete the planned eight runs in each day. They found the tank temperatures and concentrations hard to change quickly. In the end, the experiment was conducted over five days.

From Table 19.6, there are many promising treatments with no pinskip rejects in the rack. Using MINITAB and a full model, we get the Pareto chart of effects shown in Figure 19.7. There are three related interactions (*CD, CE,* and *DE*) that are large. We present a cube plot in Figure 19.8 for the candidates *C, D,* and *E*. The plot gives the average number of grills rejected per rack for the eight combinations of the three candidates.

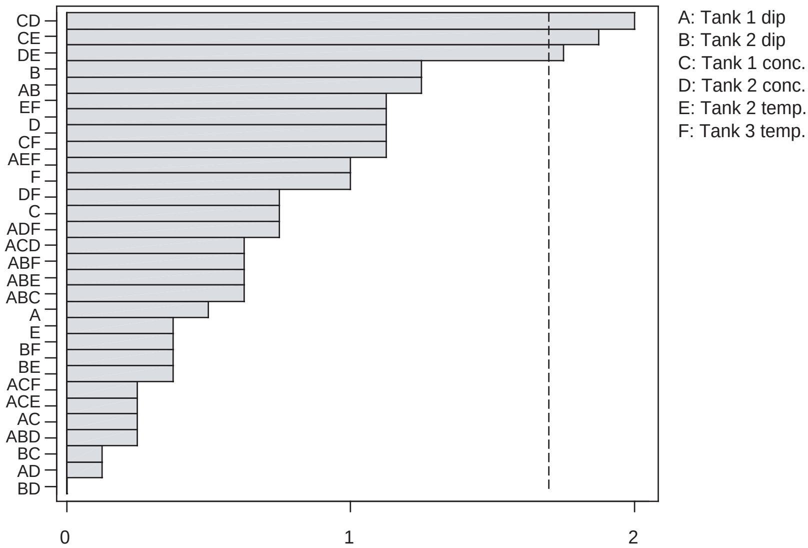


Figure 19.7 Pareto analysis of effects in pinskips experiment.

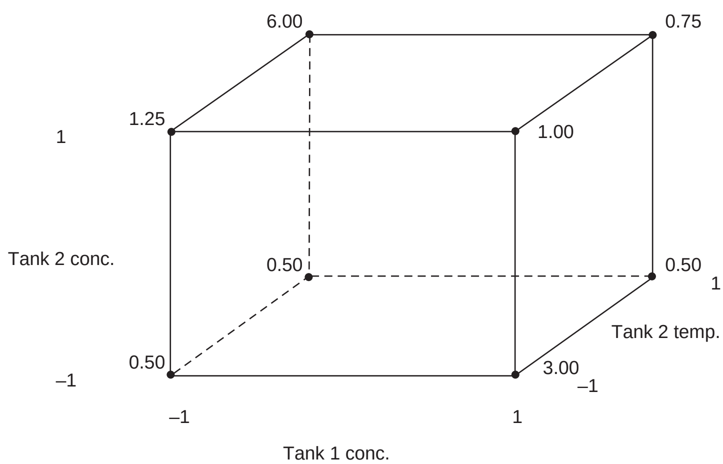


Figure 19.8 Cube plot of average number of pinskip rejects per run for tank 1 concentration, tank 2 concentration, and tank 2 temperature.

The lower the number of rejected grills the better. From the cube plot, the team was surprised to see that the process at the current levels (shown as the lower left-hand point in the plot) had an average of 0.5 rejects per rack, the best observed performance (tied with two other combinations). This was surprising since the current process averaged about 1.6 defective grills per rack.

The team concluded that there was no reason to change the process settings based on the results of the experiment. They decided to abandon the project.

We cannot say for certain what went wrong. One possibility is that none of the selected candidates interacted with the unknown dominant cause. Alternately, if we examine Figure 19.9, a plot of the number of defective grills in the rack versus the order of the runs, we see that there was a large burst of defectives during the middle portion of the experiment. If the dominant cause acts in the time-to-time family, then since each run consisted of a single rack, there was little chance of the cause acting within each run and hence no chance to see if any of the candidates could make the process robust. Defining a run as a single rack of grills is not appropriate if the dominant cause acts in the time-to-time family.

The lesson here is that the team made a number of poor decisions. They should not have jumped directly to the robustness approach. They had little process knowledge to help choose the candidates. They had no assurance that the dominant cause would act within each run or even in the week used for the experiment. They would have been better off first investigating the nature of the process variation and generating more clues about the dominant cause before selecting any particular variation reduction approach.

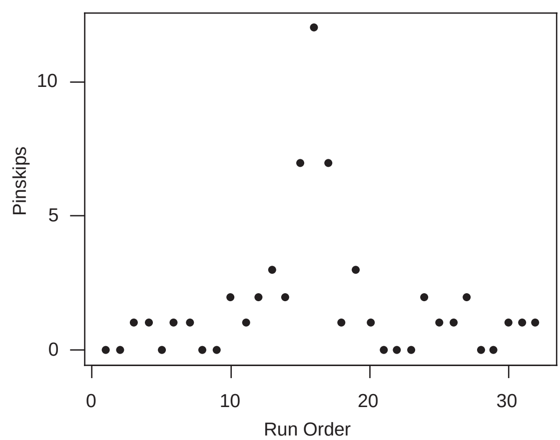


Figure 19.9 Number of defective grills versus run order.

Robustness Experiment Summary

Select candidates and levels based on knowledge of the process.

Question

Which, if any, combination of the candidate levels reduces the sensitivity of the process to variation in unknown dominant causes?

Plan

- Define a run long enough to see the full extent of variation in the output in the existing process.
- Determine the number of parts (repeats) to be measured within each run and the performance measure(s).
- Determine the number of runs.
- Select a fractional factorial design of resolution III or higher.
- Randomize the order of the runs as much as is feasible.
- Make everyone potentially impacted aware of the plan.

Data

Carry out the experiment. Record the output values, the levels for candidates, treatment number, and run order, one row for each run. Use a separate column for each repeat.

Analysis

- Plot the output by treatment to look for promising treatments. To make this plot, temporarily arrange the data with a separate row for each repeat.
- Calculate the performance measure(s) across the repeats for each run.
- Use a full model and a Pareto chart to analyze the performance measure(s) looking for large main and interaction effects.
- For large effects, construct main and interaction effects plots.

Conclusion

Identify the levels of the candidates that lead to the best performance.

19.2 ASSESSING AND PLANNING PROCESS ROBUSTNESS

To successfully implement robustness, we must:

- Identify the fixed inputs and their levels (the new process settings) that make the process robust to variation in the unknown dominant cause(s).
- Move the process center if required.

- Check that the change of settings does not produce substantive negative side effects.
- Estimate the costs of changing the settings and the new ongoing operating costs.
- Estimate the benefit of the change of settings.

If we can accomplish all these tasks, and the benefits outweigh the costs, we proceed to the validation stage of the algorithm.

Since we have little or no knowledge of the dominant cause, we are given little guidance on what candidates to select and how to pick their levels. We recommend using as many candidates as feasible and a fractional factorial design.

The definition of a run is a key step in planning the robustness experiment. To identify settings that reduce the effect of the unknown dominant cause, we need the cause to act within each run. To define a run, we can use knowledge of the time pattern of output variation from a multivari, baseline, or other investigations. In the iron silicon concentration example, the team knew that measuring the same coin several times in a short period of time would show most of the variation in the measurement system. They defined a run to be five consecutive measurements on coins poured from the same iron. In the electroplating pinskip defect example, the team had no knowledge of the time family containing the dominant cause. They specified a run as a single rack of grills. This would have been an appropriate choice if the dominant cause acted within a rack. However, if the dominant cause acted in the rack-to-rack family or slowly over time, then with this definition of a run, the experiment was doomed to fail.

If the time-based family in which the dominant cause acts is unknown, we recommend a multivari investigation before proceeding with a robustness experiment. When the dominant cause acts slowly, we need to have long runs so that the cause acts within each run. Long runs add to the cost and complexity of the experiment, so in this case, determining robust process settings may be infeasible. On the other hand, feedback control may be feasible.

For each run in the experiment, we calculate a performance measure to assess the behavior of the process output within the run. We define the performance measure based on the goal of the problem. There are many possible performance measures.[1] In the iron silicon concentration example, the performance measure was the standard deviation (actually the logarithm of the standard deviation) of the output values measured during the run. The problem goal was to reduce the measurement system variation. In the crossbar dimension example, the team chose the average burn score over the run to measure process performance. By lowering the average score, the team hoped to address the problem goal of reducing the frequency of burn defects. The team could have selected an alternate performance measure such as the proportion of parts scrapped due to burn (as in the electroplating pinskip defect example) in each run. With this choice they would have required longer runs since the output is binary.

We can define and analyze several performance measures (for example, average and standard deviation) within the same experiment. For instance, to make a measurement system more robust, we may simultaneously analyze bias and measurement variation as two performance measures. In the camshaft lobe runout example, introduced in Chapter 1, the goal was to reduce the average and variation in runout. If the team had adopted the robustness approach, in the experiment, they would have calculated both these performance measures within each run. With two or more performance measures we may be forced to make a compromise in the choice of settings.

We need to have enough repeats within each run to get a good estimate of the process performance. It is hard to give a firm rule, but more repeats are better. If measuring the output is expensive we can use a relatively long run but measure a sample of parts within the run.

After the experiment, we can assess the costs and benefits of the approach. We can estimate the performance measure for the new process settings. There are costs associated with changing the candidate levels and the ongoing operating costs at the new levels. We also need to check for negative side effects. Changing candidate levels may shift the center of the process in an undesirable direction. In that case there may be additional costs related to finding an adjuster and operating the process at a different level of the adjuster.

The robustness approach is often selected to reduce the rate of defectives. In the robustness experiment for a binary output, we need a run to be long enough so that each run will likely contain one or more defective parts. If defectives are rare, we may be able to aggravate the process to increase the defect rate for the purposes of the experiment. Then we hope that results obtained under the aggravated conditions are relevant for the standard process. For example, there were field failures of exterior electrical boxes after several years due to corrosion. The team conducted a robustness experiment on the painting process using scored panels in a salt spray chamber (a highly aggravated condition) to see if changing fixed inputs would increase the durability.

There is a strong connection between the robustness approach and the desensitization approach discussed in Chapter 16. In the latter case where we control the known dominant cause in the experiment, we can determine the interactions between the candidates and the cause directly in the experiment. With the robustness approach, we can only observe the interaction indirectly through the performance measure.

There is also a connection between the robustness and moving the process center approaches. In the electroplating pinskip defect example, the output was binary, and the goal of the problem was to reduce the proportion of defective grills. We can view the approach taken as either moving the process center or robustness. In either case, we search for changes to fixed inputs to achieve the goal without knowledge of a dominant cause.

We recommend desensitization over robustness if at all possible. In other words, we recommend first finding the dominant cause of the variation. We have had little success applying the robustness approach. In most circumstances one of the other variation reduction approaches is preferred. Selecting the robustness approach is a last hope.

The idea of analyzing a performance measure such as the within-run standard deviation was first suggested by Bartlett and Kendall (1946). Nair and Pregibon (1988) give a motivation for using log(s). Taguchi (1986) popularized process robustness and called it *parameter design*.

Key Points

- We make a process robust by changing fixed inputs to reduce the effects of unknown dominant cause(s).

- Selecting candidates for a robustness experiment is difficult due to the lack of knowledge of the dominant cause. We recommend using as many candidates as feasible in a fractional factorial design.
- We must define a run of the experiment to be long enough so that we would see the full extent of variation in the output *if* no process changes were made. We can then be assured that the unknown dominant cause acts within each run.
- We select a performance measure to reflect the goal of the problem. We calculate the value of the performance measure for each run of the experiment and use it as the response in the analysis.
- Making a process robust requires a one-time change to fixed inputs such as the product or process design or the process control plan.

Endnote (see the Chapter 19 Supplement on the CD-ROM)

1. Taguchi popularized the robustness approach. We explore his choice of performance measures, experimental plans, and analysis methods in more detail in the supplement.

Exercises are included on the accompanying CD-ROM

20

100% Inspection

If you put off everything till you're sure of it, you'll get nothing done.

—Norman Vincent Peale, 1898–1993

The simplest yet most controversial variation reduction approach is 100% inspection. We compare the value of the output characteristic of each part to inspection limits. We then scrap, downgrade, or rework any part with output value outside the inspection limits. Figure 20.1 shows how adding inspection limits reduces the output variation of accepted parts. The inspection limits are tighter than the customer-driven specification limits. To use 100% inspection, we do not need to know the dominant cause or understand the nature of the process variation.

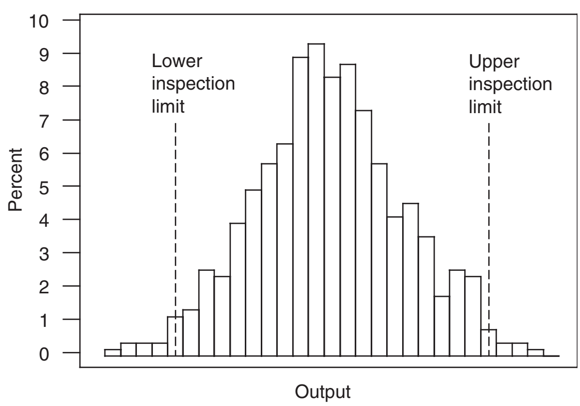


Figure 20.1 Reducing variation by adding inspection limits.

The requirements for 100% inspection to reduce variation are minimal. We need a measurement system for the output characteristic with little measurement bias and variation.

The costs associated with implementing 100% inspection are:

- The cost of measuring every part
- The cost of dealing with rejected parts

We can use the baseline investigation to help assess the second cost. For a continuous output, we use the baseline histogram to estimate the increase in the number of rejects due to adding or tightening inspection limits. 100% inspection is more effective for eliminating outliers than for reducing the standard deviation.

20.1 EXAMPLES OF 100% INSPECTION

We give several examples of the use of 100% inspection to reduce variation. The most common application is to a reformulated problem, where we apply 100% inspection to reduce variation in the dominant cause. In this context, 100% inspection is a form of error proofing.

Manifold Blocked Ports

We introduced the problem of blocked ports in cast-iron exhaust manifolds in Chapter 3. The foundry reacted to a customer complaint that a blocked port had been found on an engine. One step in the current process was the dropping of a ball bearing through the ports of each manifold to ensure that there was no blockage. This manual inspection was sometimes skipped or poorly done.

The team had difficulty establishing a baseline since the defect was so rare. They knew that blocked ports resulted from cracking in the cores during the pouring process, but they had no idea how to determined the cause of core cracking. None of the usual Statistical Engineering tools were helpful, because they could find only two manifolds with a blocked port over a one-month period.

The team decided to automate the inspection to keep the rare defective parts from being shipped. They investigated whether an ultrasound system could correctly classify ports as blocked or not. They were careful in the selection of parts for the investigation. They wanted a sample of manifolds with blocked ports that would demonstrate all conceivable ways a port could be blocked or even partially blocked. This was difficult since blocked ports were so rare. They had only the two manifolds with naturally blocked ports. To get a larger sample, they deliberately created blocked ports in 13 manifolds. They also selected 30 manifolds from regular production that had no blocked ports.

There were three fixed inputs in the ultrasound system. The team planned a full factorial experiment with eight runs. For each run of the experiment, the 45 manifolds were inspected 10 times each and classified as blocked or not. A plan with less chance of study error would use more parts and measure each part once.

For each run, the team counted:

- n1—the number of times a manifold with a blocked port was passed by the system
- n2—the number of times a manifold without a blocked port was rejected by the system

The data are:

Treatment	A	B	C	n1	n2
1	Low	Low	Low	1	7
2	Low	Low	High	2	9
3	Low	High	Low	3	3
4	Low	High	High	1	6
5	High	Low	Low	0	2
6	High	Low	High	0	12
7	High	High	Low	2	13
8	High	High	High	1	7

Since passing a manifold with a blocked port was a serious error, the team considered only treatments 5 and 6 where all manifolds with blocked ports were detected. They selected treatment 5 since it had a lower rate of rejecting manifolds without a blocked port. They changed the operation and control plan so that:

- The ultrasound system inspected all manifolds and automatically rejected any classified as having a blocked port.
- All rejected manifolds were inspected by hand to determine if a port was blocked or not.
- At the start of each shift, one of the manifolds with naturally blocked ports was inspected by the ultrasound system to ensure it was functioning properly.

In this example, defects were very rare, so the costs were limited to the cost of inspection. The inspection cost was less than the loss of goodwill when the customer found manifolds with blocked ports.

Drain Hole Cracks

In a stamping process, the baseline rate of cracks and splits in axle cover drain holes was 3%. There was 100% inspection by the press operators coupled with an off-line audit of a sample from each basket of parts. If a defective part was found in an audit, all baskets filled since the previous audit were contained and reinspected. Despite the inspection effort, there were frequent complaints from the customer about defective stampings.

The team suspected that the dominant cause of the cracks and splits defect was in the incoming steel and thus outside their control. Without further investigation, they decided to 100% inspect parts using digital cameras and image processing to detect the defects. The cost of the inspection equipment was substantial but operating costs were low.

With the new system, there was no need for the audit of inspected parts or for containment. There was an increase in scrap and rework costs. However, there were no further complaints from the customer about cracks and splits.

Broken Tooling

In a machining process, the process engineer discovered that the dominant cause of broken drills in an automated drilling operation was improperly machined parts from upstream in the process. He found that these parts occurred when the production line restarted after a shutdown and decided to use 100% inspection before the drilling operation. He installed a limit switch that detected the poorly machined parts and stopped the line when such a part was found before the drills were damaged. The cost of the inspection was low, and there were substantial savings in tooling and downtime.

This example is classic *error proofing,* 100% inspection applied to the dominant cause.

Rocker Cover Oil Leaks

In an engine assembly plant, management assigned a team to eliminate oil leaks around the rocker cover. The team used repair data to estimate that about 1.4% of the engines had such oil leaks detected at the final test stands.

There were two rocker covers per engine purchased from an outside supplier. The covers were caulked and then bolted to the engine. In the past, the plant had blamed the supplier whenever oil leaks caught the attention of management.

The team decided to look for the dominant cause using a group comparison. They set aside six engines with rocker cover oil leaks. Since each engine had two rocker covers and only one leaked, the team had six leaking and six nonleaking covers. They then measured a large number of characteristics on each of the 12 covers. However, they could not find any characteristic of the rocker cover that separated the leaking and nonleaking sides of the engine. After some thought, they decided to examine the whole process, including the assembly operation, using a component swap investigation. They were surprised to discover that the leaking engines no longer leaked after reassembly. After further investigation, they found the dominant cause was poor torque on the bolts that held the rocker cover to the engine.

They decide to put 100% automated inspection on the torque guns to signal if a rocker cover bolt was not properly tightened. This change virtually eliminated oil leaks around the rocker cover at the final test stands.

20.2 ASSESSING AND PLANNING 100% INSPECTION

To successfully implement 100% inspection, we must:

- Select appropriate inspection limits.
- Measure (with low bias and variation) the output for every part produced.
- Deal with rejects from the inspection.

- Estimate the costs of measurement, loss of volume, and dealing with rejects.
- Estimate the benefits.

If we can accomplish all these tasks, and the benefits outweigh the costs, we proceed to the validation stage of the algorithm.

Many processes already have 100% inspection because the shipment of a defective or out-of-specification part has critical consequences for the customer. For a part such as the exhaust manifold, we consider improving the existing inspection system. In the rocker cover example, the team added a 100% inspection system to the process to check for proper torque.

We can apply 100% inspection if the output can be compared to the inspection limits or standards (sometimes called *boundary samples*) with small measurement error. The key prerequisite is a stable, nondestructive measurement system with low variation and bias.

100% inspection does not eliminate defects or out-of-specification parts. We need to change the control plan to deal with the rejects. In one disastrous example, rejected parts were set aside in a box to be reworked. However, because of poor organization and labeling, the box was shipped to a customer.

Applying 100% inspection to the output is rarely the most cost-effective approach. There is less variation in the parts shipped to the customer, but there are increased costs due to the larger number of rejects. We think 100% inspection is best suited for situations where other variation reduction approaches have not proven feasible or effective. Two potential applications are processes with rare defects or when inspection and rework costs are low. In the manifold blocked port example, the defect was so rare that the team could not discover the dominant cause of the defect using empirical methods. The only feasible approach was to design an effective low-cost 100% inspection system to detect the rare defect.

Applying 100% inspection to a dominant cause is called *source inspection* (Shingo 1986). See Shimbun (1988) for numerous examples. In this case, the costs of dealing with the parts that fall outside the inspection limits on the cause are likely to be small, since we are applying the inspection upstream from the final output. In the rocker cover oil leak example, it was much cheaper to detect and repair the poorly tightened bolts than to deal with the leaks after the engine had been assembled.

We often go directly to source inspection on a dominant cause without reformulating the problem in terms of the cause. This is an example of the Fix the Obvious approach described in Chapter 14.

Most successful applications of 100% inspection use automated measurement. Human inspectors make mistakes at the best of times, and with rare defects, they are unable to remain focused. Using multiple inspectors is not a solution since each inspector may become complacent with the belief that any problem will be found by one of the others.

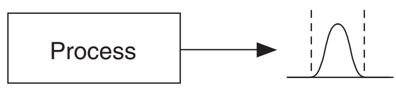
100% inspection is poorly regarded as a variation reduction approach in the quality improvement literature. For example, one of Deming's 14 points exhorts industry not to rely on mass inspection to "control" quality (Deming, 1992). Despite its unfavorable image, we see 100% inspection applied frequently because immediate process improvement is required and no other improvement approach is feasible without more process knowledge and investigation.

A common modification of 100% inspection is inspection sampling where not every part is measured. One alternative is to define lots that are accepted or rejected based on the quality of a sample taken from the lot. Accepted lots are shipped, and rejected lots are 100%

inspected or otherwise disposed. If we know that lot-to-lot variation is large relative to within-lot variation, inspection sampling can reduce variation. We can improve an existing inspection sampling scheme by redefining a lot, changing the inspection limits, or changing the lot acceptance criteria. Compared to 100% inspection, inspection costs are smaller. However, overall variation will not be reduced to the same degree. If the dominant cause acts part to part, inspection sampling is a poor approach. This is obvious for processes with rare sporadic defectives. Deming (1992, chap. 15) showed that if the process is stable, either no inspection or 100% inspection is optimal. See also VanderWiel and Vardeman (1994).

Key Points

- We use 100% inspection when defects are critical and rare, so that it is difficult and costly to determine a dominant cause, or when the inspection costs are very low.
- 100% inspection applied to a reformulated problem is one form of error proofing.



- Most successful applications of 100% inspection rely on automated inspection since human inspectors make mistakes.

21

Validating a Solution and Holding the Gains

An investment in knowledge always pays the best interest.

—Benjamin Franklin, 1706–1790

We have now reached the final stage of the Statistical Engineering algorithm. We have a proposed solution to the problem, a change to one or more fixed inputs of the process. We may propose changes to process settings, the addition or deletion of a process step, the use of a new supplier, changes to the process control plan, and so on. Although this sounds silly, we remind you that the process cannot be improved without making changes to fixed inputs.

There are two remaining tasks:

- *Validate* the proposed solution to see that the goal is met and that there are no substantial negative side effects.
- If the solution is validated, ensure that the process change is made permanent so that the gains are preserved.

21.1 VALIDATING A SOLUTION

We must validate a solution to ensure that under the proposed process changes, the process performance meets the goal of the project or, at the least, that the benefits outweigh the costs. Ideally, we should repeat the baseline investigation to compare the performance before and after the process change. We also need to check carefully for unexpected negative side effects due to the changes in the process.

Depending on the costs and risks, we may initially use a small investigation for a preliminary validation before conducting a full investigation. We do not want to make it too difficult to reverse changes to the process or product until we are confident the expected benefits will materialize. During the validation investigation, we monitor the data as they are being collected. There is no sense in continuing if it becomes clear that the proposed solution is ineffective.

In many problems, we assess the baseline performance of the process using historical data collected over weeks and months. We cannot afford to wait that long to validate the solution. We can use knowledge we have gained in the process investigations to suggest how long we need to observe the original process until we would see the full extent of variation. We plan the validation investigation to last at least that long so that we expect to see the new full extent of variation with the proposed process changes.

We may find a proposed solution inadequate for a number of reasons related to taking shortcuts in the algorithm with the hope of saving time and money. First, the solution may be based on a cause that is not dominant. In that case, the process improvement will be small (see Chapter 2). If the algorithm was followed closely, this cannot occur, since the algorithm requires verification of the dominant cause. Second, we may not meet the goal because of optimistic assumptions. For example, we assume a perfect adjustment method in assessing feedback control that is not realized in practice.

If the solution leads to an improvement that is not sufficient to meet the goal, we need to decide whether or not to continue with the implementation. We can make the decision by reevaluating costs and benefits. To meet the goal, we need to go back to reconsider the possible approaches and look for further changes to fixed process inputs.

If we have reformulated the problem, we must assess the process change against the baseline for the original output. We will have made assumptions about the links between the output and the dominant cause based on uncertain knowledge. When we validate using the original output, we can check that these assumptions were correct.

We give two examples of validation.

Crossbar Dimension

In the production of an injection-molded contactor crossbar, the problem was excessive variation in a crossbar dimension. In the baseline investigation, the team established the full extent of variation of the dimension (measured as the deviation from nominal) as –0.3 to 2.3 thousandths of an inch, with standard deviation 0.46. After some investigation (see Chapter 12), they discovered that barrel temperature was the dominant cause of variation. The team proceeded using the desensitization approach (see the Chapter 16 exercises). They found that increasing the barrel temperature set point reduced the effect of the variation in barrel temperature. Since this change increased the crossbar dimension center, they used a known adjuster to reduce the average dimension to the target value zero.

In a preliminary validation investigation, the team found that increasing the average barrel temperature resulted in burn defects. They did not search for a dominant cause of the burn defect. Instead, they used the robustness approach to find process settings that eliminated the defect while at the same time allowing the increased barrel temperature. See Chapter 19 for further discussion.

The team next proceeded to a full validation with the proposed process settings. In the validation investigation, 300 parts were selected over two shifts. This plan matched the baseline investigation. The team measured the crossbar dimension and inspected each part for the burn defect. The data are given in the file *crossbar dimension validation*. The histogram of the crossbar dimension from the validation investigation, given in Figure 21.1, shows the reduced variation. The standard deviation in the crossbar dimension was reduced to 0.23, and the burn defect occurred on only 2 of the 300 parts. With the new settings, the process performance met the project goal.

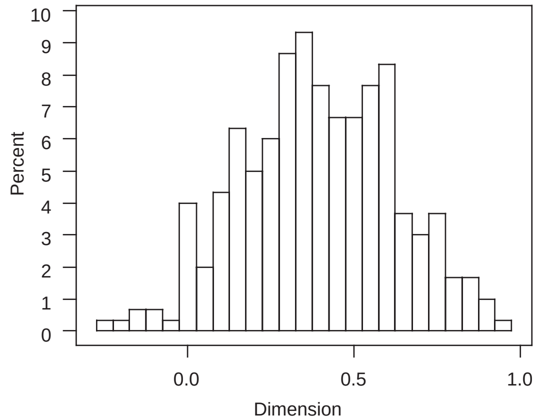


Figure 21.1 Histogram of crossbar dimension in the validation investigation.

Engine Block Porosity

In the engine block porosity example introduced in Chapter 3, the goal was to reduce scrap due to porosity found on a machined surface of an engine block. The baseline scrap rate was about 4%, based on several months of data. There was no formal numerical goal. The team devised a porosity score used throughout the problem solving. They did not conduct a baseline investigation with this new score.

To solve the problem, the team recommended a new core wash supplier. They proceeded with validation in two steps. First, they switched to the new core wash for a single shift and evaluated the results, which were very promising. There were no obvious side effects. They then used the new wash for a week, again monitoring the process carefully for porosity defects and unexpected side effects. The scrap rate due to bank face porosity was less than 1%. The team committed to proceeding with the new core wash. We can see the long-term improvement in Figure 21.2 since the change was made in month 13. The cost savings were several hundred thousand dollars per year.

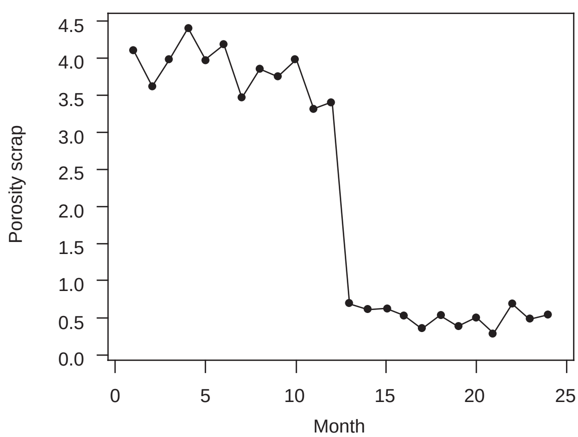


Figure 21.2 Porosity scrap rate by month.

21.2 HOLDING THE GAINS OVER THE LONG TERM

After validating a solution and implementing a process change, we need to ensure that the improvement is preserved over the long term. Most aspects of holding the gains are management issues outside the scope of this book.

In our experience, there are two imminent dangers:

- The recommended process change is not fully implemented or later reversed.
- Other fixed inputs are (later) changed in a way that reduces the effectiveness of the solution.

The difficulty in ensuring the implemented change is not reversed depends on the nature of the change. To change the design of a product or process, we should adhere to formal design change procedures that can be expensive and take considerable time. Such changes are not easily reversed. Changes to the process control plan, on the other hand, are easy to make but also easy to undo or ignore. In these cases, we need to document the changes and ensure that appropriate training is conducted. We also recommend process audits to verify that control plan changes continue to be used.

In many processes, fixed inputs change slowly over time due to wear or aging of equipment. We have discussed several problems where the obvious fix requires maintenance activity. Because at the time the problem has a high profile, the maintenance is carried out and the variation is reduced. Over time, as the original problem is forgotten, maintenance effort decreases and the projected gains slip away. In such cases, the team can increase the chance of holding the gains by implementing a formal monitoring procedure. This is a simple form of feedback control added to the proposed solution. We can use a summary of process performance over time such as a run chart, a regular calculation of process capability, or a control chart (Ryan, 1989; Montgomery, 1996). In all cases, we need a plan to act if evidence from the monitoring suggests the gains are disappearing.

At the end of a project, we need to preserve any new knowledge of process behavior that can be of use to a wider audience. We need to document any process or design changes with the supporting reasons for the change. We do not want improvements undone during future cost-reduction exercises or in the solution to another problem.

We can retain information by:

- Logging design changes in a design guide
- Documenting of all variation reduction projects on a searchable company intranet

Documenting projects and problems is useful but expensive and time-consuming. In our experience most process engineers do not enjoy this activity. See knowledge management books such as Davenport et al. (1997) and O'Dell et al. (1998) for suggestions.

We give three examples that illustrate the issues.

Truck Pull

As described in Chapter 10, in the early phases of the problem to improve the truck alignment process, the team examined right caster data stratified by the four alignment measurement machines that operated in parallel. The team was surprised to see the persistent material differences among averages of the four gages. Because trucks were assigned an alignment gage haphazardly, the gage was a cause of variation in right caster.

The team took immediate action to recalibrate the four gages to remove the systematic differences. To prevent a recurrence, they established a monitoring program to compare the daily averages of each measured characteristic for each gage. If they found significant differences, they recalibrated the four gages. Without such monitoring, there is every reason to believe that the problem would recur.

Fascia Dimension

At a fascia supplier, management assigned a team to address complaints from their customer, a car assembly plant, about difficulties installing the fascias. After consulting with the customer, the team determined that the average of a key dimension was off target and that there were too many large fascias. We discussed this example in the exercises to chapters 6 and 15.

From the baseline investigation, the team estimated the dimension average, measured from nominal, to be 7.3, with a full extent of variation of 2 to 12. After some further investigation, they proposed changing the cure and cycle times in the molding process to move the process center. They made changes to the control plan and carried out a short validation investigation. The average dimension was reduced to 3.1, with a range of 1 to 6. Within a week in actual production, the new control plan was being ignored. The line operators realized that the proposed solution was not as effective as initially thought.

Mistakes were made in this problem. The main difficulty was that the baseline investigation was conducted over too short a time period and did not capture the full extent of output variation. The team also failed to notice an upward drift in the dimension over the course of the baseline investigation. Subsequently, the validation investigation also covered too short a time period. Because of the poor plan and analysis for the baseline investigation, the team was led to an inappropriate approach and solution.

Wheel Bearing Failure Time

Based on a Pareto analysis of warranty claims, management assigned a team to solve the problem of premature failure of wheel bearings within the warranty period of three years. The goal was to reduce the failure rate from 3% to less than 0.3%. Most of the wheel bearing failures occurred in trucks driven primarily on dirt roads in wet and cold conditions. The team used the robustness approach to find a design change to the bearing assembly that increased the average failure time. During the investigation, the team discovered the failure rate had increased after the removal of a dust shield, recommended by an earlier cost-reduction program. No one had examined the potential costly side effects of removing the dust shield, despite a design change protocol that required the assessment of such side effects before the change could be implemented.

21.3 COMPLETING THE IMPLEMENT AND VALIDATE STAGE

We summarize the key tasks necessary to complete the Implement and Validate Solution and Hold the Gains stages of the Statistical Engineering algorithm. We must:

- Implement the solution (that is, change one or more fixed inputs).
- Conduct a validation investigation using the original output to show:

  —The problem goal has been met.

  —There are no substantial negative side effects.
- Establish monitoring and audits to preserve the change, if necessary.
- Document any changes to the product or process design or to the control plan.
- Document the lessons learned.

After completing all these tasks, we can celebrate our success!

Key Points

- We need to validate process changes to ensure the long-term reduction of process variation.
- In the solution validation, watch for unexpected negative side effects on other output characteristics.
- To have a lasting impact, the implemented change needs to be made permanent.

Exercises are included on the accompanying CD-ROM

References

Abraham, B., and J. Ledholter. 1983. *Statistical Methods for Forecasting.* New York: John Wiley & Sons.

Agrawal, R., 1997. Variation Transmission in Multistage Industrial Processes. Statistics Ph.D. thesis, University of Waterloo, Ontario, Canada.

Agrawal, R., J. F. Lawless, and R. J. MacKay. 1999. "Analysis of Variation Transmission in Manufacturing Processes, Part II." *Journal of Quality Technology* 31:143–154.

Amster, S., and K. L. Tsui. 1993. "Counterexamples for the Component Search Procedure." *Quality Engineering* 5:545–552.

Asher, R. K., Sr. 1987. "Statistical Process Control." *Products Finishing* 51:20–22.

Aström, K. J. 1970. *Introduction to Stochastic Control.* Mathematics in Science and Engineering Series. Vol. 70. New York: Academic Press.

Aström, K. J., and B. Wittenmark. 1989. *Adaptive Control.* Reading, MA: Addison-Wesley.

Automotive Industry Action Group (AIAG). 1995a. *Measurement Systems Analysis.* Second edition. Southfield, MI: AIAG.

Automotive Industry Action Group (AIAG). 1995b. *Statistical Process Control.* Second edition. Southfield, MI: AIAG.

Automotive Industry Action Group (AIAG). 1998. *Quality Systems Requirements QS-9000.* Third edition. Southfield, MI: AIAG.

Bartlett, M. S., and D. G. Kendall. 1946. "The Statistical Analysis of Variance: Heterogeneity and the Logarithmic Transformation." *Journal of the Royal Statistical Society. Series B*, 8:128–138.

Bhote, K. R., and A. K. Bhote. 2000. *World Class Quality.* Second edition. Saranac Lake, NY: American Management Association (AMACOM).

Bisgaard, S., and H. Fuller. 1995a. "Sample Size Estimates for Designs with Binary Responses." *Journal of Quality Technology* 27:344–354.

———. 1995b. "Analysis of Factorial Experiments with Ordered Categories as the Response." *Quality Engineering* 8: 199–207.

———. 1996. "Correction to 'Sample Size estimates for Designs with Binary Responses' (1995V27 p344-354)." *Journal of Quality Technology* 28: 496.

Box, G.E.P. 1988. "Signal to Noise Ratio, Performance Criteria, and Transformations (with discussion)." *Technometrics* 30:1–40.

———. 1999. "Statistics as a Catalyst to Learning by Scientific Method Part II: A Discussion." *Journal of Quality Technology* 31: 16–29.

Box, G.E.P., W. G Hunter, and J. S. Hunter. 1978. *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building.* New York: John Wiley & Sons.

Box, G.E.P., and G. M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control.* Revised edition. Englewood Cliffs, NJ: Prentice-Hall.

Box, G.E.P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time Series Analysis: Forecasting and Control.* Third edition. Englewood Cliffs, NJ: Prentice-Hall.

Box, G.E.P., and T. Kramer. 1992. "Statistical Process Monitoring and Feedback Adjustment: A Discussion." *Technometrics* 34:251–285.

Box, G.E.P., and A. Luceno. 1997. *Statistical Control by Monitoring and Feedback Adjustment.* New York: John Wiley & Sons.

Boyles, R. A. 2001. "Gage Capability for Pass-Fail Inspection." *Technometrics* 43: 223–229.

Brassard, M. 1988. *The Memory Jogger: A Pocket Guide of Tools for Continuous Improvement.* Second edition. Metheun, MA: Goal/QPC.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen. 1984. *Classification and Regression Trees.* Boca Raton, FL: Chapman and Hall/CRC.

Breyfogle, F. W., III. 1999. *Implementing Six Sigma: Smarter Solutions Using Statistical Methods.* New York: John Wiley & Sons.

Burdick, R. K., and G. A. Larsen. 1997. "Confidence Intervals on Measures of Variability in R&R Studies." *Journal of Quality Technology* 29:261–273.

Cameron, A. C., and P. K. Trivedi. 1998. *Regression Analysis of Count Data.* Cambridge, England: Cambridge University Press.

Canadian Oxford Dictionary: The Foremost Authority on Current Canadian English. 2002. Edited by Katherine Barber. Oxford, England: Oxford University Press.

Cleveland, W. S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74:829–836.

Coleman, D. E., and D C. Montgomery. 1993. "A Systematic Approach to Planning for a Designed Industrial Experiment." *Technometrics* 35:1–26.

Cotter, C. S. 1979. "A Screening Design for Factorial Experiments with Interactions." *Biometrika* 66:317–320.

Davenport, T. H., and L. Prusak. 1997. *Working Knowledge: How Organizations Manage What They Know*. Harvard, MA: Harvard Business School Publishing.

Deming, W. E. 1992. *Out of the Crisis.* Seventh printing. Cambridge, MA: MIT Center for Advanced Engineering Study.

De Mast, J., 2003. "Quality Improvement from the Viewpoint of Statistical Method." *Quality and Reliability Engineering International* 19:255–264.

———. 2004. "A Methodological Comparison of Three Strategies for Quality Improvement." *International Journal of Quality and Reliability Management* 21(2): 198–213.

De Mast, J., K.C.B. Roes, and R.J.M.M. Does. 2001. "The Multi-Vari Chart: A Systematic Approach." *Quality Engineering* 13:437–448.

De Mast, J., W.A.J. Schippers, R.J.M.M. Does, and E. Van den Heuvel. 2000. "Steps and Strategies in Process Improvement." *Quality and Reliability Engineering International* 16:301–311.

Del Castillo, E. 2002. *Statistical Process Adjustment for Quality Control.* New York: John Wiley & Sons.

Duncan, A. J. 1986. *Quality Control and Industrial Statistics.* Fifth edition. Homewood, IL: Richard D. Irwin.

Evans, J. R., and W. M. Lindsay. 1993. *The Management and Control of Quality.* Second edition. St. Paul, MN: West.

Farnum, N. R. 1994. *Modern Statistical Quality Control and Improvement.* Belmont, CA: Duxbury.

Feder, P. I. 1974. "Some Differences Between Fixed, Mixed, and Random Effects Analysis of Variance Models. *Technometrics* 6:98–106.

Goldratt, E. M. 1992. *The Goal.* Second revised edition. Great Barrington, MA: North River.

Grubbs, F. E. 1954. "An Optimum Procedure for Setting Machines or Adjusting Processes." *Industrial Quality Control*, July. Reprinted in the *Journal of Quality Technology* 15(4):186–189.

Hahn, G. J. 1984. "Experimental Design in the Complex World." *Technometrics* 26:19–31.

Hamada, M., and J. A. Nelder. 1997. "Generalized Linear Models for Quality Improvement Experiments." *Journal of Quality Technology* 29:292–304.

Harrington, H. J. 1987. *The Improvement Process: How America's Leading Companies Improve Quality.* New York: McGraw-Hill.

Harry, M. J. 1997. *The Vision of Six Sigma: A Roadmap for Breakthrough.* Fifth edition. Phoenix, AZ: TriStar.

Harry, M., and R. Schroeder. 2000. *Six Sigma: The Breakthrough Strategy Revolutionizing the World's Top Corporations*. New York: Doubleday.

Hastie, J. H., and R. J. Tibshirani. 1990. *Generalized Additive Models.* Boca Raton, FL: Chapman and Hall/CRC.

Hosmer, D. W., and S. Lemeshow. 2000. *Applied Logistic Regression.* Second edition. New York: John Wiley & Sons.

Hoyle, D. 2001. *ISO 9000 Quality Systems Handbook.* Fourth edition. Woburn, MA: Butterworth and Heinemann.

Ingram, D. J. 1993. "A Statistical Approach to Component Swapping." *American Society for Quality Control (ASQC) Quality Congress Transactions*, Milwaukee, WI, 85–91.

Ishikawa, K. 1982. *Guide to Quality Control.* Second revised edition. Tokyo: Asian Productivity Organization.

Jenkins, G. M. 1983. "Feedforward-Feedback Control Schemes." In *Encyclopedia of Statistical Sciences,* edited by S. Kotz and N. L. Johnson. New York: John Wiley & Sons.

Juran, J. M. 1988. *Juran on Planning for Quality.* New York: Free Press.

Juran, J. M., and F. M. Gryna, Jr. 1980. *Quality Planning and Analysis.* Second edition. New York: McGraw-Hill.

Juran, J. M., F. M. Gryna, and R. S. Bingham, eds. 1979. *Quality Control Handbook.* Third edition. New York: McGraw Hill.

Kalbfleisch, J. G. 1985. *Probability and Statistical Inference.* Vol. 2, *Probability.* Second edition. New York: Springer-Verlag.

Kotz, S., and N. Johnson. 2002. "Process Capability Indices: A Review 1992–2000." *Journal of Quality Technology* 34:2–53.

Kume, H. 1985. *Statistical Methods for Quality Improvement.* Tokyo: Association for Overseas Technical Scholarships (AOTS).

Lawless, J. F., R. J. MacKay, and J. A. Robinson. 1999. "Analysis of Variation Transmission in Manufacturing Processes, Part I." *Journal of Quality Technology* 31:131–142.

Ledolter, J., and A. Swersey. 1997a. "Dorian Shainin's Variables Search Procedure: A Critical Assessment." *Journal of Quality Technology* 29:237–247.

Ledolter, J., and A. Swersey. 1997b. "An Evaluation of Pre-control." *Journal of Quality Technology* 29:163–171.

Lewis, J. P. 2002. *Fundamentals of Project Management: Developing Core Competencies to Help Outperform the Competition.* Second edition. Saranac Lake, NY: American Management Association (AMACOM).

Liberatore, R. L. 2001. "Teaching the Role of SPC in Industrial Statistics." *Quality Progress*, July, 89–94.

Logothetis, N. 1990. "A Perspective on Shainin's Approach to Experimental Design for Quality Improvement." *Quality and Reliability Engineering International* 6:195–202.

Mackertich, N. A. 1990. "Pre-control vs. Control Charting: A Critical Comparison." *Quality Engineering* 2:253–260.

McCallagh, P., and J. L. Nelder. 1989. *Generalized Linear Models.* Second edition. Boca Raton, FL: Chapman Hall/CRC.

Mease, D., V. N. Nair, and A. Sudjianto. 2004. "Selective Assembly in Manufacturing: Statistical Issues and Optimal Binning Strategies." *Technometrics* 46:165–175.

Meyer, R. D., D. M. Steinberg, and G.E.P. Box. 1996. "Follow-up Designs to Resolve Confounding in Multifactor Experiments (with discussion)." *Technometrics* 38:303–332.

MINITAB. 2000a. *MINITAB User's Guide 1: Data, Graphics, and Macros.* Release 13 for Windows. State College, PA: MINITAB.

———. 2000b. *MINITAB User's Guide 2: Data Analysis and Quality Tools.* Release 13 for Windows. State College, PA: MINITAB.

Montgomery, D. C. 1996. *Introduction to Statistical Quality Control.* Third edition. New York: John Wiley & Sons.

———. 2001. *Design and Analysis of Experiments.* Fifth edition. New York: John Wiley & Sons.

Montgomery, D. C., E. A. Peck, and G. G. Vining. 2001. *Introduction to Linear Regression Analysis.* Third edition. New York: John Wiley & Sons.

Nair, V. N. 1992. "Taguchi's Parameter Design: A Panel Discussion." *Technometrics* 34:127–162.

Nair, V. N., and D. Pregibon. 1988. "Analyzing Dispersion Effects from Replicated Factorial Experiments." *Technometrics* 30: 247–257.

Nelson, L. S. 1985. "Sample Size Tables for Analysis of Variance." *Journal of Quality Technology* 17:167–169.

Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman W. 1996. *Applied Linear Statistical Models*. Fourth edition. Chicago: Irwin.

Nikolaou, M. 1996. "Computer-aided Process Engineering in the Snack Food Industry." *Proceedings Chemical Process Control Conference* (CPC-V), Tahoe City.

Odeh, R. E., and M. Fox. 1975. *Sample Size Choice: Charts for Experimenters with Linear Models.* Second edition. Homewood, IL: Irwin.

O'Dell, C. S., N. Essaides, and N. Ostro. 1998. *If Only We Knew What We Know: The Transfer of Internal Knowledge and Best Practice*. New York: Free Press.

Oldford, R. W., and R. J. MacKay. 2001. Stat 231 course notes, Fall 2001, University of Waterloo, Ontario.

Parmet, Y., and D. M. Steinberg. 2001. "Quality Improvement from Disassembly-Reassembly Experiments." *Communications in Statistics: Theory and Methods* 30:969–985.

Phadke, M. S. 1989. *Quality Engineering Using Robust Design.* Englewood Cliffs, NJ: Prentice-Hall.

Prett, D. M., and C. E. Gracia. 1988. *Fundamental Process Control.* Boston: Butterworths.

Quinlan, J. 1985. "Product Improvement by Application of Taguchi Methods." *Third Supplier Symposium on Taguchi Methods*, American Supplier Institute, Dearborn, MI, 367–384.

Robinson, G. K. 2000. *Practical Strategies for Experimenting.* New York: John Wiley & Sons.

Ross, P. J. 1988. *Taguchi Techniques for Quality Engineering: Loss Function, Orthogonal Experiments, Parameter and Tolerance Design.* New York: McGraw-Hill.

Ryan, B. F., B. L. Joiner, T. Ryan, Jr. 2000. *Minitab Handbook.* Fourth edition. Belmont, CA: Duxbury.

Ryan, T. P. 1989. *Statistical Methods for Quality Improvement.* New York: John Wiley & Sons.

Satterthwaite, F. E. 1954. "A Simple, Effective Process Control Method." Rath & Strong Inc. Report 54-1, Boston, MA.

Scholtes P. R. 1998. *The Team Handbook: How to Use Teams to Improve Quality.* Madison, WI: Oriel.

Seder, L. A. 1950a. "Diagnosis with Diagrams: Part I." *Industrial Quality Control,* January, 11–19.

———. 1950b. "Diagnosis with Diagrams: Part II." *Industrial Quality Control*, February, 7–11.

———. 1990. "Diagnosis with Diagrams." *Quality Engineering* 2:505–530 (reprinted from original in *Industrial Quality Control,* 1950).

Senge, P. M. 1990. *The Fifth Discipline: The Art and Practice of the Learning Organization.* New York: Currency/Doubleday.

Shainin, D., and P. Shainin. 1988. "Better Than Taguchi Orthogonal Tables." *Quality and Reliability Engineering International* 4:143–149.

———. 1989. "Pre-control Versus $\bar{X}$ & R Charting: Continuous or Immediate Quality Improvement?" *Quality Engineering* 1:419–429.

Shainin, R. D. 1992. "Technical Problem Solving Strategies: A Case Study." 46th *Annual Quality Congress Proceedings*, American Society for Quality Control (ASQC), Milwaukee, WI, 876–882.

———. 1993. "Strategies for Technical Problem Solving." *Quality Engineering* 5(3):433–448.

Shimbun, N. K., ed. 1988. *Poka-yoke: Improving Product Quality by Preventing Defects.* Cambridge, MA: Productivity Press.

Shingo, S. 1986. *Zero Quality Control: Source Inspection and the Poka-yoke System.* Stanford, CT: Productivity Press.

Snee, R. D. 2001. "My Process Is Too Variable: Now What Do I Do?: How to Produce and Use a Successful Multi-vari Study." *Quality Progress*, December, 65–68.

Spiers, B. 1989. "Analysis of Destructive Measuring Systems," *Forty-third Annual Quality Congress Proceedings*, American Society for Quality Control (ASQC), Milwaukee, WI, 22–27.

Steiner, S. H., and R. J. MacKay. 1997–1998. "Strategies for Variability Reduction." *Quality Engineering* 10: 125–136.

Taguchi, G. 1986. *Introduction to Quality Engineering, Designing Quality into Products and Processes.* Tokyo: Asian Productivity Association.

———. 1987. *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs.* White Plains, NY: UNIPUB/Kraus International.

Taguchi, G., and Y. Wu. 1985. *Introduction to Off-line Quality Control.* Nagaya, Japan: Central Japan Quality Control Association.

Taylor, W. A. 1991. *Optimization and Variation Reduction in Quality.* New York: McGraw-Hill.

Tippett, L.H.C. 1934. *Applications of Statistical Methods to the Control of Quality in Industrial Production.* Brighton, UK: Manchester Statistical Society.

Todd, R. H., D. K. Allen, and L. Alting. 1994. *Manufacturing Processes Reference Guide*. New York: Industrial Press.

Traver, R. W. 1985. "Pre-control: A Good Alternative to $\bar{X}$-R Charts." *Quality Progress*, September, 11–14.

Traver, R. W. 1995. *Manufacturing Solutions for Consistent Quality and Reliability.* New York: American Management Association.

Tukey, J. W. 1959., "A Quick, Compact, Two-sample Test to Duckworth's Specifications." *Journal of Quality Technology* 1:31–48.

VanderWiel, S. A., and S. B. Vardeman. 1994. "A Discussion of All-or-None Inspection Policies." *Technometrics* 36: 102–109.

Wheeler, D. J., and R. W. Lyday. 1989. *Evaluating the Measurement Process.* Second edition. Knoxville, TN: SPC.

———. 1990. *Understanding Industrial Experimentation.* Second edition. Knoxville, TN: SPC.

———. 1992. "Problems with Gage R&R Studies." *American Society for Quality Control (ASQC) Quality Congress Transactions*, Milwaukee, WI, 179–185.

Wu, C.F.J., and M. Hamada. 2000. *Experiments: Planning, Analysis, and Parameter Design Optimization*. New York: John Wiley & Sons.

Zaciewski, R. D., and L. Nemeth. 1995. "The Multi-vari Chart: An Underutilized Quality Tool." *Quality Progress* 28:81–83.

Index

Note: All page references following CD refer to text on the compact disk.

D

E

F

G

H

I

J

K

L

M

N

O

S

T

U

V

W

Z

Case Studies

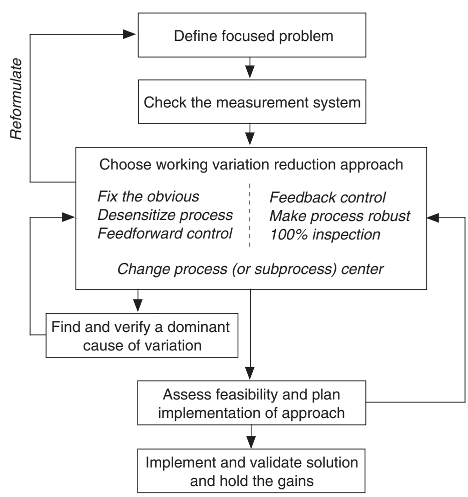
A problem well defined is a problem half solved.

—Charles Kettering, 1876–1958

We provide three case studies to illustrate the use of the Statistical Engineering variation reduction algorithm. We hope that the case studies help you to:

- Better understand the stages of the algorithm and how to move among them.
- Understand the considerations in the choice of a working approach.
- Select appropriate investigations and analysis tools in the search for a dominant cause.
- Assess the feasibility of an approach and implement it.

As W. Edwards Deming somewhat paradoxically said, "You cannot learn by example." We recognize that there are no other processes or problems for which the case studies are perfect models. We recommend, when reading the case studies, that you think about where your own processes and problems are similar to and different from those in the cases.



Statistical Engineering variation reduction algorithm.

Case Study I

Brake Rotor Balance

The value of an idea lies in using it.

—Thomas Edison, 1847–1931

An iron foundry produced veined brake rotors (Figure I.1) that were machined at a separate location. The machining plant 100% inspected the rotors for balance and welded a weight into the veins if the imbalance was too severe. We call a rotor needing added weight a balance reject.

The historic rate of balance rejects was approximately 25%. The foundry initiated the project because the reject rate jumped to 50%. This increase in rework coincided with a move from a four-cavity to a six-cavity (called a four-gang and six-gang) core mold to increase productivity in the foundry. The cores were set in the mold to create the veins when the rotor was cast.

Figure I.1 Brake rotor.

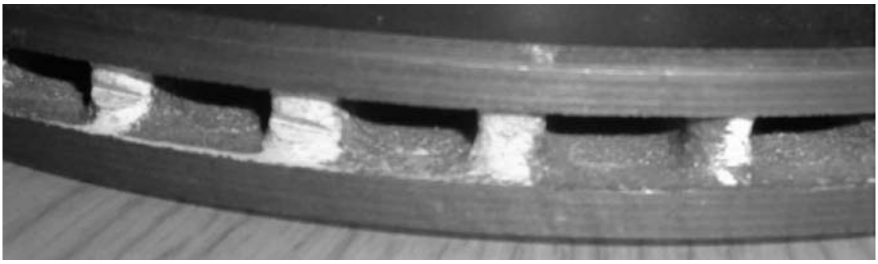


Figure I.2 Side view of a brake rotor showing veins.

The foundry was convinced that the change to the six-cavity mold was not the cause of the increase in balance rejects. Their confidence was based on their previous experience and because a full analysis of the six-cavity mold had shown all dimensions well within specification.

The increased reject rate could not be explained by any other changes made at either the foundry or the machining operation. As it stood, each party blamed the other. To address the increased rework, the machining operation planned to add another rework station. The foundry formed a team with the goal of reducing the reject rate to at most the historical level.

To determine imbalance, the machining plant measured the center of gravity (a distance and direction from the rotor center), which was then translated into a weight (in ounces) and orientation needed to balance the rotor. If needed, the weights were welded to the veins on the inside of the rotor, as seen in Figure I.2. A balance reject was any rotor needing weight greater than 0.5 ounces. To focus the problem, the team selected the balance weight as the output. They knew that if they could reduce the weight, they could eliminate the rework, regardless of the orientation.

To establish the baseline in terms of the balance weight, the team selected 300 machined rotors spread out over the previous week's production. The data are given in *brake rotor balance baseline*. The baseline histogram and run chart are given in Figure I.3. In the baseline, 46% of the rotors required rework. The run chart of the balance weights suggests no obvious pattern over time.

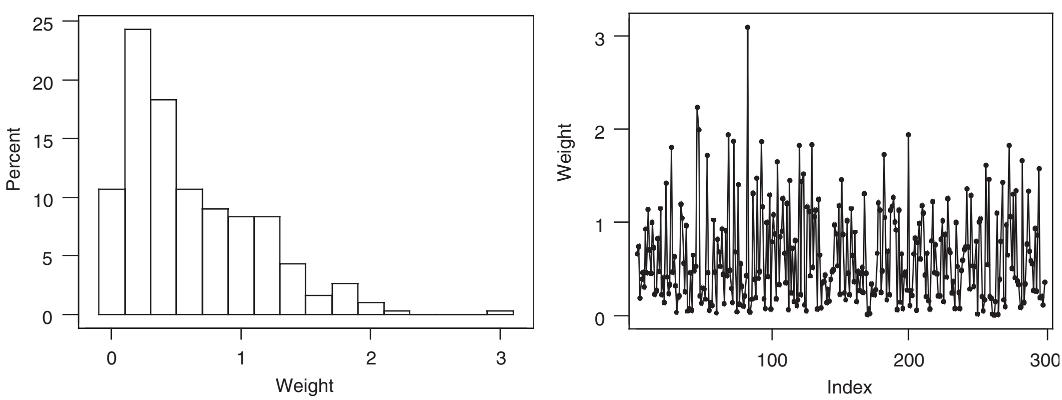


Figure I.3 Histogram and run chart of balance weights in the baseline investigation.

Summarizing the baseline results we have:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
weight	300	0.6215	0.4700	0.5813	0.5079	0.0293

Variable	Minimum	Maximum	Q1	Q3
weight	0.0100	3.0900	0.2100	0.9625

There was one outlier in the baseline sample. If we exclude this rotor, the standard deviation is reduced slightly to 0.49. Ignoring this casting, the full extent of variation of the imbalance weight is about 0 to 2.25. The team set the goal to reduce variation in the balance weight so that at least 75% of the rotors had weight less than 0.5.

The team moved on to the Check the Measurement System stage. They wanted to ensure they had a reliable way to measure balance. There were three gages in parallel used to measure balance weight. See the process map in Figure I.4.

For the investigation, the team selected three rotors with initial measured weights of 0.10, 0.54, and 1.12. They measured the three rotors twice using each of the three gages on three separate days. There was little operator effect since the gages were mostly automated. There were 54 measurements in total. The data are given in the file *brake rotor balance measurement*.

We see in Figure I.5 that the measurement system can easily distinguish the three rotors.

We calculate the summary measures by rotor (part).

Variable	rotor	N	Mean	Median	TrMean	StDev
weight	1	18	0.1083	0.1150	0.1088	0.0685
	2	18	0.5028	0.5050	0.4994	0.0697
	3	18	1.0172	1.0350	1.0175	0.0584

Variable	rotor	SE Mean	Minimum	Maximum	Q1	Q3
weight	1	0.0162	-0.0100	0.2200	0.0500	0.1700
	2	0.0164	0.4000	0.6600	0.4275	0.5600
	3	0.0138	0.9100	1.1200	0.9700	1.0525

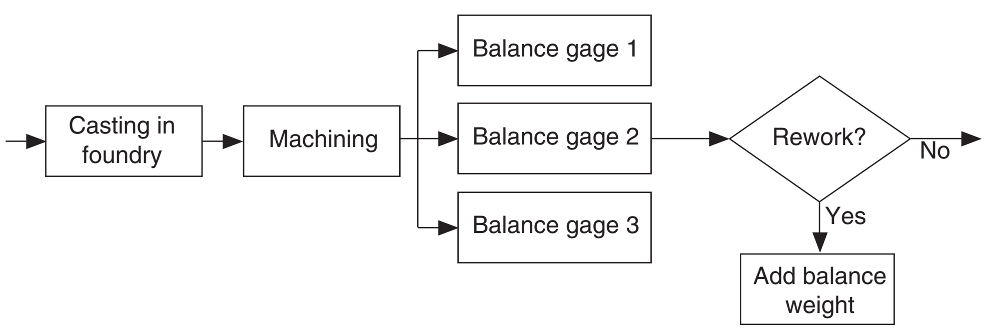


Figure I.4 Brake rotor process map.

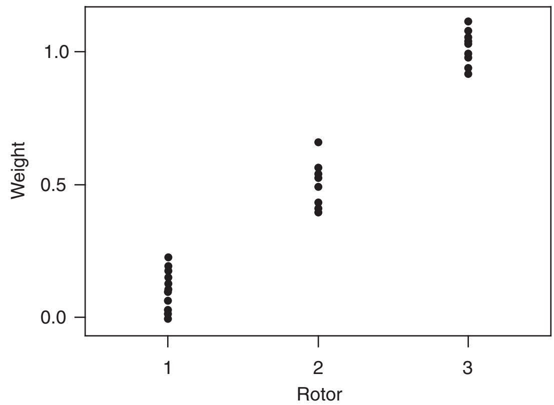


Figure I.5 Brake rotor balance measurement investigation results.

There are no obvious outliers in these data. We estimate the standard deviation of the measurement system as

$$\frac{\sqrt{0.0685^2 + 0.0697^2 + 0.0584^2}}{3} = 0.06$$

The measurement system was judged to be acceptable since this is much less than the overall standard deviation, estimated in the baseline investigation to be 0.51. The discrimination ratio D is

$$\frac{\sqrt{0.51^2 - 0.06^2}}{0.06} = 8.4$$

Next, the team had to select a working approach. They first considered the non-cause-based approaches. They ruled out 100% inspection, since that was the current approach and was too costly. They also eliminated feedback control, since there is no strong pattern in the variation over time in the baseline data and they had no idea of how to adjust the weight. Robustness or Move the Process Center (equivalent approaches in this case) were possibilities but, without more process knowledge, were not likely to succeed. The team decided to search for the dominant cause of variation in the balance weight.

The team first looked at easily available data to see what causes could be eliminated. They recorded, on a defect concentration diagram, the location (in increments of 30°) of the welded rework weight for the 140 balance rejects from the baseline investigation. The dots on the part schematic in Figure I.6 show the nonsymmetrical pattern of balance weights observed. Since the machining process is rotationally symmetric and the casting process is oriented, the team eliminated all causes in the machining operation. With this simple investigation the team made tremendous progress with little cost and in a short time.

The team next planned a group comparison. They selected 30 balance rejects (average weight 1.02) and 30 balanced brake rotors (average weight 0.15) for the comparison. They then measured 26 foundry-determined characteristics on each machined rotor. Note that

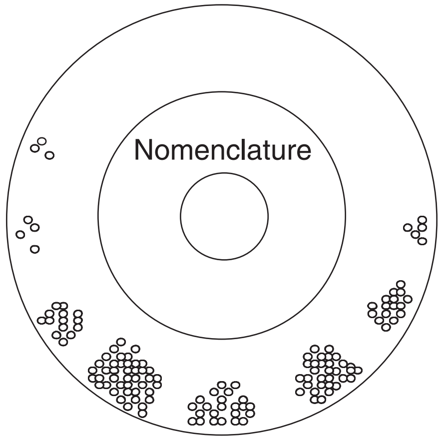


Figure I.6 Concentration diagram of weight locations.

machining destroyed many foundry characteristics. The input characteristics included vein thickness at eight different locations (four orientations at inner and outer edges of the rotor), three derived thickness variation measures (variation at inner and outer locations, and the overall variation), core offset at two orientations, four diameters related to the core size, and five hole diameters (on the inner part of the rotor, see Figure I.1). The data are given in the file *brake rotor balance comparison*.

The team identified two input characteristics, thickness variation and core position (offset), that were substantially different for balanced and unbalanced rotors. See Figure I.7. The results for thickness variation were more compelling than for core position. However, based on engineering knowledge, both inputs were plausible dominant causes of imbalance variation.

The team decided to verify these suspects hoping that they could then reformulate the problem, since the suspects could be measured in the foundry. This would save time and effort in future investigations since they would no longer need to trace rotors between the

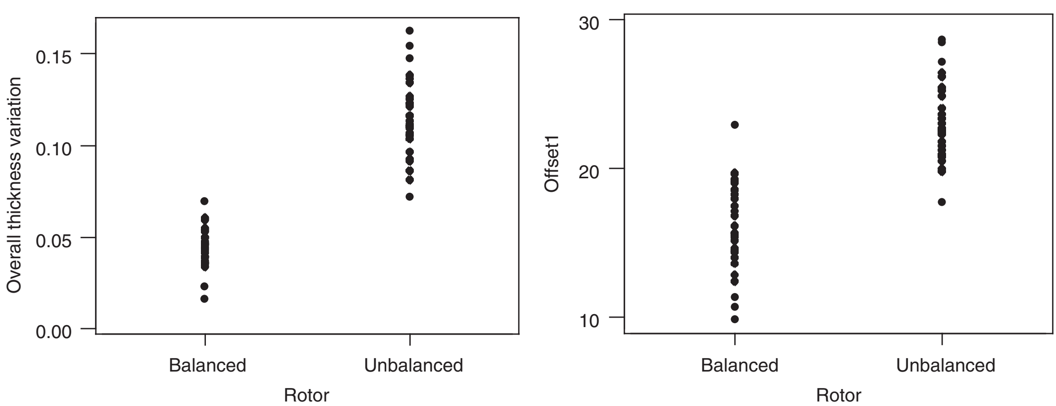


Figure I.7 Plot of thickness variation and offset 1 in balanced and unbalanced rotors.

Table I.1 Brake rotor verification experiment plan and results.

Treatment	Tooling	Core position	Thickness variation	Run order	Average balance weight
1	4-gang	Offset	30-thousandths	8	0.56
2	4-gang	Offset	Nominal	1	0.17
3	4-gang	Nominal	30-thousandths	3	0.44
4	4-gang	Nominal	Nominal	7	0.08
5	6-gang	Offset	30-thousandths	2	1.52
6	6-gang	Offset	Nominal	5	0.37
7	6-gang	Nominal	30-thousandths	4	1.34
8	6-gang	Nominal	Nominal	6	0.03

foundry and the machining operation. They planned and conducted a verification experiment to confirm that core thickness variation and core position were substantive causes of the balance weight variation and that the six-cavity mold was not. They used two levels for each input and a full factorial design. The team selected the nonnominal levels for core position and thickness variation at the high end of their normal range of variation. We give the eight treatments and input levels in Table I.1. For each run, they planned to produce eight castings. The order of the treatments was randomized.

The 64 experimental castings were tagged and shipped to the machining plant to be processed and measured. We give the data in the file *brake rotor balance verification* and the average weight over the eight castings for each run in Table I.1.

We plot the weights by treatment in Figure I.8. We see that some treatments have very little variation and result in a low average weight. We also see roughly the full extent of the variation in weight given by the dashed line on Figure I.8.

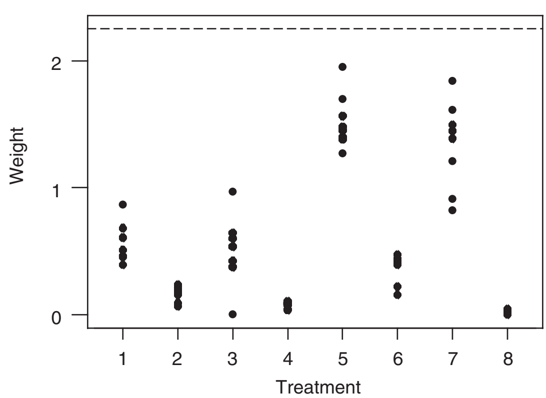


Figure I.8 Plot of weight by treatment.

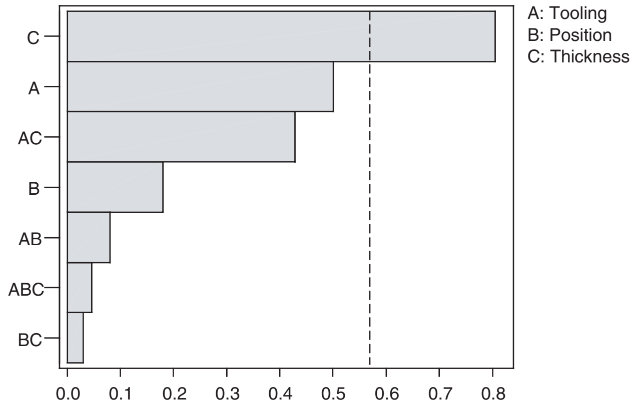


Figure I.9 Pareto chart of the effects for brake rotor verification experiment.

From the Pareto plot of the effects, given in Figure I.9, we see the effect of core position is small, so we eliminate it as a suspect. Since the tooling and thickness variation have a relatively large interaction, we look at the effect of these two suspects simultaneously.

We give the main and interaction effects plots in figures I.10 and I.11, respectively.

From the interaction plot, the team concluded that low thickness variation using the four-cavity mold produced the optimal results (the weights required were so small that the balance specification was met without rework). Thus, the dominant cause of the imbalance problem was in the core molding process.

This search for the dominant cause is summarized in the diagnostic tree shown in Figure I.12.

The team made the obvious fix and recommended that the foundry go back to the original four-cavity core mold. When this change was implemented, the rate of balance rejects immediately dropped to its historical levels. The team had met the original project goal.

The major lesson learned in the project was the effect of the thickness variation on the balance weight. The verification experiment showed that thickness variation in the cores was a dominant cause of balance weight variation in the *original process* that used the four-cavity

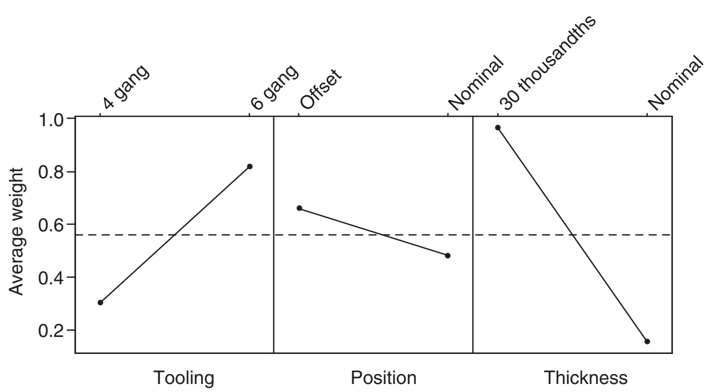


Figure I.10 Main effects plot for brake rotor verification experiment.

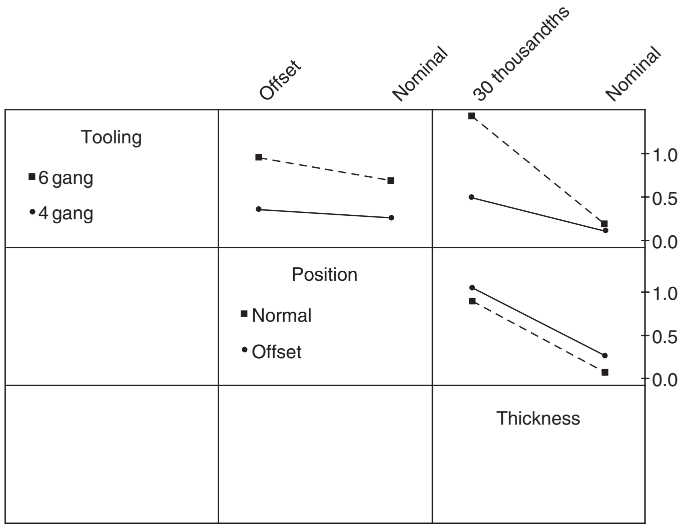


Figure I.11 Interaction plot for brake rotor verification experiment.

mold. The team was puzzled about the interaction between the thickness variation and the number of cavities in the mold. Knowledge of a dominant cause provided the opportunity to improve the process further. There was no immediately known method for adjusting the core-making process to reduce thickness variation. The team considered looking for an adjuster in the core-making process but rejected this approach since they did not expect to be successful.

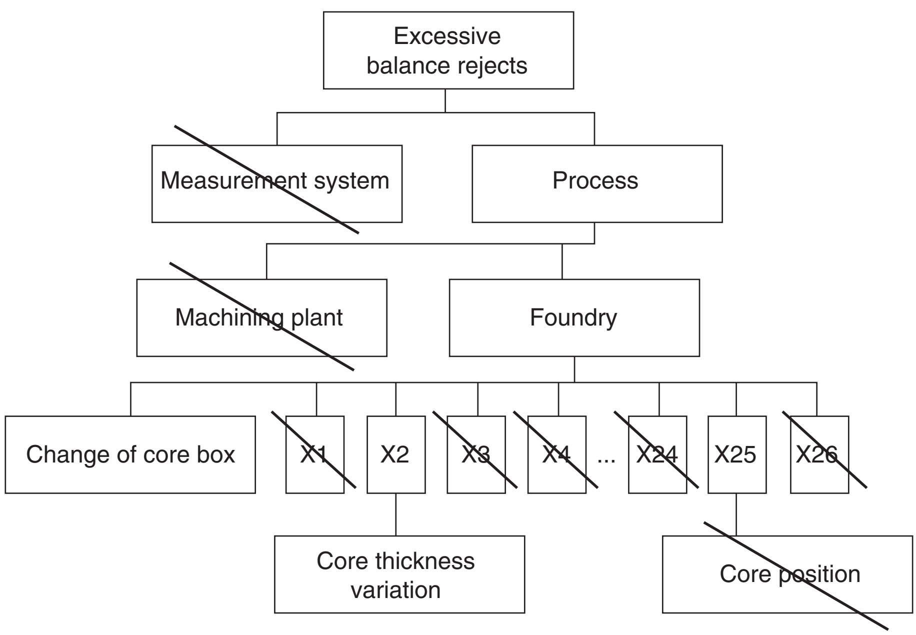


Figure I.12 Summary of the method of elimination for brake rotor example.

The team had the opportunity to implement a new core-making process for the veined rotor. The equipment was already available in the plant but not in use. The team knew that the cold box process was dimensionally stable, and they expected much less thickness variation with this process. This is the Fix the Obvious approach. With the implementation of the cold box method, the process was greatly improved. Over the next four months the rate of balance rejects dropped to 0.2%, a large reduction from the 50% at the start of the project. The machining plant eliminated the expensive rework stations and scrapped the few balance rejects in the new process.

Highlights

Some strengths and weaknesses of this case are:

- The use of the available data together with the knowledge of the symmetry of the machining operation to eliminate all causes in the machining plant.
- In the group comparison, the team could have selected two smaller sets of rotors that were more extreme with respect to the balance weights.
- The carefully planned and conducted verification experiment.
- The application of the knowledge gained about thickness variation (the dominant cause) to select the cold box process.

Case Study II

Rod Thickness

The ancestor of every action is a thought.

—Ralph Waldo Emerson, 1803–1882

A plant manufactured approximately 12,000 connecting rods per day for use in an engine assembled in the plant. The rod, illustrated in Figure II.1, connects the piston (at the small or pin end of the rod) to the crankshaft (at the large or crank end of the rod). The plant received forged blanks and machined the rods in a large number of process steps.

Management identified the rod line for a variation reduction project because the overall scrap cost was greater than budget. The yearly scrap cost was in excess of $400,000, and the scrap rate was 3.2% over the previous four months. Management set a goal to reduce the scrap rate to 1.6% in its annual business plan. The rod line production manager assigned a team to the project.

Looking at scrap records, the team found that scrap occurred at several stages in the process and for several reasons. To focus the problem, they used Pareto analysis on the records for one month. The results, in Figure II.2, showed that 65% of the scrap occurred at a grinding operation. At this operation, the team discovered that about 90% of the scrap was due to rods with their crank end thickness less than specification. The team focused their attention on reducing variation in rod thickness.

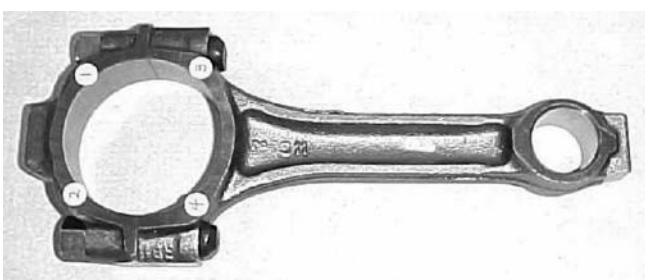


Figure II.1 A connecting rod.

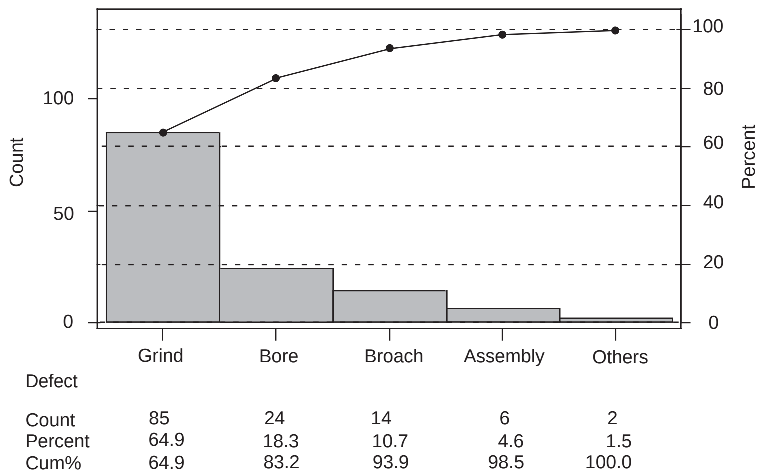


Figure II.2 Pareto chart for scrap by operation for the rod line.

At the grinder, the final thickness of the rod was set in a two-pass operation. An operator loaded the parts into the machine. In the first pass, one side of the rod was ground in three steps. The rod was then turned over by the machine operator, and the second side was ground. The grinder had a rotary table with 20 pallets that passed under four different grinding wheels. An internal control system automatically adjusted the grinding wheels based on thickness measurements taken by a series of gages internal to the grinder.

After grinding, an in-line gage measured the thickness of every rod at four positions (given by the white circles on the crank end of the rod, also faintly numbered 1 through 4, as shown on Figure II.1). The specifications were 0.91 to 0.96 inches at each location. The gage automatically rejected to a rework station a rod with any of the four thickness measurements not meeting the specification. At the rework station, an operator remeasured the rejects using a different gage and scrapped undersized rods.

The rod line ran on three shifts, five days per week. To establish the baseline, the team sampled 200 rods chosen in two batches of 20 rods for each of five days on the day shift only. They recorded the thickness measurements for the four positions using the in-line gage that did not normally store the data. We give the data in the file *rod thickness baseline*. Thickness is given as a deviation from 0.9 in thousandths of an inch.

Figure II.3 gives a histogram of thickness, where the dashed vertical lines are the specification limits. Of the 200 rods, 10 rods had thickness values less than the lower specification limit for at least one of the positions. This was somewhat higher than expected, given the historical scrap rate. There are no obvious outliers. A numerical summary of the baseline data (across all positions) is:

```
Variable         N     Mean    Median    TrMean     StDev   SE Mean
thickness      800   34.575    36.000    34.840    11.023     0.390

Variable   Minimum  Maximum        Q1        Q3
thickness    2.000   59.000    28.000    43.000
```

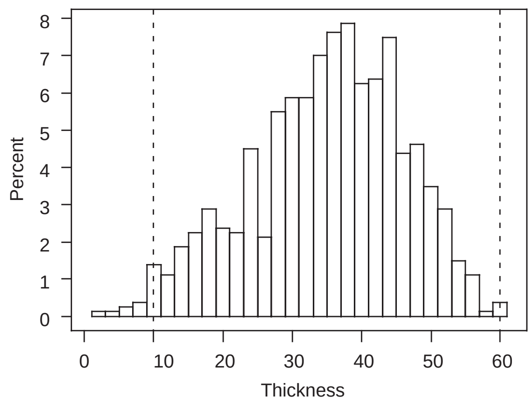


Figure II.3 Baseline histogram of rod thickness (dashed vertical lines give the specification limits).

The team set the objective to produce all rods within the thickness specification. If this ambitious goal could be achieved, they would eliminate 90% of the scrap at the grinder or 58% of the total rod line scrap, hence meeting the project goal.

The process was well centered, so to meet the goal the team needed to reduce the standard deviation to 8.5 from around 11 thousandths of an inch, while keeping the process centered on target. The full extent of the thickness variation was 2 to 59 thousandths. In Figure II.4 we show the baseline data over time (batch), which suggests there was no systematic drift in the process.

The next step was to assess the measurement system. The in-line gage used four sets of transducers to measure rod thickness at the four positions. In effect there were four gages. The team selected three rods to roughly cover the full extent of the thickness variation.

The team was concerned about a relative bias among the four gages. Since it was impossible to measure the same position on the different gages, they sent the three rods to

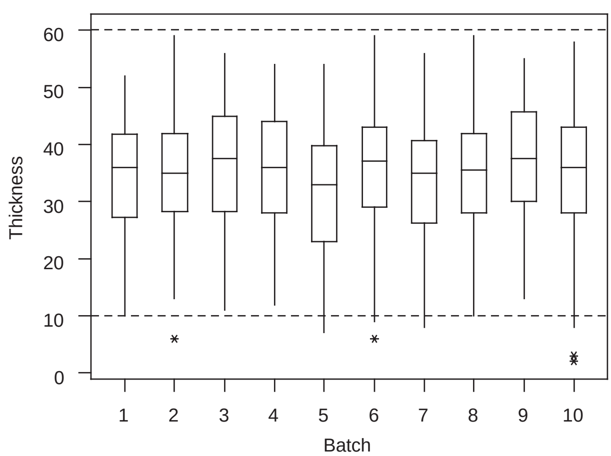


Figure II.4 Box plot of rod thickness by batch for baseline investigation (dashed horizontal lines give the specification limits).

Table II.1 True values of rod thickness by position.

Rod	Position 1	Position 2	Position 3	Position 4
1	54.6	44.7	32.4	59.0
2	22.6	33.8	11.7	22.9
3	38.3	40.6	21.4	41.8

a precision laboratory to have the thicknesses determined with little measurement bias and variation. The precision lab values are given in Table II.1.

The team chose two operators, one from the day shift and one on the afternoon shift. For three days, each operator measured the three rods three times at each of the four positions. In total, there were 216 measurements, 56 per gage. Since we know the true values, we give the measurement errors in the file *rod thickness measurement*.

A summary of the measurement errors for each position (gage) are:

```
Variable   position     N       Mean     Median     TrMean      StDev
error      1           54      0.037      0.400      0.021      1.564
           2           54     -0.033      0.250     -0.013      1.384
           3           54     -0.019     -0.400     -0.013      1.326
           4           54      0.081      0.100      0.075      1.775

Variable   position  SE Mean    Minimum    Maximum         Q1         Q3
error      1           0.213     -3.300      5.400     -0.775      1.400
           2           0.188     -3.700      3.300     -0.800      0.400
           3           0.180     -3.400      2.600     -0.700      0.775
           4           0.242     -3.800      4.000     -1.000      1.125
```

There is no evidence of substantial bias in any of the four gages, since the average errors are close to zero.

Because there is no relative bias among the gages and we have the measurement errors, we estimate the measurement system variation by the standard deviation of the 216 measurement errors. As given, we obtain the estimate 1.512.

```
Variable      N       Mean     Median     TrMean      StDev    SE Mean
error       216      0.017      0.200      0.012      1.512      0.103

Variable     Minimum    Maximum         Q1         Q3
error         -3.800      5.400     -0.800      1.100
```

The baseline variation was 11.023. We can estimate the process variation, using Equation (7.2), by $\sqrt{11.023^2 - 1.512^2} = 10.92$ and the discrimination ratio as $D = 10.92 \div 1.512 = 7.2$. The team decided that the in-line gage was not a dominant cause of the variation and was adequate to move to the next stage of the Statistical Engineering algorithm.

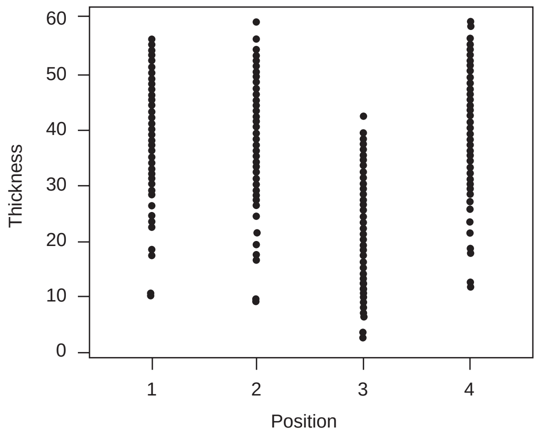


Figure II.5 Box plot of thickness by position from the baseline data.

When considering the choice of a working variation reduction approach, the team quickly rejected the approaches that did not require knowledge of a dominant cause. The process was already well targeted, there was no evidence of an exploitable time pattern in thickness, 100% inspection was already present, and making the process robust seemed difficult. The team decided to search for a dominant cause.

The team first looked at the available data from the baseline investigation to see what family of causes could be eliminated. We show a box plot of thickness by position in Figure II.5. There is a difference between position 3 and the other positions. However, the variation within positions was close to the full extent of variation, so position is not the dominant cause. All but one of the scrapped rods in the baseline sample was undersized at position 3. The team did not know of an obvious fix to move the thickness center for position 3.

The team decided to carry out a multivari investigation to compare pallet-to-pallet, position-to-position, day-to-day, and part-to-part families. They planned to select three consecutive rods from six different pallets (pallets occur in pairs, so three pairs of pallets were chosen) on four different days. They measured thickness at all four positions to confirm the findings from the baseline investigation. In total, they sampled 72 rods and made 288 measurements. The data are given in the file *rod thickness multivari*.

We show the results of the multivari analysis for the pallet-to-pallet, position-to-position, and day-to-day families in Figure II.6. We see that none of these families is home to a dominant cause, although again the thickness was lowest at position 3. To examine the part-to-part family, we form a new input group corresponding to all 96 combinations of day, position, and pallet. We use ANOVA to isolate the variation within the groups. The results are:

```
Analysis of Variance for thickness

Source      DF        SS        MS       F      P
group       95   14444.32    152.05    3.93  0.000
Error      192    7433.33     38.72
Total      287   21877.65
```

We estimate the within group variation due to causes in the part-to-part family as $\sqrt{38.72} = 6.22$, a substantial component of the baseline variation 11.023. The team concluded:

- There was substantial part-to-part variation within the same pallet, position, and day.
- Position 3 was systematically thinner than the other positions.

However, later, when the team examined the histogram of the multivari data in Figure II.6, they noticed that the range of variation in thickness from the multivari did not cover the full extent of variation. There were no undersized rods with thickness values below the lower specification limit of 10. As a result, they decided that their conclusions from the multivari were premature.

This was an unexpected result. The plan for the multivari investigation matched the baseline investigation quite closely. By sampling rods over a number of days, the team had hoped that they would observe undersized rods. The lack of undersized rods suggested that either the dominant cause acts in pallets not chosen for the multivari or that it acts only occasionally in time. Based on this thinking, the team decided to conduct another multivari investigation.

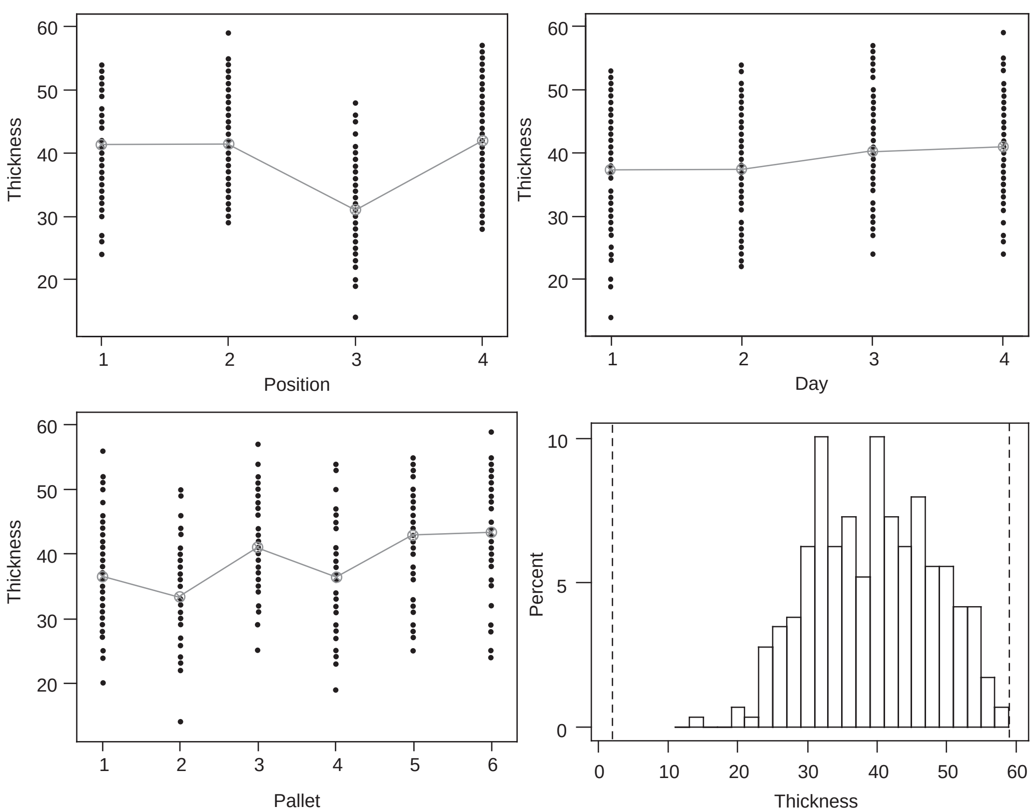


Figure II.6 Summary of the data from the first multivari investigation (vertical dashed lines give the full extent of variation).

In the second multivari, they planned to sample three consecutive rods from a different set of six pallets than those used in the first multivari investigation. This time they repeated the sampling scheme five times within a single shift. This shorter time frame seemed sufficient from the baseline results, as shown in Figure II.4. In total, the team sampled 90 rods and made 360 measurements. The data are given in the file *rod thickness multivari2*.

The team first checked that their new sampling plan generated the full extent of thickness variation. A histogram of the new multivari data, shown in Figure II.7, shows that the full extent of variation was captured. The dominant cause of thickness variation must have acted during the course of the investigation.

We can see the effects of the families of variation for time, pallet, and position in the one-input-at-a-time multivari charts in Figure II.8. We see that a dominant cause of thickness variation is acting time to time. Within each time period, the variation is substantially less than the full extent of variation. All of the undersized rods occur at time 3. The differences among the positions is smaller than previously.

We expect causes in the rod-to-rod (that is part-to-part) family to act haphazardly. To examine the rod-to-rod family, we define a new input group that uniquely numbers the 120 combinations of position, time, and pallet. The within-group standard deviation (Pooled StDev in the MINITAB one-way ANOVA) is 4.29. The rod-to-rod family is not dominant.

During the multivari investigation, 40 undersized rods were produced in a matter of two minutes at time 3. The tool setter suspected that there was a problem with the feedback control scheme that automatically adjusted the grinding wheels. For the finishing wheel, the controller used the measurements from an internal gage and the following rules:

- Measure every part at all four positions.
- If the thickness at any position for two consecutive parts exceeds 50-thousandths of an inch, lower the finishing grinding wheel for three seconds at a fixed rate.
- Ignore the measurements from the next three parts until the adjustment has taken effect.

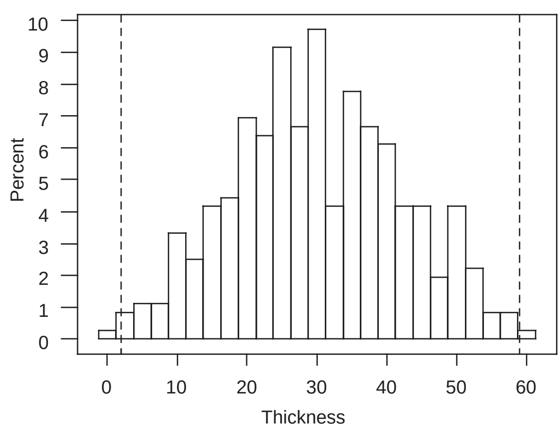


Figure II.7 Histogram of data from the second multivari investigation.

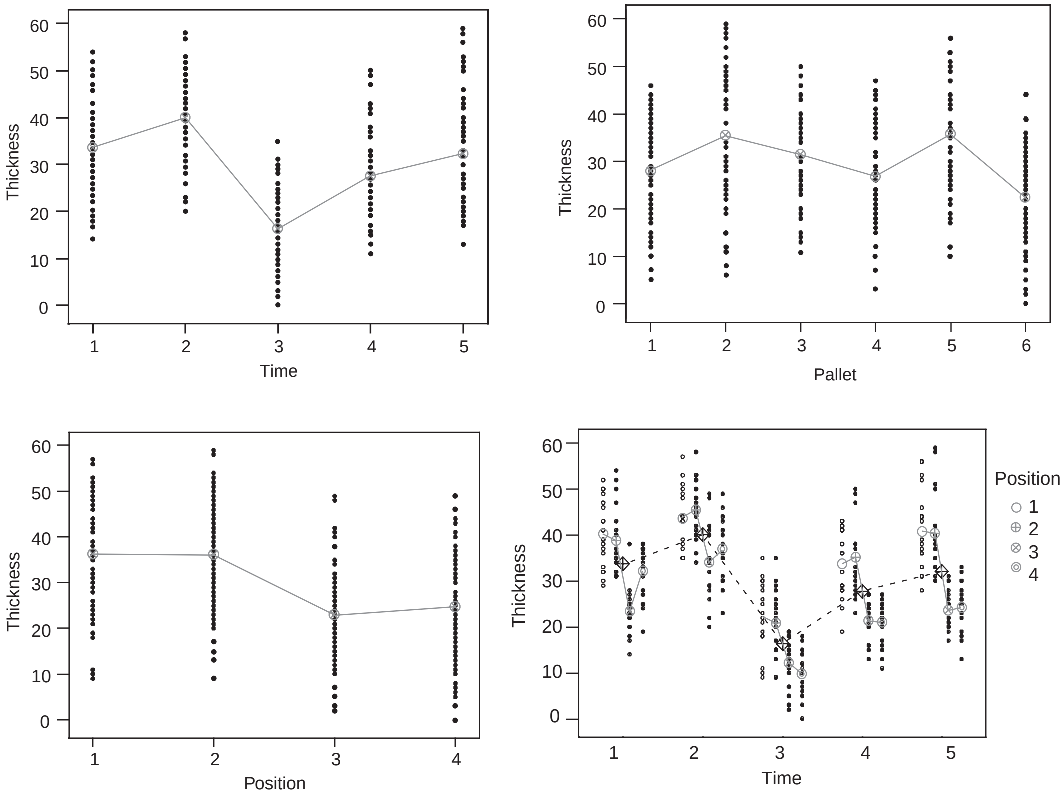


Figure II.8 Multivari charts using data from the second multivari investigation.

The last rule was necessary because of the time delay between the grinding by the finishing wheel and the measurement of thickness by the internal gage. The rule was based on the feedback timer and did not count parts. Upon closer inspection, the team discovered that, occasionally, the last rod ground before compensation was treated as the first rod after the wheel was lowered, thereby fooling the equipment into thinking a second compensation was needed. Combined with the systematic difference among the positions, the double compensation was a dominant cause of undersized rods.

The team adopted the Fix the Obvious approach. They adjusted the logic of the controller to count the parts processed after compensation instead of using the timer. This change prevented the double compensation. The team also looked for ways to adjust the process to better center the four positions. They initiated maintenance on the alignment of the grinding wheels to reduce the position-to-position differences. Since they expected the equipment to deteriorate over time, they set up a monitoring procedure to detect when the position-to-position differences became large. Each day, the operator recorded the thicknesses for five consecutive parts and plotted the position averages on a run chart. The process engineer checked the chart on a regular basis to see if there were systematic differences among the four positions.

In the validation stage, the team assessed the effect of the process changes by comparing performance before and after the changes. We give a histogram of thickness values observed when the plan for the original baseline investigation was repeated in Figure II.9.

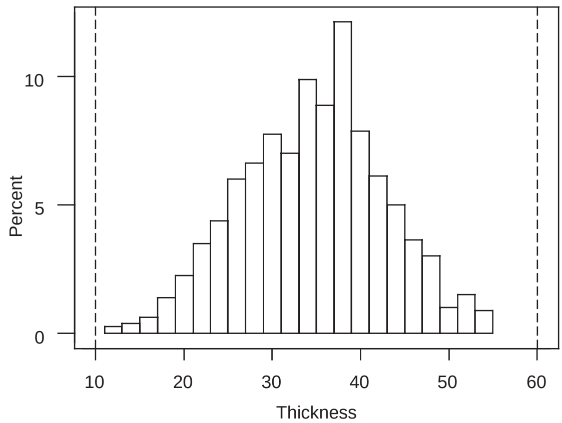


Figure II.9 Histogram of rod thickness from validation investigation.

The data are given in the file *rod thickness validation*. No undersized rods were observed in the validation investigation.

The scrap rate substantially decreased. In the first month after eliminating the double compensation, the scrap rate was 1.7%. The project team fell short of the initial goal, but the project was judged to be successful by the rod line management.

Highlights

Some strengths and weaknesses of this case are:

- The team did a good job of focusing the problem, linking the problem goal "eliminate undersized rods" to the project goal "reduce rod line scrap by at least half."
- The team adopted a good solution to the assessment of the four gages in the measurement system investigation. In this case, they could not measure the same part/position on each gage, so there was no way to tell if there were relative biases among the gages without measuring rods with known thicknesses. Note that a relative bias could have explained why the average thickness for position 3 was smaller than the other positions.
- The team did not make the best use of the information from the baseline investigation. Given that the full extent of thickness variation occurred within a single day (actually a shift), they should have planned the first multivari within a single shift. The time-to-time variation is captured by repeatedly sampling within the shift, as in the second multivari investigation.
- The team made the mistake of not initially checking that the first multivari investigation captured the full extent of thickness variation.
- The team was wise to change the control plan to monitor the process for position-to-position differences, since these were expected to recur, given the nature of the grinding process.

Case Study III

Crankshaft Main Diameter

The true creator is necessity, which is the mother of our invention.

—Plato, 427–347 B.C.

An engine plant machined approximately 1500 V8 crankshafts per day on three shifts. There was 100% inspection at an automated final gage that measured more than 60 characteristics to ensure that the customer, the engine assembly operation, received a high-quality product. We show a schematic of the crankshaft in Figure III.1.

At the beginning of this project, the monthly scrap rate ranged between 6 and 8%, averaging 7.2% over the previous four months. The first-time-through rate, the ratio of the number of parts that were accepted by the final gage to the total number of parts processed, was highly variable, and sometimes as low as 20%. Parts rejected at the final gage were scrapped or reworked and remeasured using an off-line gage.

The goals of this project were to reduce the overall scrap rate to 4.5% or less and increase the first-time-through rate at the final gage to at least 75%. Management hoped to achieve these goals without any substantial capital expenditure.

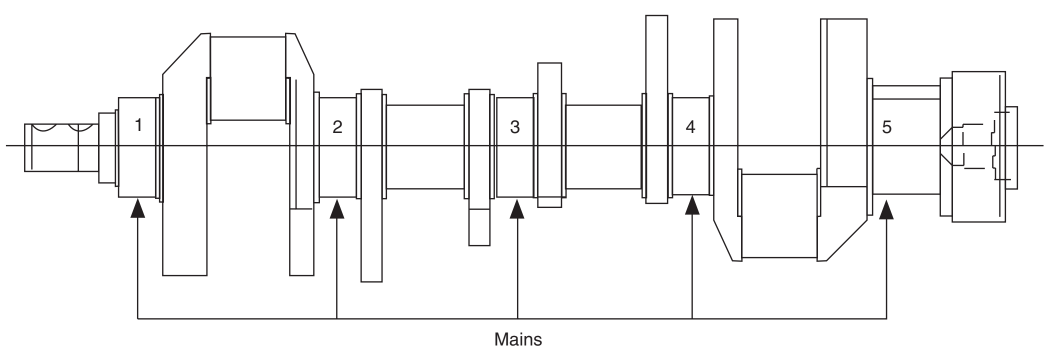


Figure III.1 V8 crankshaft showing five mains.

The team began by determining a more specific goal that could support the overall project goal. The reason was recorded for every crankshaft rejected by the final gage. The team used Pareto analysis to determine that about 85% of rejects and 73% of the scrap were related to the main diameters. The five main diameters are numbered in Figure III.1.

The final gage measured the diameters at three positions—front, center, and rear—on each of the five mains. The specification limits for diameter were ±4.0 thousandths of an inch measured from nominal. The rejects related to main diameter were due to undersized, oversized, and excess taper from the front to the rear of the main. A crankshaft with any main diameter less than –4.0 was scrapped. Taper was the difference between front and rear diameters on each main. The taper specifications were ±2.0 thousandths of an inch. Most parts rejected for taper could be reworked.

To achieve the project goal, the team decided to concentrate efforts on reducing variation in main diameter to eliminate scrap due to undersized diameters and to reduce rework due to taper and oversized diameters.

The team extracted baseline data on main journal diameters for 2000 crankshafts over four days (500 per day selected haphazardly throughout the day) from the final gage. The data are given in the file *crankshaft main diameter baseline*. Because of the large number of measurements, the data are stored in three different formats on the same worksheet:

- In the first 21 columns by crankshaft
- In the next three columns by individual measurements for diameter
- In the last two columns by individual measurements for taper

In the baseline investigation, the scrap rate was 6.9%, with 53% rework, of which 40% was due to excess taper. We show the histograms of all measured diameters and calculated taper values for the 2000 crankshafts in Figure III.2.

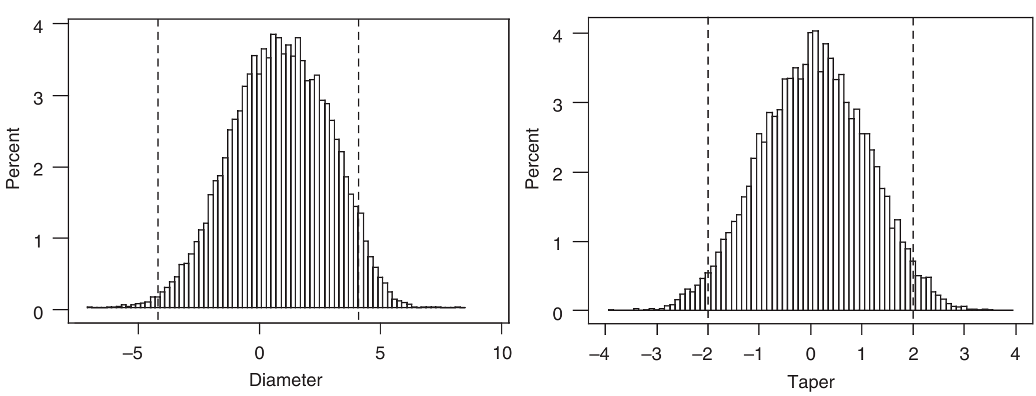


Figure III.2 Baseline histograms of main diameters and taper over all mains and positions (dashed vertical lines show the specification limits).

There are many diameters and taper values outside the specification limits. Numerical summaries of the baseline data are:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
diameter	30000	0.8810	0.9037	0.9005	1.9933	0.0115
taper	10000	0.0619	0.0686	0.0652	1.0367	0.0104

Variable	Minimum	Maximum	Q1	Q3
diameter	-6.9285	8.1401	-0.5168	2.3502
taper	-3.9438	3.5072	-0.6595	0.7870

The team saw that if the process was centered on target, they needed to reduce the standard deviation of diameter to less than 1.30 to meet their goal of producing no diameters out of specification. The full extent of diameter variation was about –5.1 to 6.9 (0.88 ±3*1.99) thousandths of an inch.

There are so many diameter observations that it is difficult to look for patterns over time. We give the box plot of diameter by day in Figure III.3. The team also looked at plots of diameter over time for each of the individual positions. There are no day-to-day patterns.

The next step was to assess the final measurement system. There were 15 different gages and the team could not measure the same part on each of them; for example, they could not measure the number 1 main front diameter on the number 3 rear gage. These no-contact gages used a common air pressure system. The team selected six different crankshafts to be measured three times each on three days. The team chose six crankshafts to roughly cover the full extent of variation because here they were simultaneously investigating 15 different measurement systems. It was not worth the effort to find three crankshafts that gave the full extent of diameter variation for all 15 positions simultaneously. There was no operator effect because the measurement process was automated. There were a total of 54 measurements on the front, center, and rear positions for each of the five mains. The data are given in the file *crankshaft main diameter measurement*.

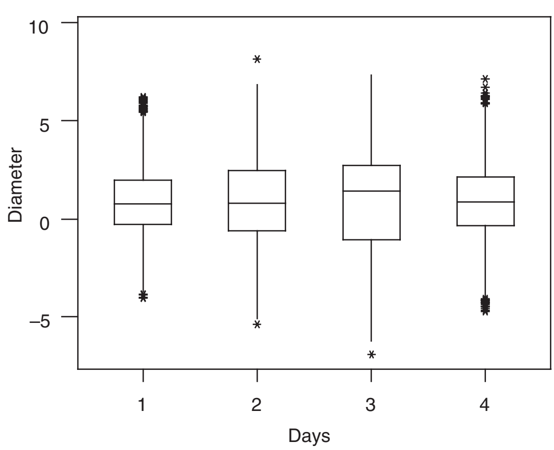


Figure III.3 Box plots of final diameter by day.

The estimates of the measurement variation (standard deviation) of the 15 gages ranged from 0.28 to 1.06. The MINITAB ANOVA results for the two extreme cases are:

```
Analysis of Variance for 1rear
Source      DF        SS        MS        F        P
crankshaft   5    125.43     25.09    22.37    0.000
Error       48     53.82      1.12
Total       53    179.25
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N      Mean     StDev ---+---------+---------+---------+---
1            9     4.450     0.990                           (----*---)
2            9     1.697     1.148          (----*---)
3            9     0.265     1.035 (----*---)
4            9     4.326     1.055                           (---*---)
5            9     2.237     1.050              (---*---)
6            9     3.763     1.070                       (----*---)
                                   ---+---------+---------+---------+---
Pooled StDev =     1.059              0.0       1.6       3.2       4.8

Analysis of Variance for 4front
Source      DF        SS        MS        F        P
crankshaft   5   59.9397   11.9879   153.84    0.000
Error       48    3.7403    0.0779
Total       53   63.6800
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N      Mean     StDev -+---------+---------+---------+-----
1            9   -0.1608    0.2674                             (-*)
2            9   -2.1253    0.1886             (*-)
3            9   -0.8184    0.2653                        (*-)
4            9   -1.4794    0.3258                  (-*)
5            9   -1.9434    0.3204              (-*)
6            9   -3.4972    0.2850 (-*)
                                   -+---------+---------+---------+-----
Pooled StDev =    0.2791          -3.6      -2.4      -1.2       0.0
```

Because there were 15 gages, there may have been relative biases among the gages. These biases would contribute to the variation in diameter across all mains as determined in the baseline. We cannot assess these biases with the data from the measurement system investigation. The team was concerned about how to determine if the measurement system was a dominant cause of the overall variation. They looked at the box plot of diameter by position from the baseline investigation, given in Figure III.4, and noted that there were systematic position-to-position differences.

These differences could be due to the process or the gages. The team decided to assess each gage against the baseline variation within its own position. We give the results in Table III.1.

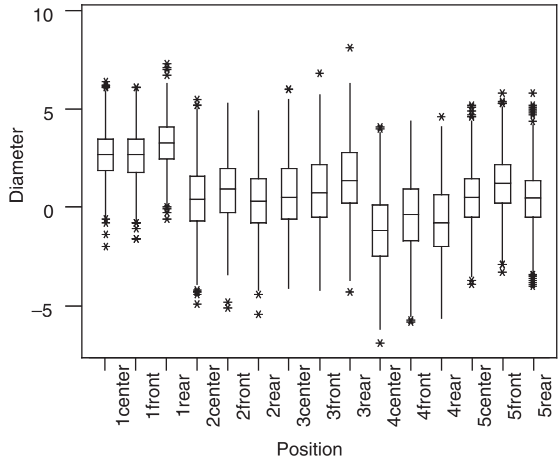


Figure III.4 Diameter by position from baseline data.

Table III.1 Measurement discrimination ratios by position.

Position	Baseline standard deviation	Measurement system variation	Discrimination ratio
1 center	1.17	0.96	0.71
1 front	1.20	0.97	0.72
1 rear	1.18	1.06	0.51
2 center	1.65	0.86	1.65
2 front	1.64	0.76	1.89
2 rear	1.65	0.83	1.70
3 center	1.80	0.86	1.84
3 front	1.78	0.92	1.67
3 rear	1.78	0.85	1.84
4 center	1.78	0.28	6.18
4 front	1.77	0.33	5.28
4 rear	1.78	0.29	5.95
5 center	1.46	0.69	1.88
5 front	1.45	0.56	2.37
5 rear	1.45	0.74	1.69

For some positions, the measurement system was a dominant cause of the variation. This was a shocking discovery, since the gage was thought to be reliable based on the required gage R&R investigations. The team needed to improve the measurement system immediately. The small discrimination ratios helped to explain why some operators remeasured scrapped crankshafts a second time. Given the observed measurement variation, it was not surprising that these crankshafts were sometimes found to be acceptable. Surprisingly, no one found the need to remeasure crankshafts that were first time passes through the measurement system.

To improve the final gage, the team decided to look for a dominant cause of measurement variation. They started with the existing data from the measurement investigation. They examined the data from positions 4 front (one of the best gages) and 1 rear (one of the worst) in more detail. In Figure III.5, we give plots of the estimated measurement errors (measurements minus the position average) by day. We see for 1 rear that the measurement error is substantially different from one day to the next. The dominant cause of the measurement variation for position 1 rear acted in the day-to-day family.

To explore this behavior further, the team measured the same crankshaft once per day for 19 days. The data are given in the file *crankshaft main diameter measurement stability*. The results were striking and surprising. We present results only for the front position for all five mains here. The results were similar for the other positions. In Figure III.6, and in the subsequent numerical results, we see that the final gage was unstable. The day-to-day variation in the gage was much greater than expected based on the existing short-term R&R results.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
1front	19	3.380	3.381	3.351	0.943	0.216
2front	19	1.599	1.309	1.583	1.255	0.288
3front	19	-1.319	-1.205	-1.305	0.778	0.178
4front	19	-1.553	-1.814	-1.573	0.712	0.163
5front	19	2.790	2.721	2.794	0.794	0.182

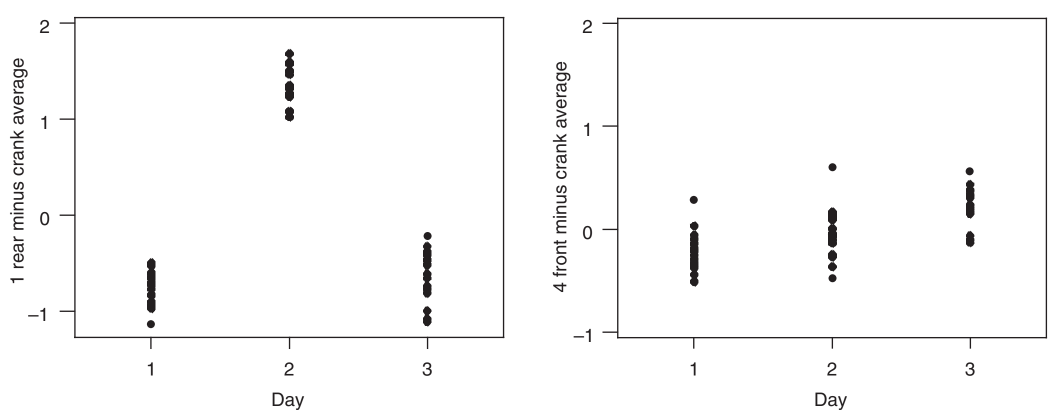


Figure III.5 Diameter minus main average by day.

Variable	Minimum	Maximum	Q1	Q3
1front	1.946	5.301	2.656	3.985
2front	-0.428	3.915	0.730	2.511
3front	-2.905	0.018	-1.827	-0.730
4front	-2.402	-0.355	-2.188	-0.858
5front	1.438	4.071	2.184	3.485

The team decided to address the measurement system instability with feedback control. The controller was feasible because:

- Quick adjustment of the measurement system was available by changing an offset.
- The shifts in the measurement process were persistent, as seen in Figure III.6.

The team developed a procedure to monitor the stability of the gages and a reaction plan. Each day, an operator measured a reference part and plotted the results on a control chart, one for each of the 15 positions. The centerlines and adjustment limits were based on the within-day measurement variation (that is, short-term variation) from the initial measurement investigation. If a plotted point fell outside the control (adjustment) limits, the gage was cleaned and remastered. The control charts provide an ongoing record of the performance of the gages. If the gages performed consistently over time on the reference part, the team had confidence in using data from the gages to make process decisions. After the feedback system was implemented, the team validated the improvement by repeating the initial measurement investigation. They found that discrimination ratios, calculated as in Table III.1, all exceeded 3.0.

The team was ready to proceed to the next step of the Statistical Engineering algorithm, confident in the measurement system and knowing they had already improved the process substantially. They were tracking the first-time-through rate at the final gage and saw a marked increase. The team considered the possible variation reduction approaches and decided to look for a dominant cause of the variation in main diameter.

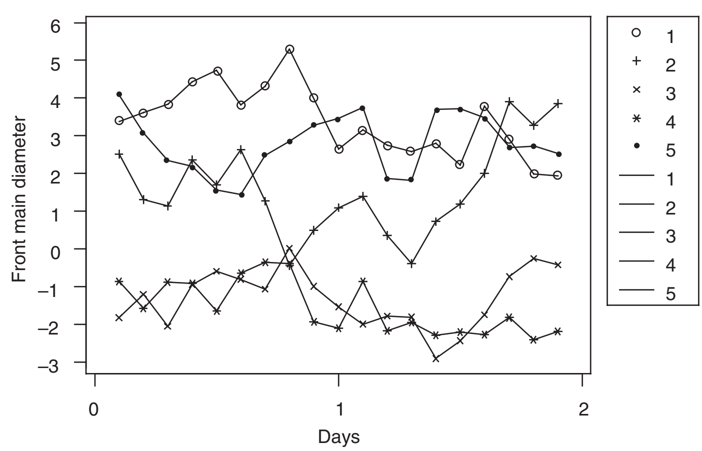


Figure III.6 Front position diameter for the five mains when measuring the same crankshaft.

Because of the faulty measurement system, they debated whether they should repeat the baseline investigation. They decided to proceed without doing so. They continued to use the original full extent of variation, –5.1 to 6.9, recognizing that this was likely too wide given the improvements they had made.

We show a simplified flowchart of the crankshaft production system in Figure III.7, highlighting the operations that affected the main diameters. The team believed initially that most of the variation in main diameters at the final gage was caused by differences in the four journal grinders. There were little data to support this belief.

The team decided to explore the families of variation defined by the process steps using a variation transmission investigation. The main diameters were measured with an in-process contact gage between the grinding and lapping operations. The team planned to measure a sample of parts with the in-process gage and then track the parts through the process and remeasure them at the final gage. They could use this variation transmission investigation to eliminate the lapper or the upstream process as the home of the dominant cause.

Based on their experience with the final gage, the team decided to next investigate the in-process gage. They carried out an investigation of this system with the same plan used for the final gage. We do not give the data here. They found the measurement variation to be relatively small (estimated standard deviation 0.31) and no evidence of instability.

The team knew that both the in-process and final measurement systems were used in an informal way to control the grinders. For that reason, the team next checked the relative bias of the two systems. They measured six crankshafts on the final gage. Then, they returned the parts to the in-process gage and remeasured the diameter. Using the same six parts, they repeated this process two more times. In the data, the intermediate (in-process) diameters have been increased by 2.25 units to reflect the fact that the nominal diameter is different at the two gages by 2.25 thousandths. The data are given in the file *crankshaft main diameter relative bias*.

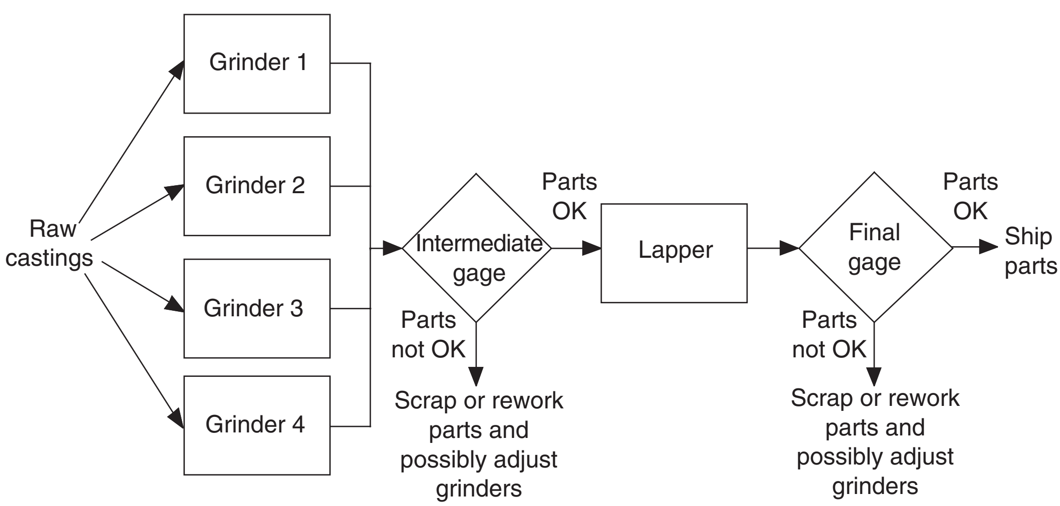


Figure III.7 Crankshaft production process map.

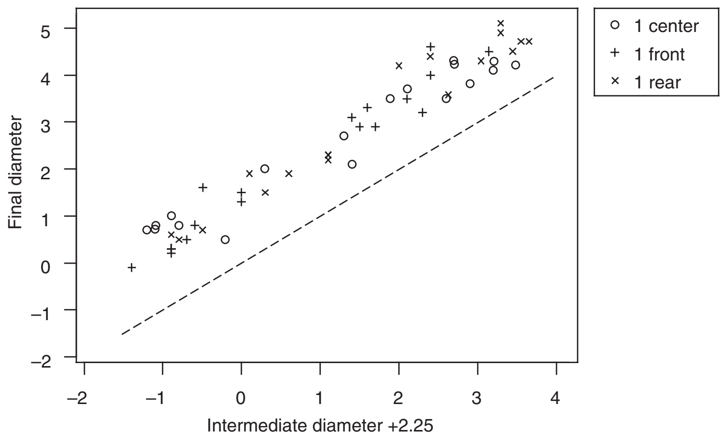


Figure III.8 Scatter plot of final versus adjusted intermediate diameter for the first main.

From the scatter plot for the first main shown in Figure III.8, we see that there is a bias between the two gages. There are similar patterns on the other mains. The team changed some of the offsets on the in-process gage to remove the bias.

The team now had confidence in the two measurement systems and recognized that they had made another improvement in the process. The first-time-through rate increased again. They proceeded with the planned variation transmission investigation to separate the effects of the lapper from those of upstream operations.

The team selected four crankshafts from each of the four grinders on four different days. Using the two gages, they measured the diameters of the 64 parts before and after lapping. The data are given in the file *crankshaft main diameter variation transmission*.

The final diameters varied between –4.3 and 5.5. This is somewhat less than the full extent of variation but, given that the process had been improved, the team was confident that the dominant cause had acted during the investigation.

In Figure III.9, we plot the final versus intermediate (in-process) diameters across all mains and positions. We see that that the intermediate diameter is a dominant cause of variation in the final diameter. The lapper transmits the upstream variation.

The team was not surprised because the lapper was a so-called dumb machine. It had no gauging or compensation and lapped for a fixed number of rotations on each main. The team expected the lapper to remove about 2.25 thousandths from the diameter. A numerical summary of the difference in diameters is:

```
Variable      N      Mean     Median     TrMean     StDev    SE Mean
diameter    960   -4.1755    -4.1000    -4.1815    0.5121     0.0165
difference

Variable     Minimum   Maximum        Q1         Q3
diameter     -5.6000   -2.4000   -4.5000    -3.9000
difference
```

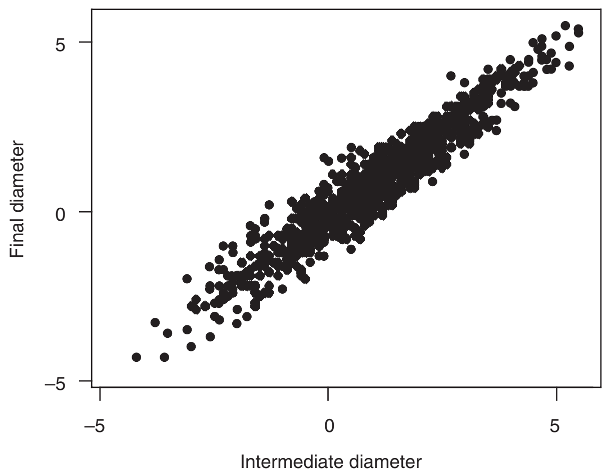


Figure III.9 Final diameter versus intermediate diameter over all positions.

Because of the offset between the two gages, the average change in diameter is –4.17 + 2.25 = –1.92, close to what was expected. The standard deviation of the change in diameter, 0.51, is small.

In Figure III.10, we look at the change in diameter by main and, more critically, by position within main. We see that the lapper removed systematically different amounts of material within mains, especially for mains 2, 4, and 5. The lapper is home to a dominant cause of the taper variation.

All parts met the taper specification ±2.0 thousandths before the lapping but not after. A summary of the before and after lapper taper is:

```
Variable        N      Mean     Median     TrMean      StDev    SE Mean
int. taper    320   -0.2440    -0.1800    -0.2313     0.7449     0.0416
final taper   320    0.1996     0.2900     0.2162     0.9421     0.0527

Variable       Minimum    Maximum         Q1         Q3
Int. taper     -1.9600     1.2500    -0.8675     0.3900
final taper    -2.6500     2.3000    -0.5650     0.9375
```

Using the results from Figure III.10, the team arranged for maintenance on the lapper to balance all shoes so that, on average, the change in diameter was consistent from main to main and from position to position within each main. They also changed the control plan for the lapper. Once per day, an operator measured a crankshaft before and after lapping, using the final gage, and plotted the change in taper for each main on a control chart. Any changes in lapper performance could be quickly identified and remedied as required.

The team next returned to the problem of excess diameter variation. They had identified the intermediate diameter as a dominant cause of the variation. The dominant cause was not verified since the conclusion made physical sense and seemed clear-cut. To select a working

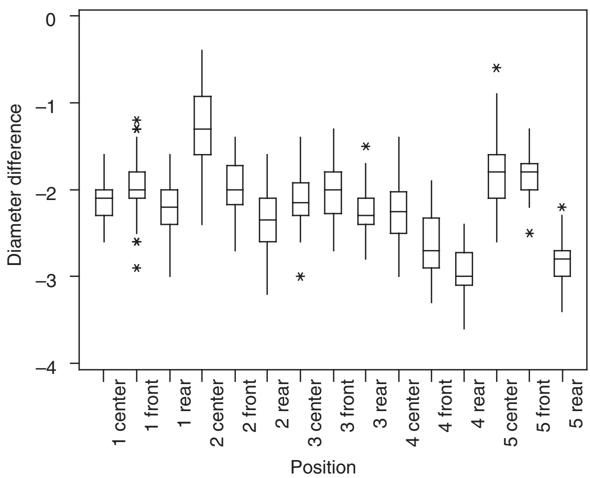


Figure III.10 Diameter difference (final minus intermediate) by position.

approach, the team briefly considered using the lapper to adjust the final diameter based on the observed intermediate diameter. They rejected feedforward control because of the cost. They would need a smart lapper that could remove a varying amount of material on each main after receiving input from the intermediate gage. They also rejected desensitization, since they did not believe any change to the lapper settings could mitigate the variation in the intermediate diameter. They decided to reformulate the problem in terms of the dominant cause.

The team used the data from the variation transmission investigation to set the baseline for the intermediate diameter. We give a numerical summary as follows and the histogram in Figure III.11.

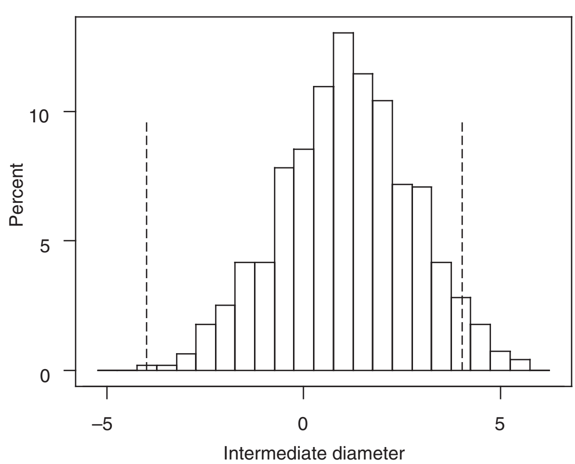


Figure III.11 Histogram of intermediate diameter (dashed lines are the specification limits ±4.0).

```
Variable        N      Mean     Median     TrMean      StDev    SE Mean
intermediate  960    1.0716     1.1000     1.0802     1.6808     0.0542
diameter

Variable        Minimum    Maximum         Q1         Q3
intermediate    -4.2000     5.5000    -0.1000     2.2000
diameter
```

The team made several observations:

- The within-process specification limits ±4.0 thousandths were the same as the final diameter specifications. There was no allowance for the addition of any variation at the lapper.
- The grinder operators ran the process on the high side to avoid scrap at the intermediate gage.
- The baseline standard deviation at the intermediate grinder was 1.68 and the full extent of variation of the intermediate diameter was –4.2 to 5.5 thousandths.

To set the goal for the reformulated problem, the team used the data from the variation transmission investigation to fit a regression model relating the final diameter to the intermediate diameter across all mains. From a portion of the MINITAB results, we have

```
The regression equation is
final diameter = -0.145 + 0.968 intermediate diameter

Predictor        Coef     SE Coef          T        P
Constant     -0.14496     0.01953      -7.42    0.000
intermed     0.968367    0.009803      98.79    0.000

S = 0.5102      R-Sq = 91.1%     R-Sq(adj) = 91.1%

Analysis of Variance

Source           DF          SS          MS        F        P
Regression        1      2540.6      2540.6  9758.54    0.000
Residual Error  958       249.4         0.3
Total           959      2790.0
```

The residual standard deviation is 0.51. If we could hold the intermediate diameter fixed, this is the estimate of the standard deviation in the final diameter. The slope of the regression equation is 0.968. This gives the expected change in average final diameter for each unit change in intermediate diameter.

The original goal was to reduce the standard deviation of the final diameter to about 1.30. Using the formula from Chapter 2, we have the equation for the required standard deviation at the intermediate diameter

$$1.30 = \sqrt{0.968^2(\text{required stdev})^2 + 0.51^2}$$

Solving, we find that the required standard deviation for the intermediate diameter is 1.24. The team set a goal to reduce the standard deviation of the intermediate diameters to 1.25 or less.

The team had already completed the Check the Measurement System stage for the in-process gage, so they proceeded to consider a working approach. The team expected that they could find a dominant cause of the variation. However, they first considered tightening the within-process specifications, that is, using 100% inspection. The current practice was to scrap parts if any diameter was less than –4.5 and to process all others with the hope that the lapper might bring the part back into specification. The team considered reducing the specifications to ±3.0 and insisting that any part outside of specification be scrapped or reworked at the intermediate gage. The team believed that the operators at the intermediate gage and the grinding process could meet these specifications on an ongoing basis. The suggestion was met with hostile resistance from the operators. Without further variation reduction effort, the team saw that by tightening the specifications, they would transfer the scrap from the final gage to the intermediate gage with little savings.

The team next considered feedback control. In the current informal system, each of the four grinder operators used their judgment plus the data from the intermediate gage. The final gage operators also informed the grinder operators when there were a large number of crankshafts out of specification. The team suspected that grinder adjustments were made only when scrap parts were produced.

The operator could quickly make an adjustment that affected all mains simultaneously but required the help of a tool setter and considerable downtime to adjust the diameter for a single main. The team decided to adopt feedback control as a working approach and to investigate further how the diameters varied over time and position.

Since taper at the intermediate gage was small and the adjuster applied simultaneously across all 15 positions, the team concentrated on the center diameter for each main. They also decided to look at one grinder only since all grinders had the same adjustment mechanism.

The team looked back at the data from the variation transmission investigation for the five center positions. We see in Figure III.12 that there were small average differences from day to day and main to main. Within each day, we see most of the full extent of variation of the diameter. The team decided to investigate the time-to-time variation within a single shift.

The team sampled and measured three consecutive crankshafts from a single grinder every 20 minutes across the eight-hour shift. The sample was about half the crankshafts produced by that grinder. They recorded the diameters from the five center positions only. During the investigation, the grinder operator was encouraged to use his usual ad hoc feedback control method. There were no adjustments during the shift. We give the data for the

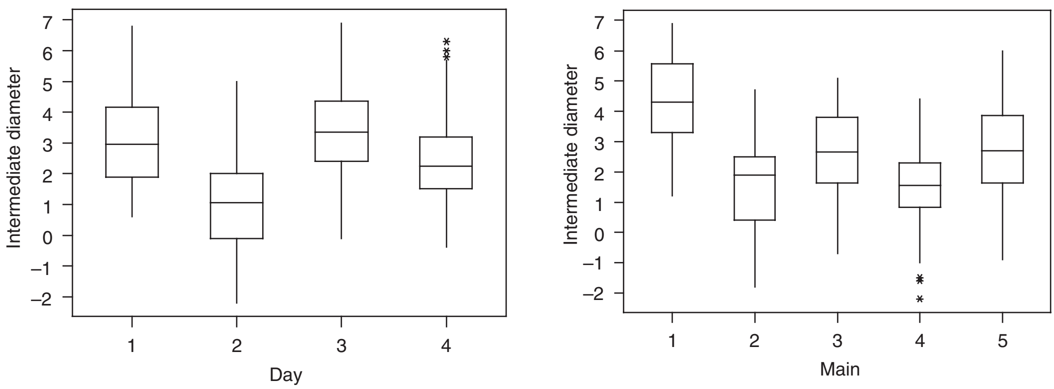


Figure III.12 Box plots of intermediate center diameter by day and main.

63 crankshafts (21 time points and three crankshafts per time) in the file *crankshaft main diameter feedback*.

The overall standard deviation of the diameters was 1.45, greater than the target value 1.25 but less than 1.68, the standard deviation of the diameter before lapping in the variation transmission study. The team used ANOVA to assess the effects of time and mains:

```
Analysis of Variance for diameter

Source      DF        SS         MS       F      P
main         4    91.187     22.797   15.96  0.000
time        20   155.538      7.777    5.44  0.000
Error      290   414.203      1.428
Total      314   660.927
```

The residual standard deviation is $\sqrt{1.428} = 1.19$. If the time-to-time and main-to-main differences could be eliminated, the within-grinder variation could be reduced to less than the target 1.25. Since the easy adjustment applied simultaneously to all mains, the team looked at the variation of the average diameter (averaged over the five mains) versus time using a multivari chart. See Figure III.13.

There was a strong trend that dominated the part-to-part variation in average diameter suggesting that feedback control would be effective. For each main, the team also plotted the average diameter (over each time period) versus time as shown in Figure III.14. There were persistent differences among the mains with main 3 being large and main 4 being small. The average diameters over all parts by main are:

main	average diameter
1	0.5048
2	0.4762
3	1.1587
4	-0.4968
5	0.6635

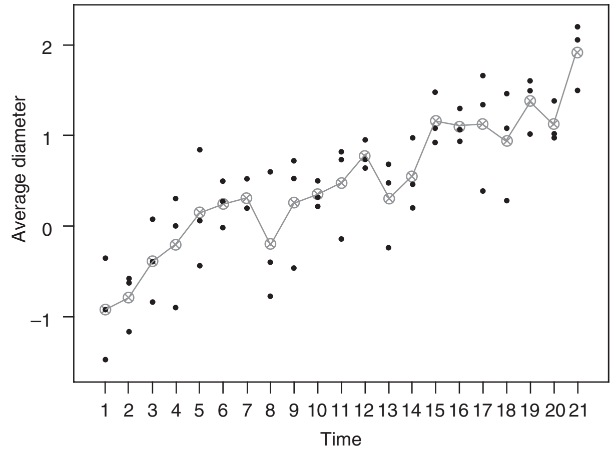


Figure III.13 Multivari chart of crankshaft average diameter versus time.

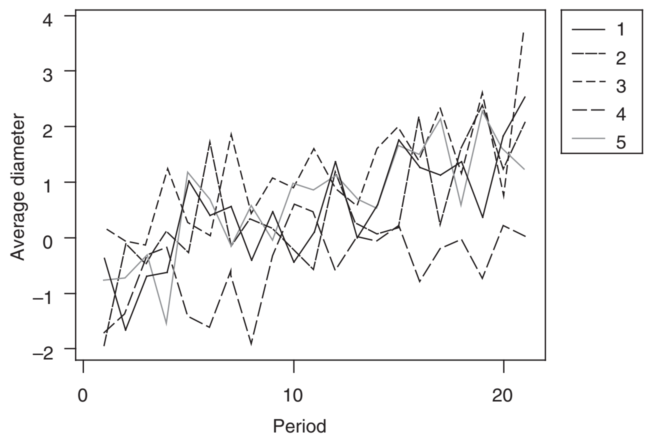


Figure III.14 Average diameter at each time period versus time period stratified by main.

The team decided to build a feedback controller using the average across all mains as the output. Separately for each grinder, they plotted a run chart of three part averages every hour. The operators were trained to look at the chart and make an adjustment if the average fell outside the range ±2.5. For political reasons, the team did not change the in-process specifications. Since the diameters drifted upwards, the target for the adjusted process was –2.0. The team carried out a small experiment on each grinder to verify that changing the adjuster produced the desired effect.

To deal with the problem of systematic differences among the mains, the team changed the section of the control plan that dealt with the setup of the grinding wheels. The tool setter had to ensure that the main averages of the first five parts after setup differed by no more than one-thousandth of an inch.

Once the changes were in place, the team sampled 100 crankshafts, 25 from each grinder, over one day from the intermediate gage. The standard deviation of the center diameters was 1.28, close to the goal of 1.25.

In summary, the team added the following changes to the process control plan:

- Monitor the stability of the final gage using a reference part and a control chart. Clean and master the gage as necessary.
- Monitor the performance of the lapper on a daily basis by measuring the change in diameter (before minus after lapping) on one part for all mains and record the results on control charts, one chart for each main. Watch for evidence of taper change. Rebalance the lapper shoes if taper becomes an issue.
- Adjust each grinder based on a feedback control scheme that uses the average across the five center positions and three consecutive parts at the in-process gage.
- Change the grinder setup procedure to ensure that all main averages are close to equal after setup.

The changes provided the process management with ongoing, timely information to better manage the process. The process engineer reacted to trends and anticipated problems before they occurred.

There were many benefits to this project. The taper rework was virtually eliminated by maintenance on the lapper. The scrap rate was reduced from 7.2% to 2.0% as measured over a one-month period. Each percent reduction in scrap saved approximately $250,000 per annum. The first-time-through rate at the final gage was held at over 90%. The crankshaft line had a large gain in productivity and a large reduction in scrap costs.

Highlights

Some strengths and weaknesses of this case are:

- Addressing both taper and diameter scrap and rework problem at the same time resulted in a complicated problem. The team needed to look simultaneously at 20 outputs (15 diameters and 5 tapers).
- The team wisely checked the stability of both the final and intermediate gages.
- To prevent recurrence of the problem, the team implemented feedback control using reference parts for ongoing monitoring of the gages and lapper.
- Not establishing a new baseline after the team made substantial improvements to the final gage was risky since it was harder to tell if a dominant cause acted in later investigations.

Exercises

Exercises

CHAPTER 1—NO EXERCISES

CHAPTER 2

2.1 The word *variation* is used in other contexts to describe a difference between a realized and target value such as in budget variation. How does this use compare to variation as discussed in Chapter 2?

2.2 We have heard the following comment many times from manufacturing engineers: "The cause of the variation is the product design—what can you expect me to do?" Discuss the comment in light of the definition of cause in Section 2.2.

2.3 Profile A is a measure of deviation of the actual from the ideal shape of a camshaft lobe over one region (A) of the lobe. The target value is zero and the upper specification limit is 250 microns. Use the data in the file *camshaft lobe runout baseline* to summarize the variation in this output. Do all lobes exhibit the same variation? Is there any time pattern in the variation?

2.4 Construct histograms and run charts for output 1 and output 2 given in the data file *chapter 2 exercise 4*. Find the average and standard deviation for each output. Assume the target value and upper specification limit for these lower-is-better outputs are 0 and 35.

a. Is the variation the same for each output?
b. Is the nature of the variation over time the same for each output?

2.5 You may convince yourself that the formulas for combining means and standard deviations given in Section 2.4 are true with the following numerical demonstration you can conduct in MINITAB. Generate two columns of 100 values sampled from some model (in MINITAB: Calc Æ Random Data Æ your choice of model and parameters, for example, Normal with mean and standard deviation 0 and 1, respectively). Then, calculate two new data columns. Let one column be the sum of the original two columns and the other the difference.

a. Find the standard deviation and average for each of the four columns.
b. Calculate the sum and differences of the averages for the first two columns. How do these compare to the average of the other two columns respectively?
c. Calculate the standard deviation for the sum and difference using the "square root of sum of squares" formula given by 2.1. How do the results compare to the standard deviations for the last two columns?

2.6 At a project review, the team presented the following summary of their investigation based on standard deviations.

Source of variation	**Percent of total**
Measurement system	30
Identified cause	50
Unidentified causes	81

a. The reviewing manager questioned the numbers in the second column of the table because they did not add to 100. Is there an error? Explain.
b. By what percentage can the process standard deviation be reduced by eliminating the contribution of the identified cause?
c. Is the identified cause a dominant cause?

2.7 In Chapter 1, we discussed a project to reduce variation in pull, an alignment characteristic of light trucks. Recall that

$$\text{Pull} = 0.23^*(\text{right caster} - \text{left caster}) + 0.13^*(\text{right camber} - \text{left camber})$$

and that the data for two months' production are stored in the file *truck pull baseline*. The data are summarized in the following table.

Output	**Average**	**Standard deviation**
Left camber	0.257	0.129
Right camber	0.249	0.130
Left caster	3.519	0.224
Right caster	4.519	0.243
Pull	0.231	0.082

a. Use the formula for pull and the results for how averages and standard deviations combine to predict the average and standard deviation for pull given by the last row in the table indirectly from the component averages and standard deviations.

b. Suppose you had the resources to reduce the variation in one of the alignment angles by 50%. Which angle would you choose? By how much, approximately, would the pull standard deviation be reduced?

CHAPTER 3

3.1 For a problem of interest to you speculate about the likely costs and feasibility of implementing each of the possible variation reduction approaches.

3.2 Variation in the location of a drilled hole in a machined casting can cause poor fits when the part is bolted to an engine. To reduce this variation, an engineer considers a variety of possible approaches.

a. A vision system is available that can measure location on 100% of the parts and reject those that it judges to be out of specification. What are the advantages and disadvantages of such an approach?
b. Institute a feedback controller by measuring two parts every hour. If hole location on either part is outside of specification, stop and adjust the process. When is such a scheme likely to be effective?
c. A third choice is to find a dominant cause of the variation. What are the advantages and disadvantages of this strategy?
d. If a dominant cause can be discovered, what options does the engineer have?

CHAPTER 4—NO EXERCISES

CHAPTER 5

5.1 Briefly discuss the advantages and disadvantages of the following—be sure to think of potential errors as described within the QPDAC framework.

a. To estimate the baseline performance of a grinding process, 100 consecutive pistons were sampled and the diameters were measured.
b. To investigate a proposed change to a chemical process, the investigators tried the change in a pilot process rather than the production process.

5.2 In the camshaft lobe BC runout problem described in Chapter 1, the team selected 50 parts (10 per day over 5 days) and measured the BC runout for each of the 12 lobes on each camshaft to quantify the baseline. The 600 runout measurements are stored in the file *camshaft lobe runout baseline original*. Conduct an analysis of these data. Are your conclusions different from those in Chapter 1? Why?

5.3 To assess a measurement system used to check the diameter of an engine bore, an investigator plans to repeatedly measure the same four (of the eight) bores on five blocks sampled from a shift of production.

a. Discuss the advantages and disadvantages of using 10 rather than 5 blocks.
b. In the investigation, all the blocks produced over one shift were available for study. Give two considerations that the investigators should take into account in making the choice of available blocks.
c. The plan was to make all measurements in a single day. Discuss the advantages and disadvantages of making the measurements over a longer time period.
d. When would the investigator be better off devoting the available resources to measuring all eight bores on fewer engine blocks?

5.4 You are a manager with the responsibility to decide if you should change the supplier for a tooling insert. You receive a report from your process engineer who has conducted an investigation into a new insert. He gives you the following verbal report and recommendation:

> Our current insert has an average life of 1105 parts. To assess the performance of the new supplier, we asked them to supply 10 inserts. We checked them the inserts out on one of our machines last week and got an average of 1300 pieces. Since the cost is the same, I think we should switch to the new inserts.

Using the QPDAC framework, think of five questions you would ask about the conduct of the investigation before you might accept the recommendation.

CHAPTER 6

6.1 In Chapter 1, we described a problem in terms of the lobe geometry of camshafts. The data are given in the file *camshaft lobe runout baseline*. Quantify the problem baseline for the following output.
a. BC runout
b. Angle error

6.2 Many programs such as Excel cannot easily handle missing observations. MINITAB is an exception. Missing values are often stored using a special numerical code (–99 is common). These special codes can result in much confusion and lead to incorrect conclusions. Consider the data *rod thickness baseline with missing observation*. In the file, there are two outputs. The output thickness_–99 uses a numerical code of –99 for missing observations, while thickness_missing uses the MINITAB missing observation symbol (*). Quantify the baseline for these two outputs. Which data summaries show the missing observation and which do not?

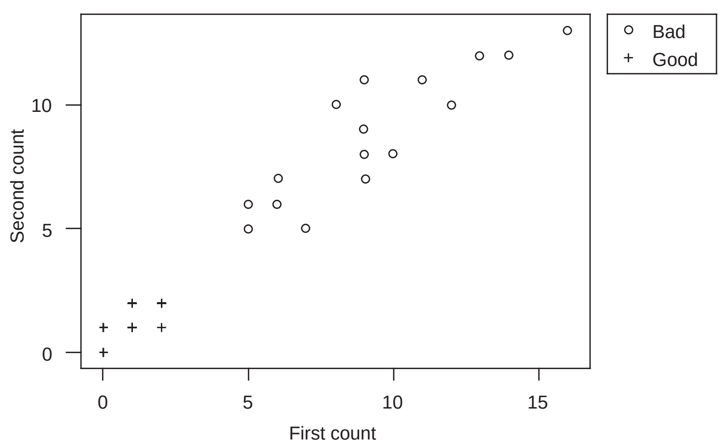
6.3 The baseline investigation for the V6 piston diameter example was described in Chapter 5. The data are given in the file *V6 piston diameter baseline*. Suppose the data were collected so that all the pistons from a given hour were collected at the start of the hour. Now the data come in subgroups as defined by hour. What summaries used in the baseline analysis are affected by the subgrouping? When taking the subgrouping into account are the conclusions any different than those derived in Chapter 5?

6.4 Based on customer complaints concerning installation difficulties, a team investigated variation of a key fascia dimension. To establish a baseline, they measured the dimension on 147 fascias sampled from one month's production. The data are given in the file *fascia dimension baseline*. Using appropriate summaries of the data, quantify the baseline. Are there any concerns?

CHAPTER 7

7.1 In a process improvement problem to improve the quality of a roof panel, the measurement system (specially designed for the project) counted the number of updings on each panel. To assess the measurement system, the number of updings on 20 bad panels and 20 good panels was counted twice. The data are given in *roof panel updings measurement*.

a. Can this investigation be used to assess the measurement variation of the counting process? Explain.
b. Can this investigation be used to assess the bias of the counting process? Explain.
c. The same operator counted all panels. Does the order in which he makes the counts matter? It is most convenient to count the same panel twice in a row. Is this a good idea?
d. A scatter plot of the first versus the second measurement is given as follows. Note that some plotting symbols correspond to more than one pair of measurements? What does the scatter plot tell you about the counting process?



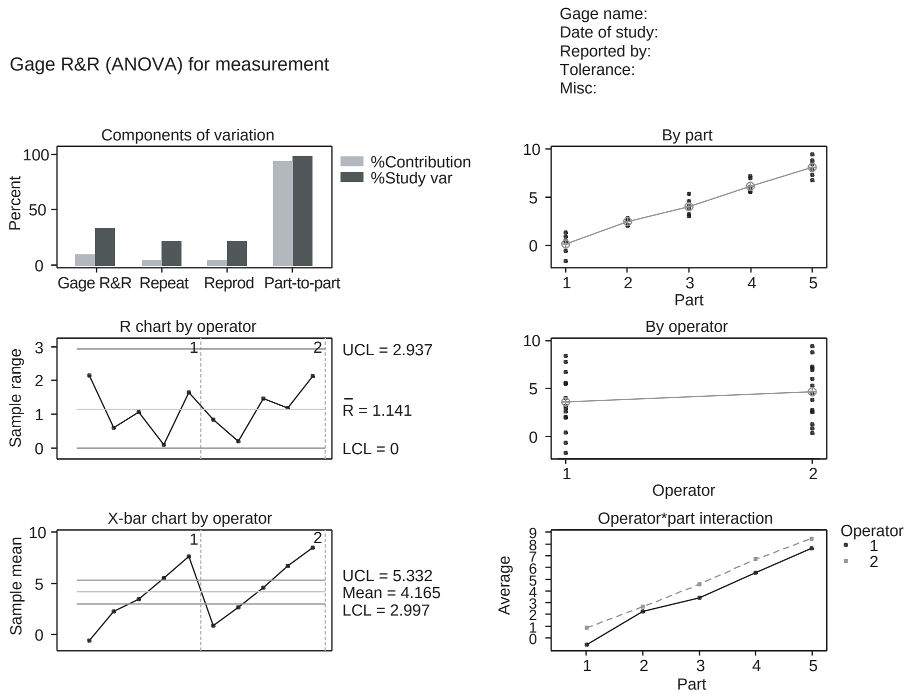
e. This investigation was conducted over one hour. What are the advantages and disadvantages of spreading the two measurements on each panel over two days?
f. Can the counting process discriminate between good and bad panels?

7.2 To monitor the process that produces engine blocks, piston bore diameters are measured on every block because they are key characteristics. Each engine block has eight bores. The bore diameter is measured at three different heights in each bore (bottom, middle, and top) and at two different orientations at each height. Because the measurement process is automated, there are no operators. A measurement investigation was conducted over a day where the diameter of every bore on four blocks was measured four times each. The main concern was out-of-round, given by 10,000 times the difference of the two diameters at a particular height. The data are given in the file *block bore diameter measurement*. From a baseline investigation the out-of-round standard deviation was 22.8.

a. Determine the discrimination ratio. Is the measurement system adequate?
b. What would have been the advantage and disadvantage of conducting the measurement investigation over a longer time period?

7.3 The following MINITAB results and graphs arise from a measurement system investigation in which two different operators measured five parts three times each. The five parts were selected with initial measured values spread out over the full extent of variation, 0 to 8. The data are given in the file *chapter 7 exercise 3*. The two operators worked different shifts so the parts were saved from one shift to the next. The results include an edited ANOVA analysis as suggested in the supplement to Chapter 7 and the default gage R&R analysis in MINITAB.

```
Analysis of Variance for measurement
Source     DF       SS       MS       F       P
part        4  230.819   57.705   81.25   0.000
Error      25   17.754    0.710
Total      29  248.573

Pooled StDev =   0.8427

Gage R&R
                              %Contribution
Source              VarComp   (of VarComp)
  Total Gage R&R      0.900      8.62
  Repeatability       0.425      4.07
  Reproducibility     0.475      4.55
    operator          0.475      4.55
Part-To-Part          9.547     91.38
Total Variation      10.447    100.00

                    StDev     Study Var   %Study Var
Source              (SD)      (5.15*SD)   (%SV)
Total Gage R&R      0.94876    4.8861       29.35
  Repeatability     0.65207    3.3582       20.17
  Reproducibility   0.68917    3.5492       21.32
    operator        0.68917    3.5492       21.32
Part-To-Part        3.08975   15.9122       95.59
Total Variation     3.23214   16.6455      100.00

Number of Distinct Categories = 5
```



a. What do the given results tell us about the bias and variation of the measurement system?
b. In the gage R&R results, the $\overline{X}$ chart by operator is out of control. What does this mean?
c. In the gage R&R results, why is the sum of the % study variation column not 100%?
d. What is the discrimination ratio (D) for this system? How does the part selection procedure influence this ratio?
e. The gage R&R is about 29%, yet D is small. Why?
f. The results suggest a small operator-to-operator difference. This observed difference may be due to a difference in method or a drift of the system over the two shifts. How can you separate these two possibilities?

7.4 To assess the variation in the system designed to measure camshaft lobe geometry over time, the same camshaft was measured daily for a month. At each measurement, a number of characteristics (for example, angle error, BC runout, taper, and so on) on each lobe were determined. The data are given in the file *camshaft lobe runout measurement stability*. Is there evidence of time-to-time variation in this measurement system?

7.5 In a process that produced V8 pistons, problems occurred when pistons in inventory were remeasured (for an audit) and found to be out of specification. Since the process used 100% final inspection, this could only occur if there was a problem with the measurement system. This was puzzling because a recent gage R&R investigation at the final gage had concluded that the measurement system was acceptable. As a result, the team decided to conduct a long-term measurement investigation.

Two pistons were chosen to span the range of diameter normally seen. Each piston was measured four times a day (spread out over the day) for 12 days. During that time the regular gage calibration was performed every four hours. The data are given in the file *V8 piston diameter measurement stability*.

a. Does the measurement system drift over time?

b. What effect does the regular gage calibration have?

7.6 Consider the brake rotor balance example described in the case studies. In the measurement investigation, three rotors were specially selected: one well balanced, another poorly balanced, and the final rotor requiring weight near the specification limit of 0.5. The three rotors were measured twice by each of the three gages on three separate days. There is no operator effect since the gages are automated. The 54 measurements are given in *brake rotor balance measurement*. The analysis given in the case study focuses on the measurement of the weight needed to balance the rotor. However, the location (or orientation) of the weight needed to move the rotor's center of gravity is also important. Can the measurement system consistently determine the orientation of the required balance weight? From the baseline investigation, the orientation of the weight was roughly uniform from 0° to 360°.

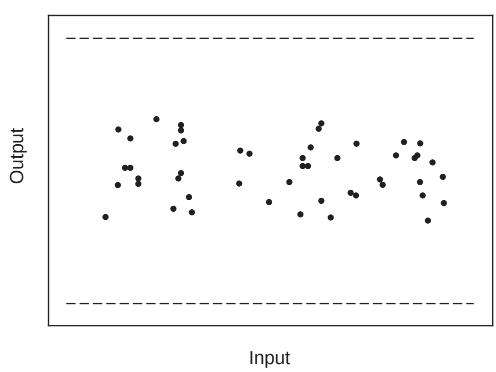
7.7 If necessary, measurement variation can be reduced by applying the Statistical Engineering algorithm. Describe how each of the seven variation reduction approaches might be used to improve a measurement system.

CHAPTER 8—NO EXERCISES

CHAPTER 9

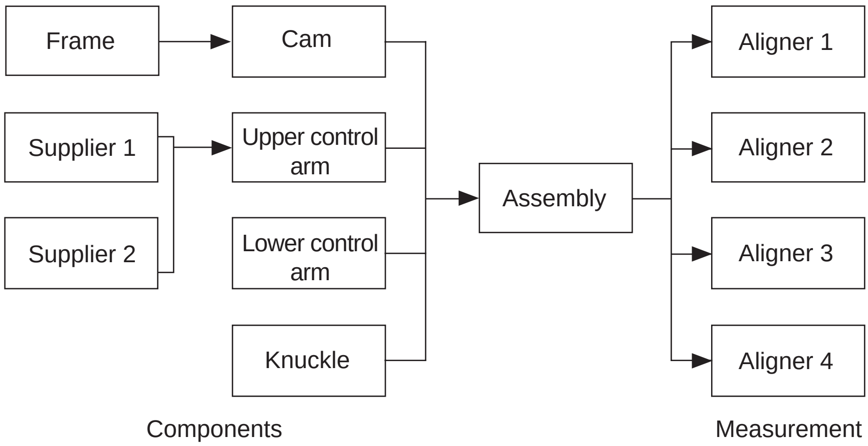
9.1 Think of a process and problem you know well. Define various families of causes.

9.2 The following plot shows the results of a process investigation aimed at finding a dominant cause. The dashed lines give the full extent of variation in the output as defined by the problem baseline. Can the input be ruled out as a dominant cause of variation?

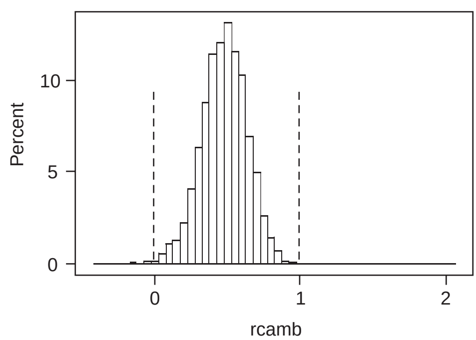


CHAPTER 10

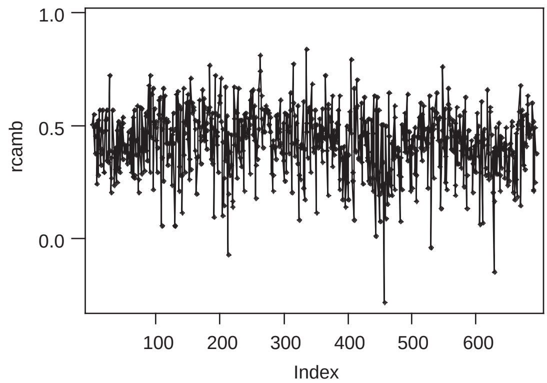
10.1 The flow chart that follows shows the major steps in an assembly process to set the wheel alignment of a truck.



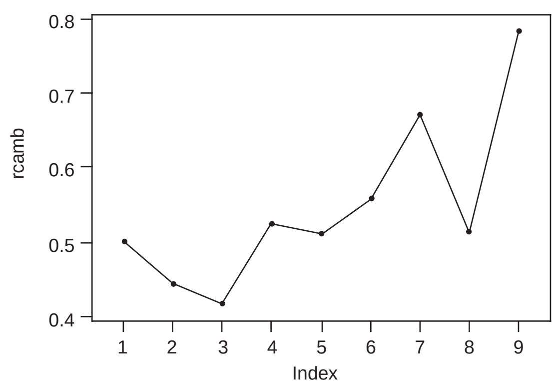
The characteristic of interest is right camber with specification 0.5 ± 0.5°. Camber is measured on every truck by one of the four gages (aligners). The process performance for right camber is shown as follows based on about 6200 consecutive trucks.



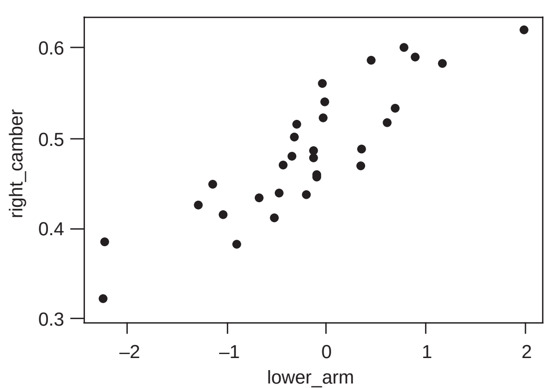
a. Based on this histogram, can the measurement system be eliminated as a dominant cause of the camber variation?
b. What data could you collect to demonstrate that a dominant cause does not act in the measurement system?
c. How could you rule out the assembly operation as the home of a dominant cause?
d. How could you eliminate differences in the suppliers of the upper control arm as the home of a dominant cause?
e. The plot that follows shows the process behavior over three shifts. What family of causes can be eliminated based on these data?



f. The following plot shows the camber variation for the first nine trucks in the data set. What families can be ruled out as the home of a dominant cause using these data?



g. In a special study, one key characteristic of the lower control arm was measured for 30 trucks. The other components were specially selected to ensure that they were well within specification. Based on the plot that follows, is the lower control arm characteristic a dominant cause of right camber variation? Explain.



10.2 Consider again the camshaft lobe runout problem introduced in Chapter 1. Each camshaft has 12 lobes with a number of characteristics of interest. In a search for a dominant cause, we may compare the lobe-to-lobe and camshaft-to-camshaft families of variation. Using the problem baseline data given in the file *camshaft lobe runout baseline,* explore the relative sizes of the two families for the following characteristics and decide which family, if any, can be eliminated.

a. Profile *A*
b. Profile *B*
c. Profile *C*

10.3 In the manufacture of an injection molded part, a key crossbar dimension exhibited excess variation. The problem baseline estimated the standard deviation of the crossbar dimension as 0.46 with full extent of variation –0.3 to 2.0. The goal was to reduce the standard deviation to less than 0.25. An investigation showed the measurement system to be highly capable.

Next the team conducted a multivari investigation where five consecutive parts were sampled every 30 minutes for four hours. Analyze the data given in *crossbar dimension multivari*. Which family of variation can be eliminated as the home of the dominant cause?

10.4 As described in Chapter 7, in a process that placed labels on bottles, the team searched for an acceptable measurement system. The file *label height measurement* contains the data from an investigation in which three operators using a hand feeler gage measured three specially chosen bottles three times on two different days. The bottles were chosen to roughly cover the range of label height values seen in the process. From a baseline investigation an estimate of the overall standard deviation was 0.022. The results of a one-way ANOVA are:

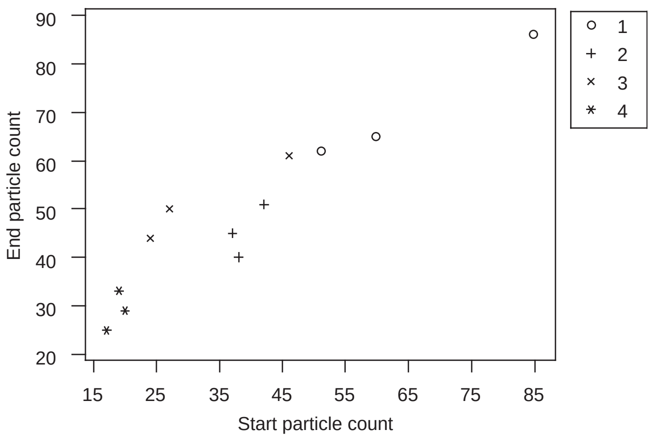
```
Analysis of Variance for height
Source     DF        SS        MS        F        P
part        2 0.0413263 0.0206631   263.10    0.000
Error      51 0.0040054 0.0000785
Total      53 0.0453317
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ----+---------+---------+---------+--
1          18   0.06966   0.00965  (-*-)
2          18   0.10950   0.00930                (-*)
3          18   0.13705   0.00749                           (-*)
                                   ----+---------+---------+---------+--
Pooled StDev =  0.00886              0.075     0.100     0.125     0.150
```

We have stdev(due to measurement) = 0.00886, and thus

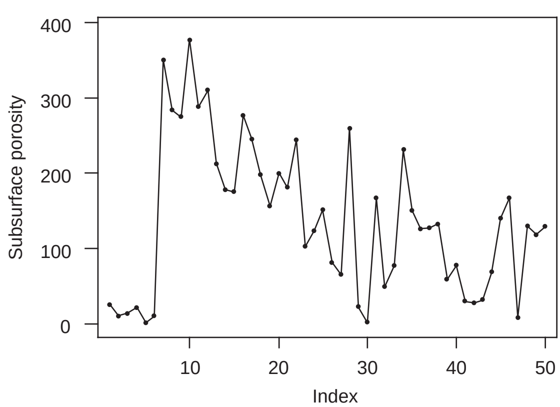
$$\mathit{stdev}(\text{due to process}) = 0.0204\left(\sqrt{(.022)^2 - (.00886)^2}\right)$$

and an estimated measurement discrimination ratio of 2.3. The team decided to improve the measurement system before addressing the original label height variation problem. Reanalyze the measurement investigation results to eliminate families of possible dominant causes of measurement variation.

10.5 A process improvement problem was initiated to reduce the number of updings on a roof panel. Updings are small outward dents in the metal surface caused by contamination. The team discovered that the dominant cause was contamination before the forming process step. In an investigation, the team measured the particle count on coils directly after the arrival from steel supplier and again after blanking and stamping (before the forming process). They measured at the tail, middle, and head of four different coils. The data are given in the file *roof panel updings variation transmission*. What does the following scatter plot tell us about the dominant cause? The plotting symbols correspond to the four different coils.



10.6 In the engine block porosity example discussed in Chapter 10, the team found the occurrence of high porosity coincided with production directly after breaks. To explore this clue further, they conducted another investigation in which the porosity of 50 consecutive blocks was measured. The first six blocks were taken from directly before the lunch break, and the next 44 blocks were the first ones produced after the break. The data are given in the file *engine block porosity run chart*. What does the observed pattern in the run chart tell us about the dominant cause?

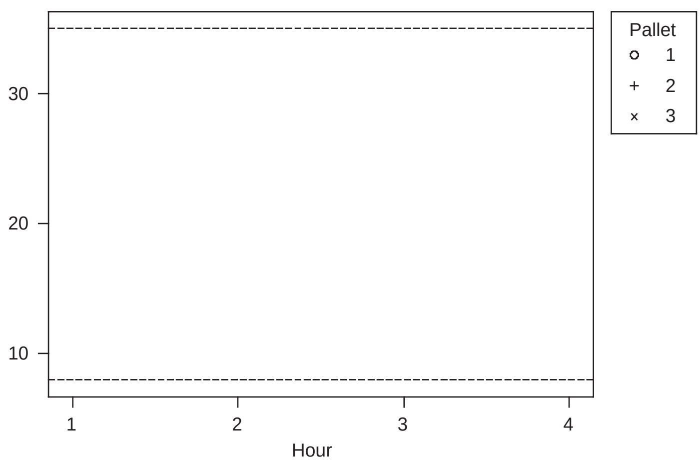


10.7 High silicon concentration in cast iron is undesirable as it was found be a dominant cause of fluidity variation. However, measuring the silicon level can be difficult. The measurement process consisted of sampling the molten iron by pouring sample coins for testing. The coins are then machined and polished before being spectrochemically analyzed. The full extent of variation in percent silicon as measured in the current process was 1 to 4%. The measurement system was investigated by repeatedly measuring three different coins that roughly covered full extent of variation in the observed percent silicon. Two operators measured each of the three coins three times on each of two days. The data are given in the file *iron silicon concentration measurement*. Analysis of the measurement results estimated the measurement standard deviation as 0.33. The corresponding estimate of the process standard deviation was 0.5; thus the discrimination is too small at around 1.5. The team needs to improve the measurement system. Using the existing measurement investigation data, are there any clues about the dominant cause of the measurement variation?

CHAPTER 11

11.1 In a multivari investigation, two consecutive pieces are selected from each of three pallets once per hour. Sketch the appearance of the multivari chart that shows all three families at the same time if a dominant cause lies in the following family. Use the following multivari chart template in which the dashed lines indicate the full extent of variation.

a. Pallet-to-pallet family
b. Part-to-part family
c. Hour-to-hour family
d. An interaction between the part-to-part and pallet-to-pallet families

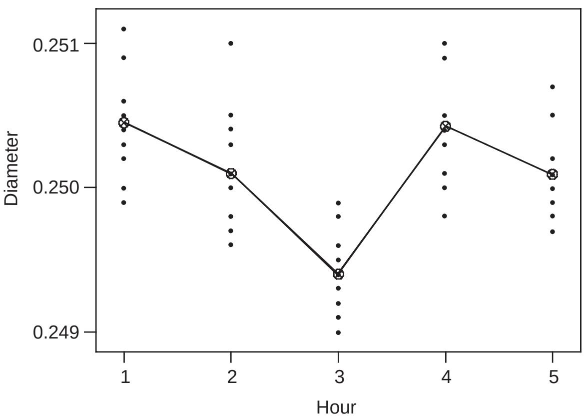


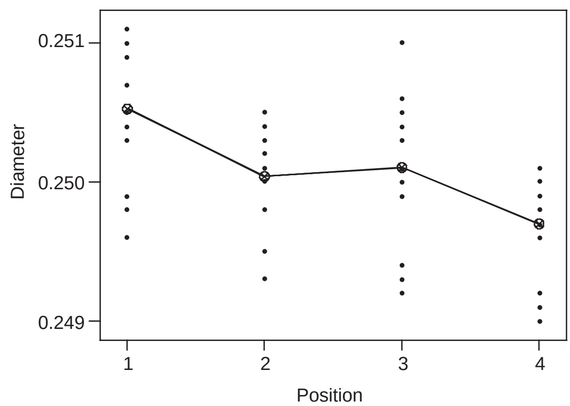
11.2 In the engine block leakers example, introduced in Chapter 1, the baseline defect rate was 2–3%. The team conducted a multivari investigation where three consecutive blocks were taken at twelve different times throughout the day. The investigation continued for three production days giving a total of 108 castings. Each block was tested for leaks. The data are given in the file *engine block leaks multivari*. What can you conclude?

11.3 At an intermediate operation the team planned a multivari investigation in which three consecutive parts were taken from each of two machines operating in parallel once every hour for two days. Consider two different processes. In the first process, the order of the parts coming from upstream is preserved, while in the second process the order is jumbled. When interpreting the resulting multivari chart (think specifically about the part-to-part family), what difference does it make which process we are observing?

11.4 In a multivari investigation, the diameter of a transmission shaft was measured at four positions (left and right side at two different orientations) for three consecutively sampled shafts each hour. The data are available in the file *transmission shaft diameter multivari*.

a. What conclusion can you draw from the multivari charts that follow?

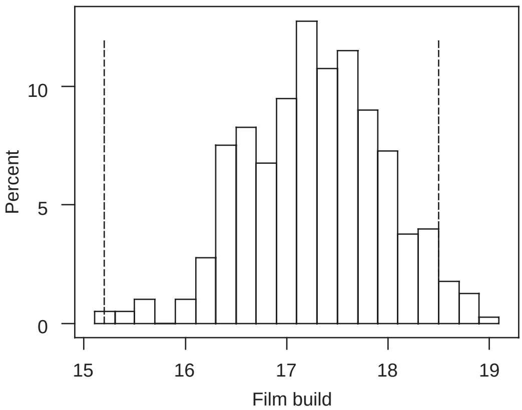


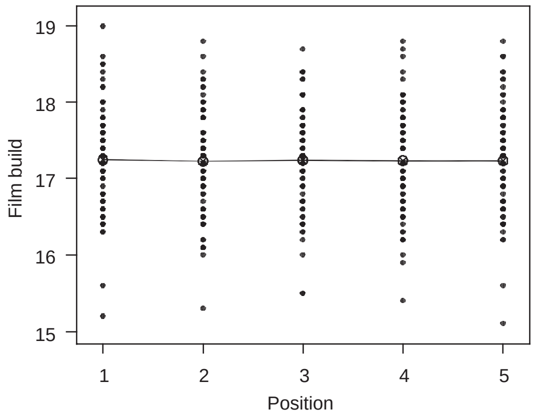


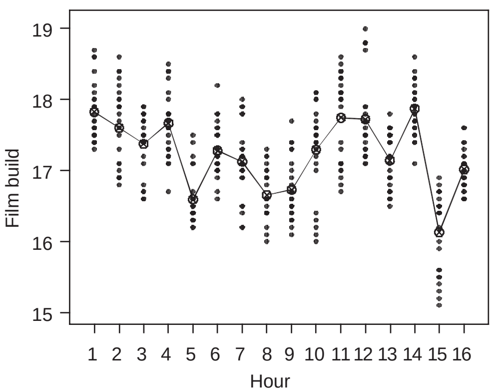
b. Using the data assess whether the dominant cause acts in the shaft-to-shaft family.

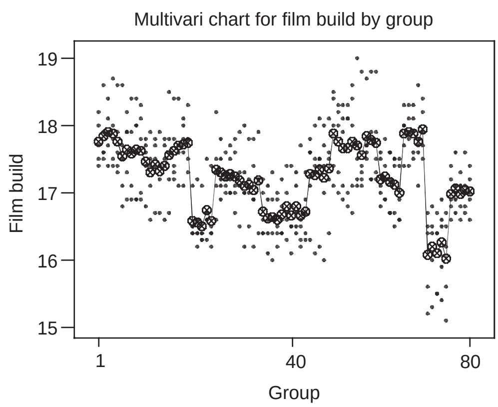
11.5 In the production of engine blocks, bore diameters are key characteristics. Bore diameter is measured at three heights and two orientations in each of the eight bores in each block. The team used Statistical Engineering to address a problem of excess bore diameter variation. The baseline investigation found a standard deviation of 3.04 and the full extent of variation of –9 to 9 as measured from nominal in microns. There were no strong differences between the different bores, heights, or positions. Another investigation concluded that the measurement process was adequate. To isolate the processing step where the dominant cause acts, the team selected 30 engine blocks haphazardly from a day's production. In the investigation the bore diameter (measured from nominal at that processing step) in the first bore at the top position and first orientation was measured at each of five processing steps in the machining part of the process. The data are given in the file *block bore diameter variation transmission*. Which processing step is home to the dominant cause?

11.6 In the paint film build example described in Chapter 3, a baseline investigation found the standard deviation was 0.315, with an average of 16.2 thousandths of an inch. The full extent of variation was 15.2 to 18.5. To search for a dominant cause, the team conducted a multivari investigation where they measured the film build at five positions on five cars in a row every hour for two shifts (16 hours). This resulted in a total of 400 film build measurements. The data are given in the file *paint film build multivari*. Based on the plots that follow, what conclusions can you draw? We define group as (hour – 1) ¥ 5+ position.



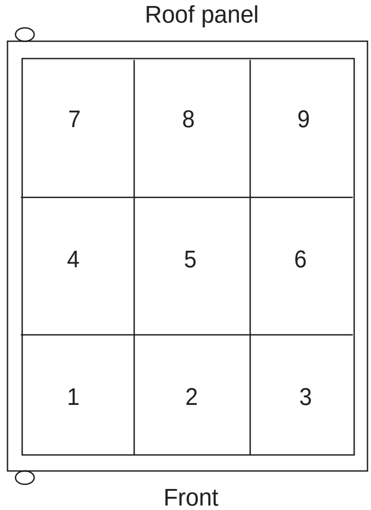






11.7 A team wanted to reduce the number of updings on a roof panel. Updings are small outward dents in the metal surface caused by contamination. A baseline investigation found that the total number of updings in 20 consecutive panels ranged between 5 and 438. To search for a dominant cause the team conducted a multivari investigation where the number of updings was counted for 20 consecutive roof panels from three sections of seven different pallets of steel sheets. Originally, the plan was to repeat this data collection over two separate days. However, the team found the full extent of variation from the baseline was observed on the first day so they stopped collecting data. The data are given in the file *roof panel updings multivari*.

a. Analyze the data using multivari charts and draw conclusions.
b. When the number of updings was counted they were classified into one of the nine locations as numbered in the schematic that follows. Analyze the multivari data using a concentration diagram based on the given schematic.



11.8 The baseline investigation for the sand core example discussed in Chapter 1 involved taking five samples over a single day of five consecutive shots of four cavities each. The data are given in *sand core strength baseline*. What conclusions can you draw?

CHAPTER 12

12.1 Vehicle plant and warranty complaints for loose spark plug wires at the spark plug end prompted an improvement project. As a result of several investigations, the family of causes related to push forces on the wires was the home of a dominant cause. A further investigation then compared eight loose and eight good connections. For each of the 16 connections, the team measured the terminal position of wires and terminal runout of the spark plug in millimeters. The data are given in the file *spark plug connection comparison*. What do the data tell us about the dominant cause?

12.2 A sunroof installation process suffered from a 90% rework rate due to a lack of flushness. Flushness is defined as the difference in height between the sunroof seal and the metal roof. It is measured using digital calipers at six points (three at the front and three at the back). A baseline investigation showed that flushness problems were most pronounced at the two front corners with full extent of variation between –3.5 to 4 mm and standard deviation 1.25 millimeters. A goal of reducing the front corner flushness variation to 0.5 and a range of –2 to 1 millimeters was established. Based on engineering knowledge, the team felt that only two types of characteristics could lead to flushness variation, namely roof crown height and attachment pad height. When the roof is adapted to allow installation of a sunroof, six installation pads are added. Based on this knowledge, the team selected six vehicles with large positive flushness and six vehicles with large negative flushness on both front corners. The sunroof modules were removed and the six attachment pad heights and roof crown height were measured at the front and back. The data are given in the file *sunroof flushness input-output*. What conclusions can you draw?

12.3 An example related to sand defects in manifolds was discussed in Chapter 12. Before the problem mentioned in Chapter 12, the team carefully monitored the process for a shift. Each of 970 manifolds was classified as either scrap (due to sand issues) or not scrap. In addition many inputs relating to each manifold, including some discrete inputs such as mold number and continuous inputs such as pour time were recorded. In some cases the linkages were difficult to establish, and the team did the best they could. The data are given in the file *manifold sand scrap comparison*. What conclusions can you draw?

CHAPTER 13

13.1 In a verification experiment there were two suspects at two levels. The low and high levels for each suspect were chosen based on the extremes from historical variation. The results of the first three runs of the experiment are shown in the following table.

Input *A*	Input *B*	Order	Output
Low	Low	2	13
Low	High	3	16
High	Low	1	17
High	High	4	?

Given that the full extent of output variation is 12 to 30, what conclusions can you draw about the dominant cause?

13.2 In the engine block porosity example discussed in the text and exercises of Chapter 10, a dominant cause of porosity acted immediately following scheduled breaks in production. Based on this evidence, the team identified two suspects: iron pouring temperature and the addition of ladle wash. During work stoppages, iron that remained in the six pouring ladles cooled off because there was no external heat source. At the start of the break, ladle wash was added to the ladles to protect the refractory (surface). The team could not easily manipulate the pouring temperature, but they could change the amount of ladle wash. They conducted a verification experiment in which they added the normal amount of wash to ladles 1, 3, and 5 and half the normal amount to the other three ladles over two lunch breaks. At each break, they measured the porosity of the first 30 blocks poured (five from each ladle). The data are given in the file *engine block porosity verification*.

a. What have we learned about the identity of the dominant cause of porosity?
b. Explain how the effects of ladle number and the presence or absence of ladle wash are confounded. Does this matter?

c. Is it a problem that in this verification experiment we have not observed the behavior of the process before lunch breaks?

13.3 The manufacture of a tube assembly required a protective nylon sleeve to be positioned and bonded to a tube. The bond strength of this tube assembly was occasionally tested using a destructive test where the sleeve was subject to increased tensile shear load until failure. In the current process, the average pull-off force was around 15 pounds, but roughly 8% of assemblies tested had a pull-off force less than the desired minimum of five pounds. The team decided to try to solve the problem by reducing the variation in pull-off force rather than by increasing the average pull-off force. A number of investigations were conducted to find the dominant cause. A multivari investigation suggested that the dominant family of causes was tube-to-tube. At this point, the team decided to conduct an experiment to search for a dominant cause using the limited process information they had gathered. They planned a factorial experiment with three suspects—clearance between the sleeve and tube, amount of adhesive, and cure time—all consistent with the tube-to-tube family clue. The team chose the low and high levels of each suspect to roughly match their range in regular production. The levels of clearance were achieved by sorting sleeves and tubes. There were two replicates of each treatment, and the run order was randomized. The data are given in the file *nylon bond strength verification* and summarized in the following table:

Treatment	Order	Clearance	Adhesive	Cure time	Bond strength
1	6, 7	Low	Low	Low	26, 28
2	14, 8	High	Low	Low	10, 10
3	2, 3	Low	High	Low	25, 26
4	12, 5	High	High	Low	7, 9
5	16, 4	Low	Low	High	24, 27
6	13, 10	High	Low	High	12, 13
7	15, 11	Low	High	High	23, 21
8	1, 9	High	High	High	7, 7

What do the results tell us about the dominant cause?

13.4 Steering knuckles are produced in a gray iron casting process. Around 2% of castings were scrapped because the percent nodularity was too small. In this example the team did not clearly establish a problem baseline. The team thought the cause must be related to the inoculation of the molten iron using a silicon-based alloy. The inoculant was added as the iron was poured to increase nodularity (and thus casting strength). Based on observing the process, the team noticed that the amount of inoculant added by the automated delivery system seemed to vary. The desired amount of inoculant was obtained by slowly shaking the inoculant onto a plate. The plate was designed to tip automatically when the required weight of inoculant was present. The team saw that the location on the plate where the inoculant fell varied, and they thought that this might influence when the plate tipped and thus how much inoculant was delivered. The team decided to verify inoculant amount as the dominant cause of nodularity variation. In the verification experiment, they produced a total of 20 castings at each of two levels of inoculant amount, 12.3 and 13.5 grams. For the experiment the inoculant was carefully weighed and added by hand. The experiment consisted of eight runs of five castings each. The order of the eight runs (four at each level) was randomized. For each of the 40 castings the percent nodularity was determined. The data are given in the file *steering knuckle strength verification* and are summarized in the table that follows:

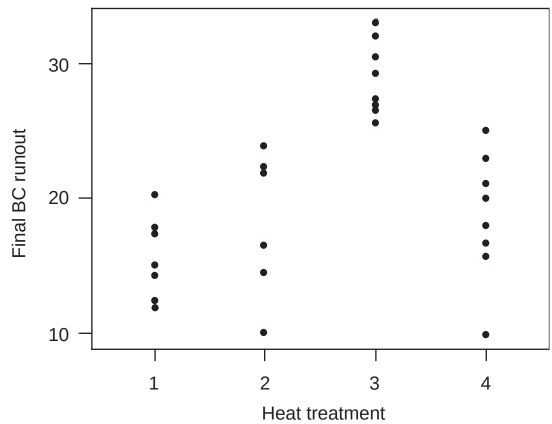
Run	Inoculant amount	Order	Percent nodularity
1	12.3	2	81.8, 79.4, 80.3, 80.6, 79.3
2	12.3	3	79.8, 77.0, 77.8, 79.3, 78.7
3	12.3	8	80.9, 82.0, 80.6, 80.6, 81.1
4	12.3	4	81.0, 79.4, 77.0, 80.6, 80.2
5	13.5	5	82.5, 86.1, 82.3, 83.5, 85.2
6	13.5	7	82.1, 84.6, 83.9, 85.0, 85.6
7	13.5	6	85.0, 87.8, 83.1, 84.0, 84.4
8	13.5	1	85.0, 84.3, 86.3, 83.8, 82.9

a. What considerations should the team have used in choosing the two levels for inoculant?
b. Why was randomizing the order of the runs important?
c. Has the team verified the dominant cause of nodularity variation?

CHAPTER 14

14.1 In the camshaft lobe runout example, the team searched for a dominant cause of variation. As discussed in Chapter 10, they conducted a variation transmission investigation where runout was measured directly before heat treatment and after the final step of the process, on the same 32 parts selected over the course of one day. In the investigation the grinder (one of eight) and heat treatment spindles (one of four) used were also recorded. The data are given in the file *camshaft lobe runout variation transmission*. They found that a dominant cause of variation was the

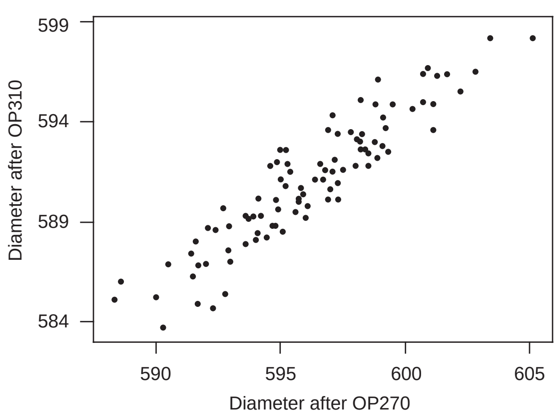
BC runout just after heat treatment and, more specifically, as shown in the plot that follows, that heat treatment spindle was a dominant cause.



In this example, the team decided not to reformulate the problem but to look for a more specific cause.

a. Discuss the advantages and disadvantages of the decision not to reformulate.
b. Suppose the team had reformulated the problem based on heat treatment spindle and that the original goal was to reduce the final runout standard deviation to less than 4.5. Using the results from a one-way ANOVA model based on heat treatment spindles, derive a goal for the new problem based on differences among spindle averages.

14.2 In Chapter 11, the team found that the piston diameter directly after operation 270 was a dominant cause of final V6 piston diameter variation. The relationship is illustrated in the scatter plot that follows. The data are given in the file *V6 piston diameter variation transmission*.



The team decided to look further for a more specific dominant cause. Suppose, however, they had wanted to reformulate the problem in terms of the operation 270 diameter. Determine an appropriate goal for the reformulated problem. Recall that the goal for the original problem was to reduce the final diameter standard deviation to less than 2.0.

CHAPTER 15

15.1 Based on customer complaints about assembly difficulty, a team investigated fascia dimension variation. A baseline investigation found that some fascias were too large. The team felt that reducing the average size of the fascias could solve the problem (that is, they adopted the Move the Center approach). They planned a (full) factorial experiment with two candidates, cycle time and cure time, each at two levels to look for an adjuster. They chose the levels for each candidate based on engineering judgment. The results of the experiment are given in the file *fascia dimension move center* and in the following table. For each treatment, the team conducted four runs producing 10 fascias for each run. The order of the 16 runs was randomized over a day. In the data, we give only the average fascia dimension from each run and not the individual values.

Treatment	Run order	Cycle time (minutes)	Cure time (hours)	Average fascia size (from nominal)
1	8, 10, 1, 14	85	5	4.50, 5.23, 5.75, 6.51
2	6, 15, 12, 2	113	5	7.12, 8.25, 9.06, 9.28
3	11, 3 , 9, 16	85	19	3.65, 3.75, 4.27, 5.34
4	13, 5, 4, 7	113	19	4.24, 6.31, 7.15, 8.22

a. Can cycle time or cure time be used as an adjuster?
b. Suppose the goal was to reduce the average fascia size to 3.0. What do you recommend?
c. What is the advantage of looking at the dimensions for all the fascias within a run, rather than the averages?

15.2 An experiment was carried out to investigate four candidates to search for an adjuster of the formability safety margin of galvanized sheet metal trunk lids. The purpose was to increase the average safety margin from the baseline value of 10.7. In the experiment, each candidate was tested at two levels, selected to be near the edge of what was physically possible—see the table that follows. Note that none of the treatments corresponded to the existing process settings.

Treatment	Run order	Tonnage	Lubrication	Blank size	Prebending	Safety margin
1	6	310	Unlubricated	949__1494_mm	No	8
2	16	375	Unlubricated	949__1494_mm	No	11
3	3	310	Lubricated	949__1494_mm	No	12
4	11	375	Lubricated	949__1494_mm	No	0
5	7	310	Unlubricated	965__1500_mm	No	13
6	15	375	Unlubricated	965__1500_mm	No	6
7	4	310	Lubricated	965__1500_mm	No	11
8	13	375	Lubricated	965__1500_mm	No	1
9	1	310	Unlubricated	949__1494_mm	Yes	18
10	12	375	Unlubricated	949__1494_mm	Yes	17
11	5	310	Lubricated	949__1494_mm	Yes	18
12	14	375	Lubricated	949__1494_mm	Yes	10
13	2	310	Unlubricated	965__1500_mm	Yes	8
14	10	375	Unlubricated	965__1500_mm	Yes	12
15	8	310	Lubricated	965__1500_mm	Yes	16
16	9	375	Lubricated	965__1500_mm	Yes	8

Press tonnage was very difficult to change so all eight runs with low press tonnage were carried out first. Within each group of eight runs, the order was randomized. The data are given in the file *sheet metal move center*.

a. Analyze the experimental data to see if any of the candidates is an adjuster.
b. Does the restriction on randomization required for this experiment make any difference to the conclusions we can draw?

15.3 In the sand core strength example introduced in Chapter 1, too many cores were breaking during handling. A suggested solution was to increase the core strength (and thereby reduce core breakage) by increasing the resin concentration. It was known that increasing the resin would result in a stronger core. However, the precise nature of the relationship—that is, how much the core strength increases for a given change in resin concentration—was not known. An experimental investigation was planned to quantify the relationship. Three levels of resin concentration (1.3, 1.6, 1.9% by weight) were chosen based on engineering judgment. In the experiment, 40 cores for each level of resin were produced; 15 were measured for strength (using a destructive test) and the remaining 25 were processed to look for casting problems.

The experiment consisted of three runs with 15 repeats. The order of the runs was not randomized. The data are given in the file *sand core strength move center*.

a. What can you conclude about the relationship between resin concentration and core strength?
b. The team used only three runs with 15 repeats for each run. Discuss the advantages and disadvantages of this plan compared with using five replicates for each treatment with three repeats each.

CHAPTER 16

16.1 In a sonic welding operation, problems arose due to poor weld strength, measured as pull-off force. The goal was to reduce the variation and increase the average pull-off force. The second goal is not addressed here. From the baseline, the full extent of variation for pull-off force was 0.9 to 3.0. The team discovered that the dominant cause acted in the time-to-time family. While they could not be more specific, the team felt that the dominant cause was related to material hardness, which was outside their control. They decided to try to desensitize the process to variation in the dominant cause.

The team planned a fractional factorial experiment with four candidates at two levels each in eight treatments. Using the results of regular process monitoring, they identified three time periods when weld strength was low, medium, and high relative to the baseline. In each period, they randomized the order and then produced a part with each of the eight treatments. The pull-off force data and plan are given in the file *sonic weld desensitization* and the table that follows. The three values in the columns Order and Pull-off force correspond to the three different time periods. The original settings of the candidates correspond to treatment 2.

Treatment	Order	A	B	C	D	Pull-off force
1	7, 2, 6	–1	–1	–1	–1	1.8, 2.1, 2.3
2	4, 6, 1	–1	–1	1	1	0.9, 2.0, 2.7
3	6, 3, 4	–1	1	–1	1	2.0, 2.4, 2.1
4	8, 1, 2	–1	1	1	–1	0.6, 1.5, 3.0
5	3, 7, 8	1	–1	–1	1	2.8, 2.9, 3.0
6	2, 4, 3	1	–1	1	–1	2.1, 3.0, 3.7
7	1, 8, 7	1	1	–1	–1	2.9, 3.2, 3.1
8	5, 5, 5	1	1	1	1	2.3, 2.5, 4.0

a. Explain why the team believed the dominant cause acted over the three runs for each treatment.

b. What levels of the candidates do you recommend to reduce the variation in pull-off force?
c. Another way to assess the results of this experiment is to summarize the output across each treatment using log standard deviation. Using this performance measure, do your conclusions differ from part b?

16.2 In the crossbar dimension example discussed in Chapter 12, the team found that the dominant cause of dimension variation was barrel temperature. Because it was hard to control in regular operation, the team decided to try to make the process less sensitive to barrel temperature variation. In the current process, barrel temperatures ranged over roughly 4°C. The team planned a half fraction factorial experiment with three candidates—target barrel temperature, injection pressure, and material—at two levels each, as shown in the following table. The current injection pressure and target barrel temperature were 1000 and 75, respectively. Note that although the variation in barrel temperature was the dominant cause, the target barrel temperature is a fixed input. Five crossbars were produced and measured in each run. For each treatment, there were two runs, one at the target barrel temperature plus 2°C and the other at the target barrel temperature minus 2°C. The data are given in the file *crossbar dimension desensitization* and in the table as follows.

Treatment	Target barrel temperature	Injection pressure	Material	Dimensions at barrel temperature –2°C	Dimensions at barrel temperature +2°C
1	75	1000	Old	0, –0.1, 0.1, –0.1, –0.2	0.5, 1.1, 0.8, 0.9, 0.7
2	75	1200	New	1.1, 0.6, 1.0, 1.4, 1.1	1.5, 1.8, 1.5, 1.4, 1.3
3	79	1000	New	1.1, 1.0, 1.3, 0.9, 0.8	1.0, 1.1, 0.8, 0.9, 1.0
4	79	1200	Old	1.2, 1.8, 1.8, 1.7, 1.9	2.3, 2.1, 2.4, 2.1, 1.9

Since the average dimension can be easily adjusted, we focus the analysis on finding a way to make the process less sensitive to barrel temperature variation.
a. What levels of the candidates do you recommend?
b. Injection pressure and material were chosen as candidates based on engineering judgment. Looking again at the results presented in Chapter 12, what motivates the choice of target barrel temperature as a possible candidate?

16.3 In Chapter 16, we describe a desensitization experiment for the refrigerator frost buildup example where each refrigerator is subjected to only two extreme levels of environmental causes. Here we consider a hypothetical experiment in which each refrigerator is exposed to a number of environmental conditions to ensure that any chosen new design works well under any conditions, not just extreme conditions.

The experimental design for the four candidates—D1, D2, D3, and D4—is the same as in Chapter 16. Here we plan to test each of the eight refrigerators (treatments) under all eight possible combinations of the usage or environmental inputs as given in the following table:

	Cause combination							
Varying input	1	2	3	4	5	6	7	8
Ambient temperature (°C)	26	26	26	26	32	32	32	32
Relative humidity (%)	70	70	90	90	70	70	90	90
Door openings per hour	4	8	4	8	4	8	4	8

The experimental plan had 64 runs. To conduct the experiment, all eight refrigerators were simultaneously placed in a test chamber and exposed to each cause combination in the given order. The cooling plate temperatures are given in the file *refrigerator frost buildup desensitization2* and in the following table:

	Candidates				Cooling plate temperatures (in cause combination)							
Treatment	**D1**	**D2**	**D3**	**D4**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	N	O	O	N	3.6	3.9	4.6	1.0	4.4	0.1	4.4	0.7
2	N	O	N	O	5.1	4.7	4.3	2.9	4.2	4.1	7.1	5.1
3	N	N	O	O	4.6	4.6	4.3	4.9	2.4	5.0	5.5	16.0
4	N	N	N	N	3.8	12.8	6.9	6.9	7.1	6.7	3.0	15.7
5	O	O	O	O	2.9	0.2	–0.2	–0.2	–0.2	–0.2	–0.2	16.0
6	O	O	N	N	0.1	1.9	0.8	1.3	5.9	5.1	0.0	14.7
7	O	N	O	N	0.7	0.8	0.1	0.1	0.4	0.2	0.1	–0.1
8	O	N	N	O	0.2	3.4	0.3	1.0	4.0	0.2	5.2	16.0

In the table, we have coded the new and original settings for the candidates as N and O, respectively. What conclusions can you draw? Remember, the goal is to desensitize cooling plate temperature to changes in the environmental conditions.

16.4 There were excessive failures in the accelerated life testing of electric motors. Using a group comparison investigation, the team found that unevenness in the commutator shaft surface was a dominant cause of these failures. The team next reformulated the problem to one of reducing the unevenness in the commutator shaft. The surface unevenness is measured on a scale of 1 (smooth) to 10 (rough). With further investigation, the team determined that the dominant cause of the variation in the (final) smoothness was the shaft profile before machining. The team adopted the Desensitization approach. They decided to conduct a fractional factorial experiment with eight treatments using four candidates. For each of the eight treatments there were two runs, one that used a shaft with a premachined smooth profile, and a second that used a rough profile. The experimental plan and data are given in the file *electric motor failure desensitization* and the table that follows. The order of the runs was randomized.

						Smoothness	
Treatment	**Depth**	**Grind time**	**Rotational speed**	**Feed rate**	**Order**	**Smooth profile**	**Rough profile**
1	Shallow	Short	1800	Slow	4, 5	2	7
2	Deep	Short	1800	Fast	6, 11	3	8
3	Shallow	Long	1800	Fast	1, 14	1	9
4	Deep	Long	1800	Slow	16, 12	2	8
5	Shallow	Short	2400	Fast	13, 9	3	2
6	Deep	Short	2400	Slow	10, 8	1	4
7	Shallow	Long	2400	Slow	3, 7	2	3
8	Deep	Long	2400	Fast	15, 2	3	5

a. What is the confounding structure of the design? What limitations does this introduce?
b. What conclusions can you draw?
c. What would be the advantages and disadvantages of measuring the time to failure using the accelerated life test for each run rather than judging the smoothness of the commutator surface after machining?

CHAPTER 17

17.1 In an investigation, 100 trucks were selected from regular production over two weeks. The frame geometry as given by four summaries (left and right front, left and right rear) and the alignment outputs left and right camber and caster were determined for all 100 trucks. The data are given in the file *truck pull feedforward*. In Chapter 17 an analysis determined that feedforward control based on frame geometry was feasible for left caster. Repeat the analysis for the other outputs: right caster, left camber, and right camber.

17.2 Engine assembly problems occurred due to a poor fit between the pistons and the engine bore. The dominant cause of poor fit was found to be variation in the clearance, the difference between the (minimum) bore diameter and the (maximum) piston diameter. To solve this problem, the team thought about using the feedforward (selective fitting) approach. The idea was to measure each piston diameter and place them into bins of similar diameter. Then, after each bore diameter was measured, a piston would be selected from the appropriate bin. To assess this proposal the diameter measurements for 469 pistons and bores, as measured from nominal, are given in the file *block bore diameter feedforward*. Quantify the expected reduction in clearance variation when using one (that is, no selective fitting), two, three, or four bins of pistons. A suggestion is to define the bins by dividing the range in piston and bore diameters (roughly –10 to 10 microns) into equal widths.

17.3 In the V6 piston diameter example discussed in Chapter 11, the team found that piston diameter after Operation 270 was a dominant cause of the final diameter. The data are given in the file *V6 piston diameter variation transmission*. This suggested that feedforward control might be a feasible approach.

a. What are the requirements for feedforward to be feasible in this context?
b. If feedforward were feasible, assess the potential benefit using the results of the variation transmission investigation.
c. Could the team also use the diameter after Operation 200, rather than the diameter after Operation 270, as the input to a feedforward controller?

CHAPTER 18

18.1 The bias of the system used to measure camshaft journal diameters tended to increase over time. The cause of this increase was not determined. Instead, the team introduced a feedback controller. At the start of each shift, a master journal with known diameter was measured. If the measured diameter deviated from the known value by a large amount, the measurement system was recalibrated.

a. How could we sensibly define a large deviation in this context?
b. What would happen to the measurement variation if the measurement device were recalibrated every time the master journal was measured, rather than only when the deviation from the known dimension was large?

18.2 In a machining process, the dominant cause of dimension variation acted in the setup family. That is, the dimension variation within a particular setup was small relative to the variation from one setup to the next. The existing control plan called for a complete process adjustment back to the target based on the first observation after each setup. There were more than 200 parts machined between setups. The baseline dimension standard deviation was 0.31. The team decided to explore a new feedback control scheme based on the average for the first five observations after each setup. In an offline investigation, they carried out 10 setups and produced 20 parts after each setup without any adjustment. The dimension data, scaled so that the target is zero, are given in the file *machining dimension feedback*.

a. Use a one-way ANOVA to estimate the standard deviation if the process could be adjusted so that the dimension averages across all setups were equal.
b. Use simulation to compare the performance of the existing feedback controller with the proposed controller that makes a complete adjustment based on the average for the first five observations after each setup.
c. In general, we may design a feedback controller by averaging the output from the first *n* observations after each setup. What considerations help you decide how many observations should be used to estimate the process average after each setup?

18.3 In a machining process, there was excess variation in the diameter of a precision ground shaft. The shaft diameter was measured for all shafts using a complex automated gage (that also measured other outputs). Upon investigation, the team discovered that the dominant cause acted in the measurement family. In particular, the measurement bias changed from day to day, consistent with the pattern observed in the baseline. To explore this bias change further the team planned an investigation where the diameter of the same shaft was measured each hour for four days. A total of 32 diameter measurements were made. The data are given in the file *precision shaft diameter feedback*, with the output being the diameter measured from nominal. The results show a gradual drift. The team speculated that the drift was caused by changes in some (unidentified) environmental conditions. They decided to reduce the measurement variation using a feedback controller.

a. What type of feedback controller (that is, what prediction equation and what adjustment rule) would you recommend in this application?
b. Suppose the team decided to use a feedback controller based on EWMA forecasts with the smoothing parameter alpha equal to 0.4. What kind of a reduction in the measurement variation could they expect?

CHAPTER 19

19.1 In the paint film build example introduced in Chapter 3, the baseline standard deviation in film build (paint thickness) was 0.67 thousandths of an inch. With this variation, to ensure a minimum film build of 15-thousandths of an inch, the process was centered at 17. The goal was to reduce the standard deviation to 0.35, thereby allowing for a reduction in the average film build.

The dominant cause of film build variation was found using a multivari investigation to act in the car-to-car family. Despite further effort, the dominant cause was not found. The team decided to adopt the process robustness approach. Based on process experience, candidates and their corresponding levels were chosen as follows:

Candidate	Low level	High level
Anode dimension	3.1	3.9
Conductivity of paint	Low	High
Temperature	30	50
Zone X voltage	450	475
Zone Y voltage	500	525

The team selected a fractional factorial resolution V experiment with the 16 treatments given as follows. To reduce the cost of the experiment, panels were used rather than cars. With this choice there was a risk of study error.

Treatment	Anode dimension	Conductivity of paint	Temperature	X voltage	Z voltage
1	3.1	Low	30	450	500
2	3.9	Low	30	450	525
3	3.1	High	30	450	525
4	3.9	High	30	450	500
5	3.1	Low	50	450	525
6	3.9	Low	50	450	500
7	3.1	High	50	450	500
8	3.9	High	50	450	525
9	3.1	Low	30	475	525
10	3.9	Low	30	475	500
11	3.1	High	30	475	500
12	3.9	High	30	475	525
13	3.1	Low	50	475	500
14	3.9	Low	50	475	525
15	3.1	High	50	475	525
16	3.9	High	50	475	500

For each run, five panels were painted. The order of the treatments was randomized. Since the dominant cause acted car to car, the team believed the unknown dominant cause would act within each run. Film build was measured at five locations on each panel. The data for one location are given in the file *paint film build robustness* and in the table that follows.

Treatment	Order	Film build	Average	Log(s)
1	14	15.6, 15.3, 15.9, 15.2, 15.8	15.56	–0.51
2	5	16.0, 16.3, 17.3, 16.2, 16.6	16.47	–0.31
3	6	15.0, 14.8, 14.9, 15.3, 16.1	15.22	–0.28
4	2	16.1, 17.6, 17.2, 16.3, 16.1	16.69	–0.16
5	9	15.7, 15.6, 15.2, 15.2, 15.7	15.49	–0.57
6	12	17.3, 17.6, 16.8, 17.5, 17.3	17.28	–0.49
7	13	16.2, 14.4, 15.4, 14.5, 15.9	15.30	–0.09
8	4	17.3, 16.6, 16.6, 16.4, 17.8	16.94	–0.25
9	7	16.1, 14.7, 16.2, 14.7, 16.2	15.59	–0.09
10	16	17.2, 15.8, 16.4, 16.0, 15.8	16.23	–0.24
11	15	15.4, 15.2, 15.4, 15.3, 15.2	15.29	–1.06
12	1	16.6, 16.4, 16.4, 16.5, 16.4	16.48	–1.00
13	3	15.1, 15.4, 15.4, 15.0, 14.4	15.05	–0.41
14	10	16.8, 16.9, 17.0, 17.3, 16.3	16.89	–0.42
15	11	15.0, 15.1, 15.0, 14.9, 14.8	14.97	–0.86
16	8	16.6, 16.7, 16.3, 16.5, 16.3	16.48	–0.79

a. Analyze the data using the standard deviation of film build over the five consecutive panels to measure performance. Is it possible to make the process robust to noise variation? What levels of the candidates do you suggest?
b. The team had a way to adjust the process center. However, we can also use the robustness experiment to look for an adjuster. Analyze the data using the average film build over the five consecutive panels to measure performance. Are any of the candidates adjusters?
c. In the experiment, the film build at a particular location on five consecutive cars (panels) was used to define a run. Suppose, instead, that the five observations came from five fixed locations on a single door. What, if any, changes are needed in the analysis presented in part a?

19.2 In a trim plant, customer complaints about seat appearance prompted management to assign a team the task of reducing shirring variation. The team proceeded without using Statistical Engineering and made a number of mistakes. Seat cover shirring was scored on a scale of 1 to 6 using boundary samples by how much cloth was gathered by the parallel stitching. Shirring scores of 1 to 4 were acceptable with 1 being the best. Scores of 5 or 6 resulted from either too much or too little shirring. A review of historical data suggested that the observed shirring score over a week covered all six possible values. Next, the team informally checked the measurement system. They found the measurement system added little variation. The team decided not to look for a dominant cause. Rather they moved directly to assessing the feasibility of making the process robust. They used brainstorming to select six candidates with two levels each as follows:

Candidate	Low level	High level
Leather thickness	0.8	1.2
Leather toughness	Pliable (soft)	Stiff (tough)
Seam width	9 mm	11 mm
Material feed	Top up	Bottom up
Steam to skin bun	Used	Not used
Bun thickness	+5 mm	–5 mm

The team planned a resolution III fractional factorial experiment with 16 runs (one for each treatment) as follows:

Treatment	Leather thickness	Seam width	Leather toughness	Machine feed	Steam	Bun thickness
1	High	Low	Tough	Top up	Yes	High
2	High	Low	Soft	Bottom up	No	Low
3	Low	High	Tough	Top up	No	Low
4	Low	High	Soft	Bottom up	Yes	High
5	High	High	Tough	Bottom up	Yes	Low
6	High	High	Soft	Top up	No	High
7	Low	Low	Tough	Bottom up	No	High
8	Low	Low	Soft	Top up	Yes	Low
9	High	High	Tough	Bottom up	No	Low
10	High	High	Soft	Top up	Yes	High
11	Low	Low	Tough	Bottom up	Yes	High
12	Low	Low	Soft	Top up	No	Low
13	High	Low	Tough	Top up	No	High
14	High	Low	Soft	Bottom up	Yes	Low
15	Low	High	Tough	Top up	Yes	Low
16	Low	High	Soft	Bottom up	No	High

Each run consisted of three seats (repeats). The runs were conducted in the treatment order given in the table. The data are given in the file *seat cover shirring robustness* and reproduced as follows:

Treatment	Order	Seat 1	Seat 2	Seat 3	Average score
1	13	3	1	2	2.0
2	16	1	2	1	1.3
3	7	2	2	2	2.0
4	6	2	2	2	2.0
5	10	2	1	1	1.3
6	1	3	1	3	2.3
7	11	4	2	1	2.3
8	15	2	2	4	2.7
9	5	1	2	2	1.7
10	3	4	5	2	3.7
11	14	3	3	2	2.7
12	9	2	3	3	2.7
13	8	1	2	2	1.7
14	2	2	2	3	2.3
15	4	1	4	2	2.3
16	12	2	3	1	2.0

a. Explain why choosing the process output as a measure of variation (that is, high scores come from either too much or too little shirring) was a poor one.
b. The goal is to find process settings that lower the average shirring score. Can we use any of the candidates to achieve the goal?
c. Each run consisted of three seats. Discuss this choice in the context of a robustness experiment.

For the last two parts of this question, suppose the first three candidates (leather thickness, leather toughness, and seam width) used in the robustness experiment were normally varying rather than fixed inputs.

d. How should the levels of the first three inputs have been chosen?
e. Discuss changes you would make to the analysis you conducted in part b.

19.3 Torsional rigidity of the weather stripping was the dominant cause of door assembly problems. Management set a goal of reducing standard deviation in torsional rigidity to 0.3. A baseline investigation found the variation in torsional rigidity was roughly 0.55 mm and that the dominant cause acted over the short term and certainly within any half hour. The team looked briefly for a dominant cause of rigidity variation without success. Next, they planned a robustness experiment with four candidates at two levels each, chosen based on engineering judgment. The candidates and levels are:

Candidate	Low level (–1)	High level (+1)
Heat (pre)	100	700
Extruder RPM	22	26
Tension (pre)	1	5
Water flow	2	6

The team planned a full factorial experiment with 16 runs, one for each treatment. The correspondence between treatments and candidate levels is given in the table that follows.

Treatment	Heat	Extruder RPM	Tension	Water flow
1	–1	–1	–1	–1
2	–1	–1	–1	1
3	–1	–1	1	–1
4	–1	–1	1	1
5	–1	1	–1	–1
6	–1	1	–1	1
7	–1	1	1	–1
8	–1	1	1	1
9	1	–1	–1	–1
10	1	–1	–1	1
11	1	–1	1	–1
12	1	–1	1	1
13	1	1	–1	–1
14	1	1	–1	1
15	1	1	1	–1
16	1	1	1	1

Each run consisted of running the process for half an hour after the candidate levels had been reached. Within each run, 10 weather-strip samples were selected spread out over the half hour. The order of the runs was randomized. The torsion rigidity of each of the 10 weather-strip samples for each treatment is given in columns s1 to s10 of the table that follows and in the file *weatherstrip torsional rigidity robustness*.

Treatment	Order	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
1	13	10.3	13.0	11.5	11.8	10.7	9.9	10.7	11.5	11.0	11.1
2	6	11.5	13.0	10.4	11.1	10.9	10.6	12.0	9.3	9.2	9.3
3	9	11.6	13.0	10.4	16.0	10.3	10.8	11.5	11.0	11.3	10.9
4	1	11.5	11.7	10.4	11.7	14.0	11.7	10.4	11.7	10.4	10.4
5	3	14.0	11.7	11.7	19.0	11.9	11.7	12.1	13.0	11.1	11.0
6	11	22.0	15.0	18.3	11.7	20.3	21.0	12.6	13.6	14.7	15.1
7	5	9.1	9.6	10.2	9.8	9.0	9.7	10.0	12.0	9.0	8.8
8	14	10.0	9.1	10.6	10.4	10.8	11.0	11.1	10.8	10.5	10.8
9	2	11.7	12.5	11.9	11.7	20.0	14.0	10.4	11.5	11.7	20.0
10	10	10.3	11.6	10.5	10.6	13.0	14.0	11.7	10.3	15.0	11.8
11	7	10.3	10.5	11.0	11.4	9.8	10.4	11.7	11.8	11.5	11.9
12	15	11.6	11.0	11.4	11.3	12.0	10.6	10.9	10.7	10.7	10.7
13	16	10.6	10.7	11.6	10.6	10.7	22.0	11.0	10.4	10.4	23.0
14	8	9.1	10.4	10.6	11.4	10.9	10.4	10.8	10.9	11.0	11.6
15	12	10.3	11.0	12.0	12.1	10.5	10.7	11.3	11.4	10.8	10.9
16	4	10.4	10.4	10.4	10.5	10.9	11.4	9.0	9.6	9.8	10.2

a. To analyze the results of this robustness experiment, what performance measure(s) do you recommend and why?

b. Analyze the experimental results using your chosen performance measure(s). What can you conclude?

CHAPTER 20—NO EXERCISES

CHAPTER 21

21.1 Discuss whether lessons learned are properly maintained in corporate memory in your organization. What could be done to improve the situation?

21.2 In the paint film build example described in Chapter 19, the team found new process settings that resulted in reduced car-to-car variation in film build. To validate the proposed solution, 80 cars were painted over one day with the settings given in the following table. These were the best settings found in the robustness investigation. The film build values from five specific positions on one door for each of the cars are available in the file *paint film build validation*.

Candidate	Setting
Anode dimension	3.5 (midpoint)
Conductivity of paint	High
Temperature	30
Zone X voltage	475
Zone Y voltage	500

a. The baseline film build standard deviation was 0.68. The problem goal was to reduce the standard deviation to 0.35, and the robustness experiment results suggested that changing settings would reduce the standard deviation to about 0.37. Has the solution been validated?

b. What, if anything, do the validation results tell us about the home of the dominant cause in the remaining variation?

21.3 In the truck pull example described in Chapter 17 and Exercise 17.1, a feedforward controller was implemented to compensate for the effect of truck-frame geometry on pull. After the feedforward system had been operating successfully for some time, management decided to review its operation. The four frame geometry measurements and left and right caster and camber were recorded for roughly a month of production consisting of over 6600 trucks. The data are given in the file *truck pull validation*.

a. The standard deviations for caster and camber before implementation of the feedforward controller can be estimated from the 100-truck investigation described in Chapter 17. From the same investigation, the team predicted the possible reduction in standard deviation using a feedforward controller. A summary is given in the following table.

Characteristic	Baseline standard deviation	Predicted reduction in standard deviation
Left caster	0.90	0.18
Right caster	0.83	0.20
Left camber	0.51	0.13
Right camber	0.41	0.10

Do the results of the investigation validate the reduction in left and right caster variation due to the feedforward controller?

b. For each of the two caster characteristics, conduct a regression analysis to see if the feedforward controller can be improved. Recall that the feedforward controller should be compensating for variation in the frame geometry.

c. Repeat the analysis in parts a and b for left and right camber.

Exercise Solutions

CHAPTER 1—NO EXERCISES

CHAPTER 2

2.1 The word *variation* is used in other contexts to describe a difference between a realized and target value such as in budget variation. How does this use compare to variation as discussed in Chapter 2?

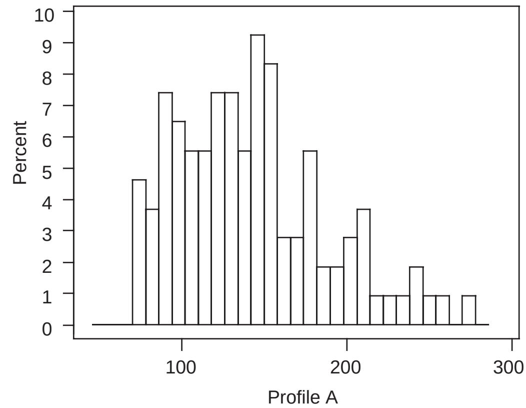
We define two types of *variation,* an off-target component and a part-to-part component. *Budget variation* only describes the off-target component of variation.

2.2 We have heard the following comment many times from manufacturing engineers: "The cause of the variation is the product design—what can you expect me to do?" Discuss the comment in light of the definition of cause in Section 2.2.

Product design cannot be the cause of part-to-part variation because it does not change from part to part. It is, however, possible that changes to the product design may solve the problem, though in most cases, the process produces good parts. This suggests improvement is possible with the existing design.

2.3 Profile A is a measure of deviation of the actual from the ideal shape of a camshaft lobe over one region (A) of the lobe. The target value is zero and the upper specification limit is 250 microns. Use the data in the file *camshaft lobe runout baseline* to summarize the variation in this output. Do all lobes exhibit the same variation? Is there any time pattern in the variation?

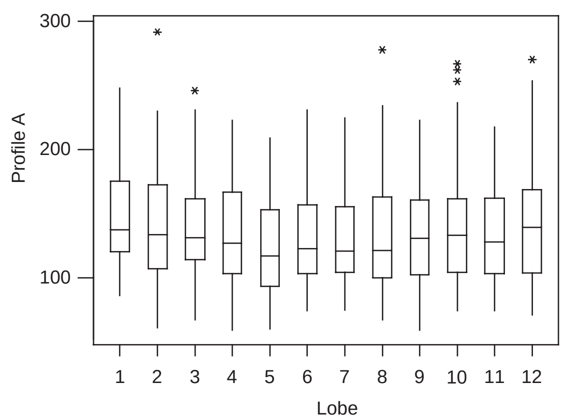
We use a combination of plots and numerical summaries. Some typical plots follow. The histogram (and the subsequent numerical summary) suggests the full extent of variation is 59 to 292.



Variable	N	Mean	Median	TrMean	StDev	SE Mean
profile A	1296	136.32	130.00	134.54	39.83	1.11

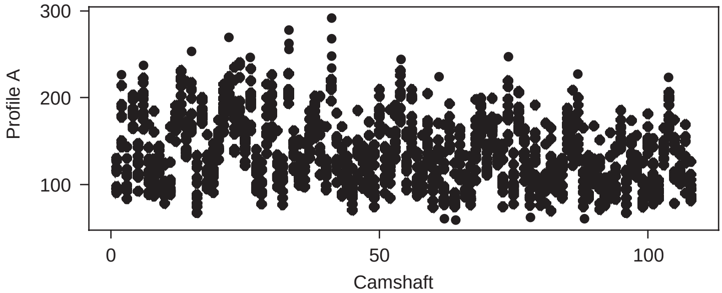
Variable	Minimum	Maximum	Q1	Q3
profile A	59.00	292.00	104.00	164.00

Stratifying by lobe we see the average and variation in profile A is roughly the same across all 12 lobes.

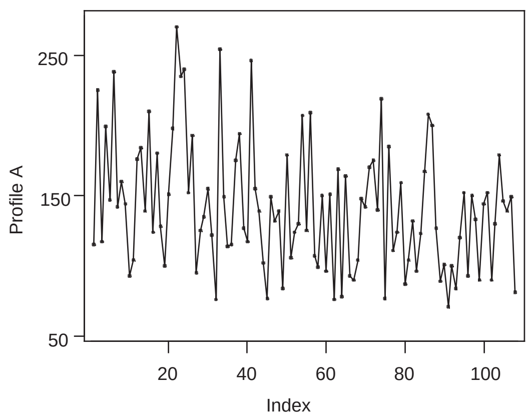


We need to be careful looking at the data by time. As stored, the 1296 profile A values come from 108 camshafts. The camshafts are given in production sequence.

We can plot profile A for each of the 12 lobes by camshaft in order of production. The plot shows the 12 values for each camshaft. There are no strong time effects that influence all lobes in the same way.

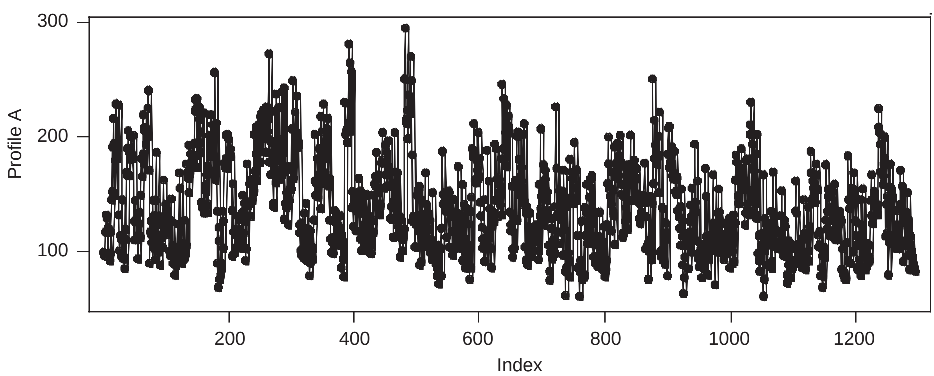


To determine if there is a time effect on individual lobes, we can look at run charts of profile A values for individual lobes. For example, for lobe 12 we get:



There are no time patterns evident for the different lobes.

A common mistake is to create the run chart of all 1296 profile A values as follows:



The observations are not ordered by time because each group of 12 values comes from the 12 lobes on a single camshaft. In this plot, large lobe-to-lobe effects could be misinterpreted.

2.4 Construct histograms and run charts for output 1 and output 2 given in the data file *chapter 2 exercise 4*. Find the average and standard deviation for each output. Assume the target value and upper specification limit for these lower-is-better outputs are 0 and 35.

a. Is the variation the same for each output?

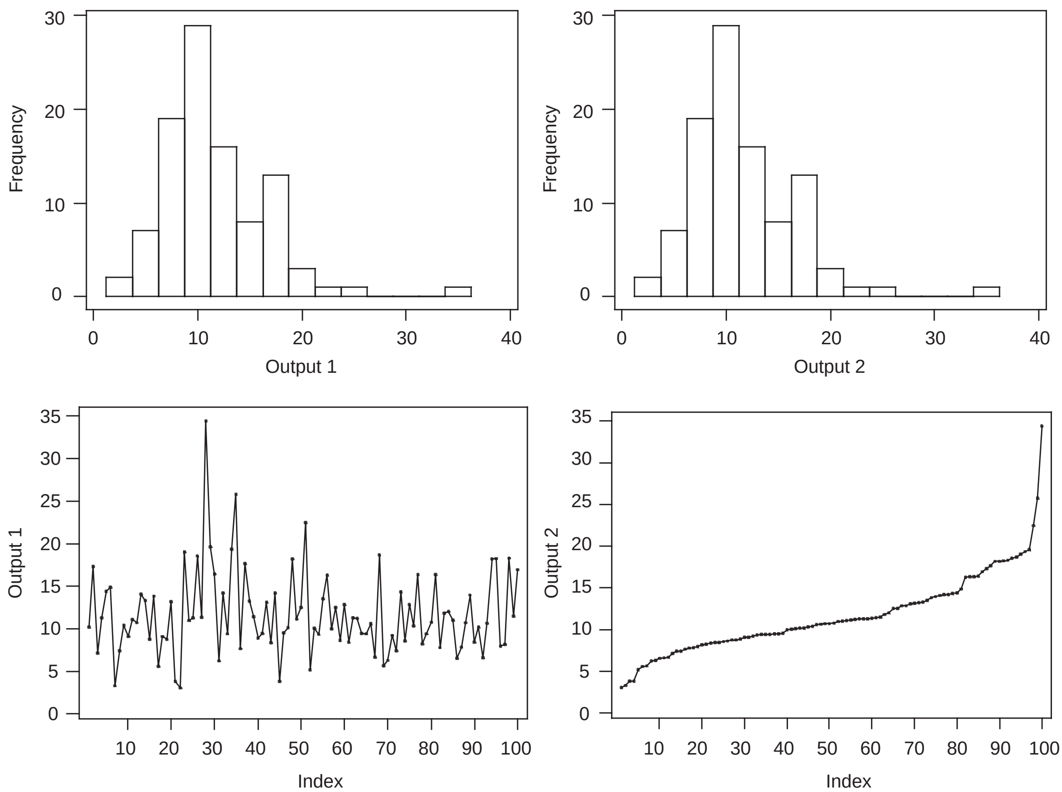
b. Is the nature of the variation over time the same for each output?

The results are:

Descriptive Statistics: output1, output2

Variable	N	Mean	Median	TrMean	StDev	SE Mean
output1	100	11.634	10.713	11.364	4.882	0.488
output2	100	11.634	10.713	11.364	4.882	0.488

Variable	Minimum	Maximum	Q1	Q3
output1	3.017	34.488	8.539	14.041
output2	3.017	34.488	8.539	14.041



a. The standard deviations and histograms are the same for both outputs.

b. Output 2 is ordered from smallest to largest over time.

2.5 You may convince yourself that the formulas for combining means and standard deviations given in Section 2.4 are true with the following numerical demonstration you can conduct in MINITAB. Generate two columns of 100 values sampled from some model (in MINITAB: Calc Æ Random Data Æ your choice of model and parameters, for example, Normal with mean and standard deviation 0 and 1, respectively). Then, calculate two new data columns. Let one column be the sum of the original two columns and the other the difference.

a. Find the standard deviation and average for each of the four columns.
b. Calculate the sum and differences of the averages for the first two columns. How do these compare to the average of the other two columns respectively?
c. Calculate the standard deviation for the sum and difference using the "square root of sum of squares" formula given by 2.1. How do the results compare to the standard deviations for the last two columns?

a. An example of the results you will get is:

```
Descriptive Statistics: C1, C2, sum, diff

Variable          N       Mean     Median     TrMean      StDev    SE Mean
C1              100      0.061      0.016      0.055      1.126      0.113
C2              100    -0.0976    -0.0853    -0.0904     0.9145     0.0915
sum             100     -0.037     -0.226     -0.048      1.571      0.157
diff            100      0.158      0.222      0.167      1.320      0.132

Variable    Minimum    Maximum         Q1         Q3
C1           -3.214      2.801     -0.787      0.802
C2          -2.3400     2.3015    -0.6814     0.5067
sum          -3.482      4.000     -1.181      1.221
diff         -3.134      3.308     -0.746      1.115
```

b. The sum and difference of the averages for the first two columns equal the average of the sum and difference columns, respectively.
c. For the sum (and difference) of C1 and C2 we have

$$stdev(\mathrm{C1} \pm \mathrm{C2}) = \sqrt{stdev(\mathrm{C1})^2 + stdev(\mathrm{C2})^2}$$

In the example data $\sqrt{1.126^2 + .9145^2} = 1.45$, which closely matches the standard deviation of the last two columns.

2.6 At a project review, the team presented the following summary of their investigation based on standard deviations.

Source of variation	Percent of total
Measurement system	30
Identified cause	50
Unidentified causes	81

a. The reviewing manager questioned the numbers in the second column of the table because they did not add to 100. Is there an error? Explain.
b. By what percentage can the process standard deviation be reduced by eliminating the contribution of the identified cause?
c. Is the identified cause a dominant cause?

a. There is no error. The percentages are calculated on the standard deviation scale. Recall that standard deviations combine using the square root formula as illustrated by Equation (2.1). The total is given by $\sqrt{0.30^2 + 0.50^2 + 0.81^2} = 1$.
b. If we eliminate the contribution of the identified cause the remaining variation is given by $\sqrt{0.30^2 + 0.81^2} = 0.85$. So the process standard deviation could be reduced by roughly 15%.
c. No.

2.7 In Chapter 1, we discussed a project to reduce variation in pull, an alignment characteristic of light trucks. Recall that

Pull = 0.23*(right caster – left caster) + 0.13*(right camber – left camber)

and that the data for two months' production are stored in the file *truck pull baseline*. The data are summarized in the following table.

Output	Average	Standard deviation
Left camber	0.257	0.129
Right camber	0.249	0.130
Left caster	3.519	0.224
Right caster	4.519	0.243
Pull	0.231	0.082

a. Use the formula for pull and the results for how averages and standard deviations combine to predict the average and standard deviation for pull given by the last row in the table indirectly from the component averages and standard deviations.
b. Suppose you had the resources to reduce the variation in one of the alignment angles by 50%. Which angle would you choose? By how much, approximately, would the pull standard deviation be reduced?

a. The derived standard deviation for pull is

$$\sqrt{0.23^2 * \left(0.224^2 + 0.243^2\right) + 0.13^2 * \left(0.13^2 + 0.129^2\right)} = 0.08$$

The derived standard deviation for pull is not exactly 0.082, because the alignment angles do not vary independently; there is a small correlation. Note that the averages of the components play no role in the standard deviation.

The derived average is:

$$0.23(4.519 - 3.519) + 0.13(0.249 - 0.257) = 0.229$$

There is some rounding error, so this value does not match the pull average in the table.

b. The largest reduction in overall standard deviation would be achieved by reducing the variation in the right caster. Reducing the variation in right caster by 50% would reduce the variation in pull to roughly

$$\sqrt{0.23^2 * \left(0.224^2 + \left(\frac{0.243}{2}\right)^2\right) + 0.13^2 * \left(0.13^2 + 0.129^2\right)} = 0.063$$

This corresponds to approximately a 20% reduction.

CHAPTER 3

3.1 For a problem of interest to you speculate about the likely costs and feasibility of implementing each of the possible variation reduction approaches.

The solution is dependent on the chosen problem.

3.2 Variation in the location of a drilled hole in a machined casting can cause poor fits when the part is bolted to an engine. To reduce this variation, an engineer considers a variety of possible approaches.

a. A vision system is available that can measure location on 100% of the parts and reject those that it judges to be out of specification. What are the advantages and disadvantages of such an approach?

b. Institute a feedback controller by measuring two parts every hour. If hole location on either part is outside of specification, stop and adjust the process. When is such a scheme likely to be effective?

c. A third choice is to find a dominant cause of the variation. What are the advantages and disadvantages of this strategy?

d. If a dominant cause can be discovered, what options does the engineer have?

a. 100% inspection would ensure that no out-of-specification parts were shipped to the customer, assuming there were no measurement errors. The inspection system may be expensive to install and run. Also, we need to specify how to handle rejected parts.

b. For feedback control to be effective, the short-term variation in hole location must be substantially smaller than the hour-to-hour variation. There must also be a way to adjust the center of the process.

c. Finding a dominant cause of hole location variation would be valuable information that may lead directly to a low-cost solution. However, finding a dominant cause may be difficult or expensive. Also, the dominant cause may be outside the control of local management or difficult to control.

d. With knowledge of a dominant cause, the engineer can consider the variation reduction approaches that require knowledge of a dominant cause, namely: fixing the obvious, desensitizing the process, and feedforward control. Any of the non-cause-based approaches are also still options.

CHAPTER 4—NO EXERCISES

CHAPTER 5

5.1 Briefly discuss the advantages and disadvantages of the following—be sure to think of potential errors as described within the QPDAC framework.

a. To estimate the baseline performance of a grinding process, 100 consecutive pistons were sampled and the diameters were measured.

b. To investigate a proposed change to a chemical process, the investigators tried the change in a pilot process rather than the production process.

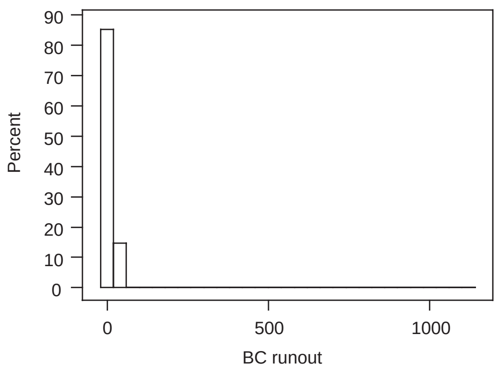
a. The given plan results in quick and easy data collection. However, there is a danger that process variation will be underestimated since 100 consecutive pistons may be more similar than 100 pistons chosen over a longer time frame (study error). Also, 100 pistons is a relatively small sample. See Table S6.1 in the supplement to Chapter 6 to get a better idea of the uncertainty in the estimate of the process standard deviation from a small sample size (sample error).

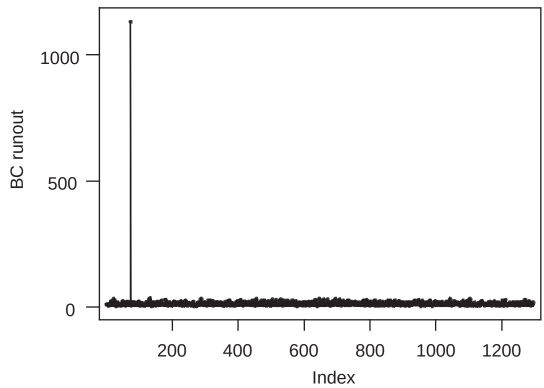
b. An investigation on the pilot process would be cheaper and easier than using the production process. The main concern is whether results from the pilot process can be scaled up to the regular process (study error.)

5.2 In the camshaft lobe BC runout problem described in Chapter 1, the team selected 50 parts (10 per day over 5 days) and measured the BC runout for each of the 12 lobes on each camshaft to quantify the baseline. The 600 runout measurements are stored in the file *camshaft lobe runout baseline original*. Conduct an analysis of these data. Are your conclusions different from those in Chapter 1? Why?

From the MINITAB results that follow, we see that the variation in BC runout as measured by the standard deviation is now much larger at 31.7. However, looking at the numerical or graphical summary, we see that the data contain a large outlier. In particular, observation number 74 is 1130. A transcription error put the decimal point in the wrong place. The value should have been 11.3. This mistake was readily identified and corrected before proceeding with the rest of the analysis.

```
Descriptive Statistics: BC runout

Variable            N      Mean    Median    TrMean     StDev   SE Mean
BC runout        1296     13.51     11.10     12.28     31.69      0.88

Variable      Minimum   Maximum        Q1        Q3
BC runout        2.60   1130.00      7.43     17.00
```





5.3 To assess a measurement system used to check the diameter of an engine bore, an investigator plans to repeatedly measure the same four (of the eight) bores on five blocks sampled from a shift of production.

a. Discuss the advantages and disadvantages of using 10 rather than 5 blocks.

b. In the investigation, all the blocks produced over one shift were available for study. Give two considerations that the investigators should take into account in making the choice of available blocks.

c. The plan was to make all measurements in a single day. Discuss the advantages and disadvantages of making the measurements over a longer time period.

d. When would the investigator be better off devoting the available resources to measuring all eight bores on fewer engine blocks?

a. Using 5 blocks rather than 10 would be cheaper and quicker. However, with fewer blocks there is a greater chance of sample error. The performance of the measurement system on the chosen 5 blocks may be different (better or worse) than its performance on other blocks.

b. The investigator needs to trade off cost and convenience with the risk of study error. The key question is whether the performance of the measurement system is likely to be different when examining blocks from other days. It may be, for example, that engine bores from a single day have similar dimensions and that the measurement system works well only for some dimensions.

c. Again the investigator needs to trade off cost and convenience with the risk of study error. Here the key question is whether the properties of the measurement system change substantially over time. Measurement systems, like other processes, change over time unless properly maintained. As we shall see in Chapter 7, we recommend, if at all possible, that measurement investigations compare the measurements over a longer time period.

d. Having fewer engines makes the logistics of the investigation easier. In the measurement investigation to repeatedly measure the bores on the same block, the team had to move the block back into the process before the gage. Measuring all the bores on fewer blocks would be a good idea if we expected the measurement system to be sensitive to the bore number (though if this were known, another option would be to focus on the known problem bores). If, on the other hand, the dimensions of all the bores on a particular block were similar, using fewer blocks might prevent us from seeing the full range of bore dimensions in the measurement investigations.

5.4 You are a manager with the responsibility to decide if you should change the supplier for a tooling insert. You receive a report from your process engineer who has conducted an investigation into a new insert. He gives you the following verbal report and recommendation:

> **Our current insert has an average life of 1105 parts. To assess the performance of the new supplier, we asked them to supply 10 inserts. We checked the inserts out on one of our machines last week and got an average of 1300 pieces. Since the cost is the same, I think we should switch to the new inserts.**

Using the QPDAC framework, think of five questions you would ask about the conduct of the investigation before you might accept the recommendation.

You should ask questions like:

Is the machine used for the investigation typical of all machines we use?

Did the manufacturer supply 10 specially good inserts? How can we tell?

How much variation was there in the number of pieces cut by the 10 new inserts?

What data were used to estimate the average life of the current inserts?

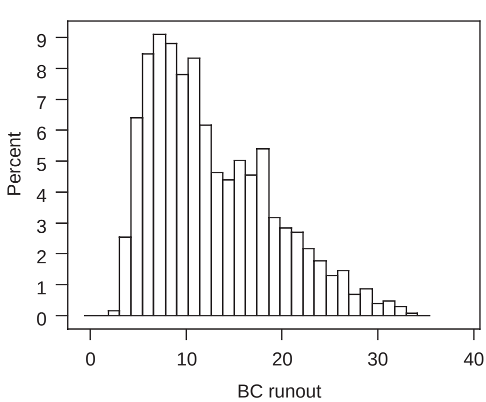
Are the environmental and production conditions in the week of the investigation of the new inserts typical?

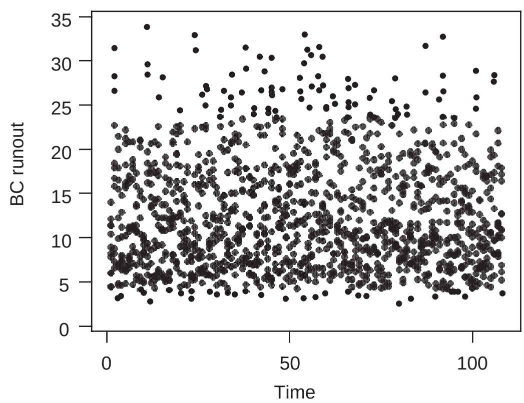
CHAPTER 6

6.1 In Chapter 1, we described a problem in terms of the lobe geometry of camshafts. The data are given in the file *camshaft lobe runout baseline*. Quantify the problem baseline for the following output.

a. BC runout
b. Angle error

a. We summarize the data using a histogram and a run chart as follows. The run chart plots the 12 lobe BC runout values for the 108 camshafts in production order.



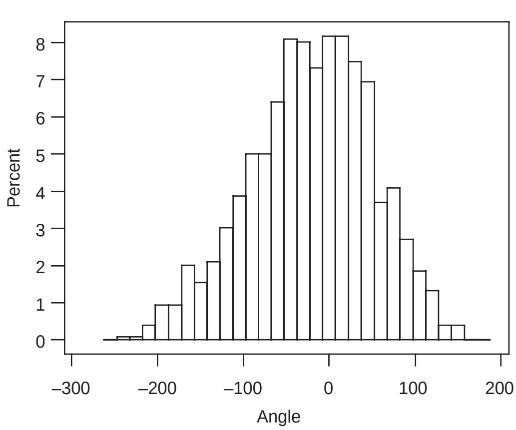


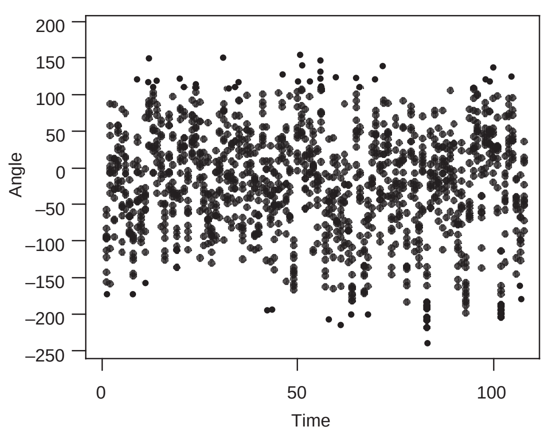
Variable	N	Mean	Median	TrMean	StDev	SE Mean
BC runout	1296	12.643	11.100	12.271	6.389	0.177

Variable	Minimum	Maximum	Q1	Q3
BC runout	2.600	33.900	7.425	17.000

The baseline standard deviation is 6.4, and the full extent of variation is 2.6 to 33.9 microns. We see no obvious patterns over time. We can also look at run charts for each lobe separately.

b. Looking at angle error we use the same summaries as in part a.





Variable	N	Mean	Median	TrMean	StDev	SE Mean
angle	1296	-21.30	-18.00	-20.04	71.50	1.99

Variable	Minimum	Maximum	Q1	Q3
angle	-241.00	155.00	-67.00	30.00

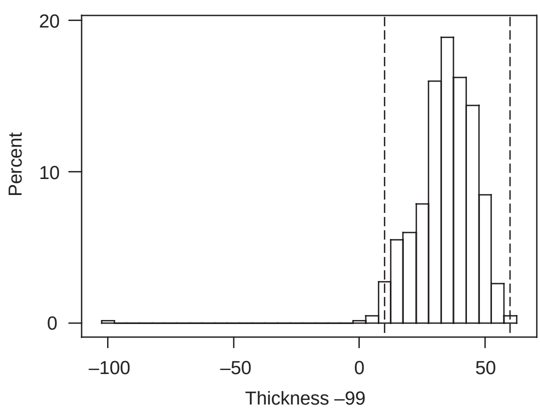
The baseline standard deviation is 71.5 and the full extent of variation is roughly –240 to 155. There is a weak pattern over time.

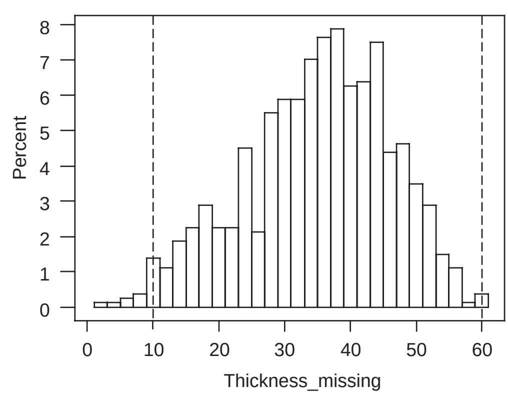
6.2 Many programs such as Excel cannot easily handle missing observations. MINITAB is an exception. Missing values are often stored using a special numerical code (–99 is common). These special codes can result in much confusion and lead to incorrect conclusions. Consider the data *rod thickness baseline with missing observation*. In the file, there are two outputs. The output thickness_–99 uses a numerical code of –99 for missing observations, while thickness_missing uses the MINITAB missing observation symbol (*). Quantify the baseline for these two outputs. Which data summaries show the missing observation and which do not?

If we forget we are using a special code for a missing observation, it is not readily apparent in the numerical summary and the –99 inflates the standard deviation by about 9%. The code for the missing observation is easily identified in the graphical display if the code is extreme relative to the usual values. Using a missing observation symbol is much preferred over a special numerical code.

Variable	N	N*	Mean	Median	TrMean	StDev
Thickness_-99	800	0	34.426	36.000	34.832	11.981
Thickness_miss	799	1	34.593	36.000	34.861	11.018

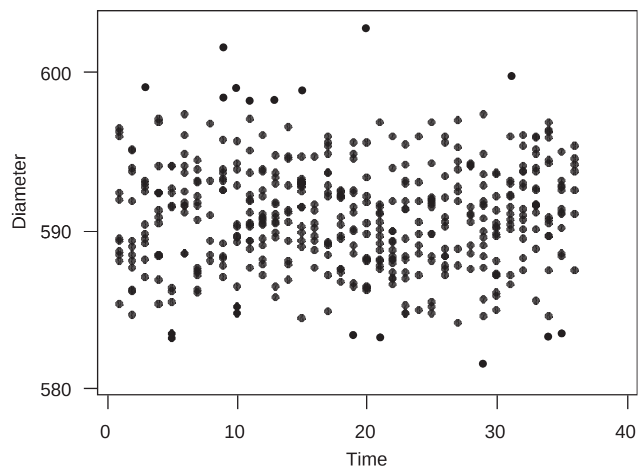
Variable	SE Mean	Minimum	Maximum	Q1	Q3
Thickness_-99	0.424	-99.000	59.000	28.000	43.000
Thickness_miss	0.390	2.000	59.000	28.000	43.000





6.3 The baseline investigation for the V6 piston diameter example was described in Chapter 5. The data are given in the file *V6 piston diameter baseline*. Suppose the data were collected so that all the pistons from a given hour were collected at the start of the hour. Now the data come in subgroups as defined by hour. What summaries used in the baseline analysis are affected by the subgrouping? When taking the subgrouping into account are the conclusions any different than those derived in Chapter 5?

The subgrouping in the data affects the interpretation of the run chart. The run chart (as given in Chapter 5) assumes the observations are equally spaced in time. Since the data are collected so that the five observations in each subgroup are much closer together in time then the observations from subsequent hours a better display would plot the diameters by hour (time).

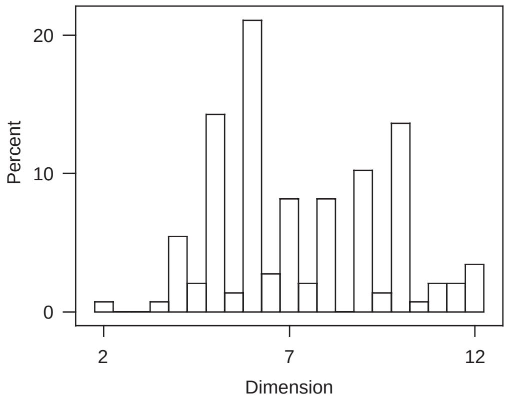


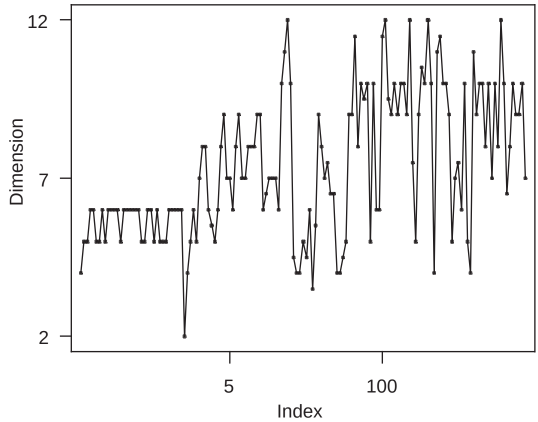
There appears to be no special pattern over time. The conclusions for the baseline would be unchanged.

6.4 Based on customer complaints concerning installation difficulties, a team investigated variation of a key fascia dimension. To establish a baseline, they measured the dimension on 147 fascias sampled from one month's production. The data are given in the file *fascia dimension baseline*. Using appropriate summaries of the data, quantify the baseline. Are there any concerns?

We summarize the baseline data numerically and using a histogram and run chart.

```
Variable       N      Mean    Median    TrMean     StDev   SE Mean
dimension    147     7.303     7.000     7.256     2.261     0.186

Variable     Minimum   Maximum        Q1        Q3
dimension      2.000    12.000     6.000     9.000
```





There is some evidence of an increase in the fascia dimension over the month of the baseline investigation. This suggests the investigation was not conducted over a long enough period of time. We need to plan a new baseline investigation.

CHAPTER 7

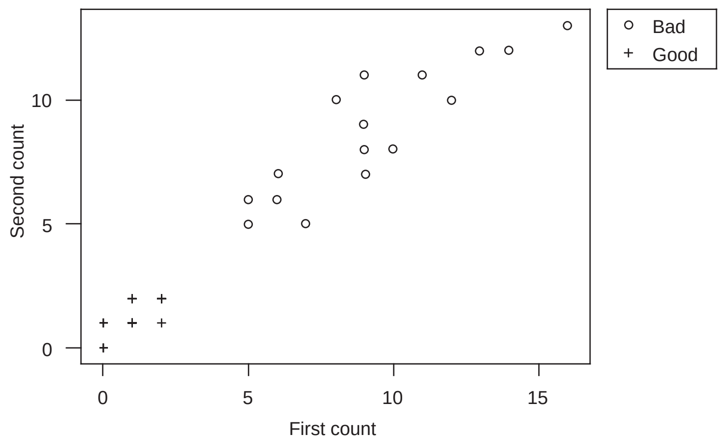
7.1 In a process improvement problem to improve the quality of a roof panel, the measurement system (specially designed for the project) counted the number of updings on each panel. To assess the measurement system, the number of updings on 20 bad panels and 20 good panels was counted twice. The data are given in *roof panel updings measurement*.

a. Can this investigation be used to assess the measurement variation of the counting process? Explain.

b. Can this investigation be used to assess the bias of the counting process? Explain.

c. The same operator counted all panels. Does the order in which he makes the counts matter? It is most convenient to count the same panel twice in a row. Is this a good idea?

d. A scatter plot of the first versus the second measurement is given as follows. Note that some plotting symbols correspond to more than one pair of measurements? What does the scatter plot tell you about the counting process?



e. This investigation was conducted over one hour. What are the advantages and disadvantages of spreading the two measurements on each panel over two days?

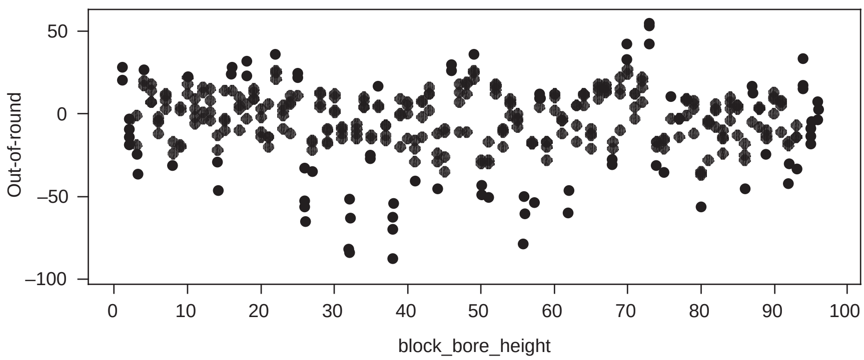
f. Can the counting process discriminate between good and bad panels?

a. Yes. By measuring each panel twice we can assess measurement variation. However, we may underestimate the variation if important causes of measurement variation do not vary sufficiently during the investigation.
b. No. We cannot assess bias since we do not know the true upding count for each panel.
c. It is best to randomize the order of the panels. The danger with measuring the same panel twice in row is that on the second count the operator will remember and be influenced by the first count.
d. The points lie close to the 45° line and there is a clear separation between good and bad panels. This suggests the measurement system is adequate.
e. Using two days for the investigation reduces the risk of study error but takes longer.
f. Yes. See part d.

7.2 To monitor the process that produces engine blocks, piston bore diameters are measured on every block because they are key characteristics. Each engine block has eight bores. The bore diameter is measured at three different heights in each bore (bottom, middle, and top) and at two different orientations at each height. Because the measurement process is automated, there are no operators. A measurement investigation was conducted over a day where the diameter of every bore on four blocks was measured four times each. The main concern was out-of-round, given by 10,000 times the difference of the two diameters at a particular height. The data are given in the file *block bore diameter measurement*. From a baseline investigation the out-of-round standard deviation was 22.8.

a. Determine the discrimination ratio. Is the measurement system adequate?
b. What would have been the advantage and disadvantage of conducting the measurement investigation over a longer time period?

a. Out-of-round is measured four times at each of 96 locations (4 blocks by 8 bores by 3 heights). To analyze the data, we define a new characteristic (called "block_bore_height") to uniquely identify each of the 96 locations. Plotting out-of-round by this new characteristic we get



The measurement variation is roughly comparable across the different blocks, bores, and heights. To estimate the measurement variation from this data we use a one-way ANOVA analysis; part of the results follow.

Analysis of Variance for out-of-round

Source	DF	SS	MS	F	P
block_bore_height	95	177032.7	1863.5	24.94	0.000
Error	288	21516.3	74.7		
Total	383	198549.0			

Pooled StDev = 8.643

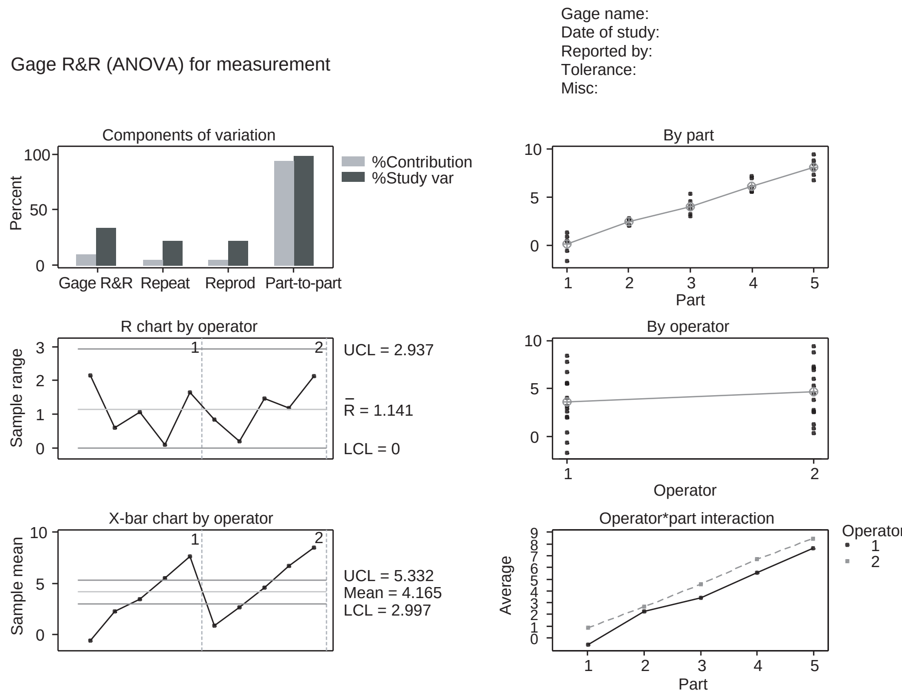
The variation due to the measurement system is estimated as 8.6. We cannot assess the stability of the measurement system here. The order of measurements was not preserved and, in any case, all the measurements were conducted over a single day.

Using the baseline standard deviation we can solve for an estimate of the standard deviation attributable to the process. We get 21.1 ($=\sqrt{22.8^2 - 8.643^2}$). The measurement system discrimination ratio is thus 2.44, and we conclude that the measurement system is not the dominant cause but that the measurement system should be improved before proceeding with the Statistical Engineering algorithm.

b. Spreading the measurements over a longer time period would have delayed reaching a conclusion about the adequacy of the measurement system. An advantage is that the longer time would have allowed slowly changing causes to act. From the baseline investigation, the team knew that the dominant cause acted within a day. Thus, a single day was long enough for the measurement investigation.

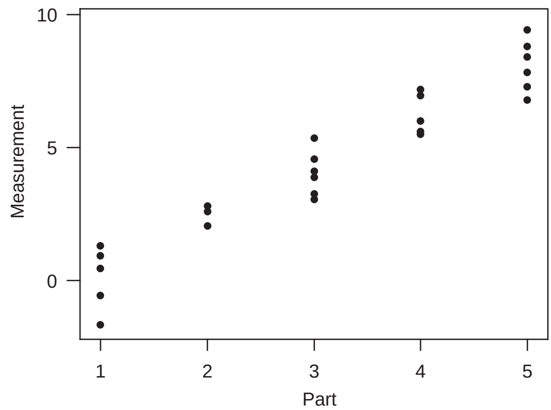
7.3 The following MINITAB results and graphs arise from a measurement system investigation in which two different operators measured five parts three times each. The five parts were selected with initial measured values spread out over the full extent of variation, 0 to 8. The data are given in the file *chapter 7 exercise 3*. The two operators worked different shifts so the parts were saved from one shift to the next. The results include an edited ANOVA analysis as suggested in the supplement to Chapter 7 and the default gage R&R analysis in MINITAB.

```
Analysis of Variance for measurement
Source     DF       SS       MS       F       P
part        4  230.819   57.705   81.25   0.000
Error      25   17.754    0.710
Total      29  248.573

Pooled StDev =   0.8427

Gage R&R
                               %Contribution
Source              VarComp  (of VarComp)
  Total Gage R&R      0.900      8.62
  Repeatability       0.425      4.07
  Reproducibility     0.475      4.55
    operator          0.475      4.55
Part-To-Part          9.547     91.38
Total Variation      10.447    100.00

                    StDev    Study Var  %Study Var
Source              (SD)     (5.15*SD)  (%SV)
Total Gage R&R      0.94876   4.8861     29.35
  Repeatability     0.65207   3.3582     20.17
  Reproducibility   0.68917   3.5492     21.32
    operator        0.68917   3.5492     21.32
Part-To-Part        3.08975  15.9122     95.59
Total Variation     3.23214  16.6455    100.00

Number of Distinct Categories = 5
```



a. **What do the given results tell us about the bias and variation of the measurement system?**
b. **In the gage R&R results, the $\bar{X}$ chart by operator is out of control. What does this mean?**
c. **In the gage R&R results, why is the sum of the % study variation column not 100%?**
d. **What is the discrimination ratio (D) for this system? How does the part selection procedure influence this ratio?**
e. **The gage R&R is about 29%, yet *D* is small. Why?**
f. **The results suggest a small operator-to-operator difference. This observed difference may be due to a difference in method or a drift of the system over the two shifts. How can you separate these two possibilities?**

a. We estimate the standard deviation due to the measurement system as 0.84, the pooled standard deviation from the ANOVA. This calculation corresponds to the formula given in Section 7.2. We cannot assess measurement bias because the true dimensions are not known. As an aside, it is always a good idea to also plot the measured values versus part number as follows.



b. The points plotted on the $\overline{X}$ chart are the average measurement for a particular part and operator. The control limits are determined by the variation of the measurements within the 10 part and operator combinations. If the measurement system is able to distinguish among the parts, there should be many points outside the control limits. Here, out-of-control signals indicate a good measurement system.

c. The column partitions the percent variation due to the measurement system and the parts with the overall standard deviation as the divisor. To combine standard deviations we need to square and add. We see that $29.35^2 + 95.59^2 = 100^2$.

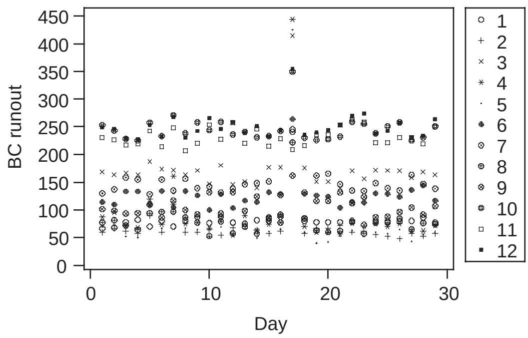
d. To calculate the discrimination ratio, we can estimate the overall standard deviation from the full extent of variation as 8 / 6 = 1.33 (we are assuming a bell-shaped histogram in this calculation). Thus we can estimate the standard deviation due to the process as 1.03 $(= \sqrt{1.33^2 - 0.84^2}\,)$. The corresponding estimate for the discrimination ratio *D* is 1.22 (= 1.03 / 0.84). The selection of the parts plays no role in the calculation of *D*. We assume that the variation within each part is the same regardless of true size.

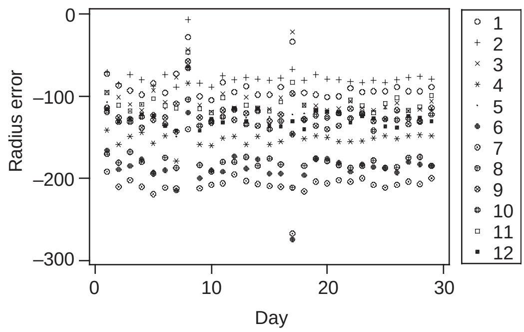
e. The denominator of the gage R&R calculation is the variation observed in the investigation. In this case, the overall standard deviation is 2.93, much larger than the estimated baseline variation, which decreases the R&R. The reason for the large overall variation is the part selection procedure.

f. With the plan, there is no way to distinguish the operator and shift effects (we say the effects are confounded). If we want to be able to separate the effects, we need another investigation in which both operators measure parts at the same time over several shifts.

7.4 To assess the variation in the system designed to measure camshaft lobe geometry over time, the same camshaft was measured daily for a month. At each measurement, a number of characteristics (for example, angle error, BC runout, taper, and so on) on each lobe were determined. The data are given in the file *camshaft lobe runout measurement stability*. Is there evidence of time-to-time variation in this measurement system?

Checking for time-to-time variation is difficult due to the large number of output characteristics and lobes. We plot each output versus day, stratifying by lobe number.





Examining a run chart of any characteristic versus the order in the MINITAB worksheet may be misleading since much of the observed pattern may be due to differences between lobes.

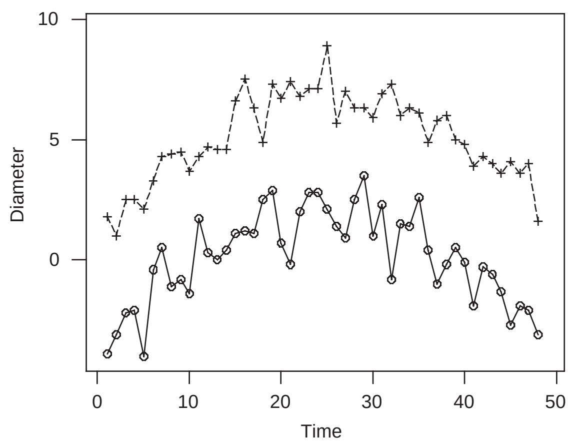
Most of the characteristics show a pattern similar to the one observed for BC runout, where the gage appears stable except for the results on day 17. Radius error has unusual results for day 8.

7.5 In a process that produced V8 pistons, problems occurred when pistons in inventory were remeasured (for an audit) and found to be out of specification. Since the process used 100% final inspection, this could only occur if there was a problem with the measurement system. This was puzzling because a recent gage R&R investigation at the final gage had concluded that the measurement system was acceptable. As a result, the team decided to conduct a long-term measurement investigation. Two pistons were chosen to span the range of diameter normally seen. Each piston was measured four times a day (spread out over the day) for 12 days. During that time the regular gage calibration was performed every four hours. The data are given in the file *V8 piston diameter measurement stability*.

a. Does the measurement system drift over time?

b. What effect does the regular gage calibration have?

a. To assess drift we plot the diameter for each piston over time.

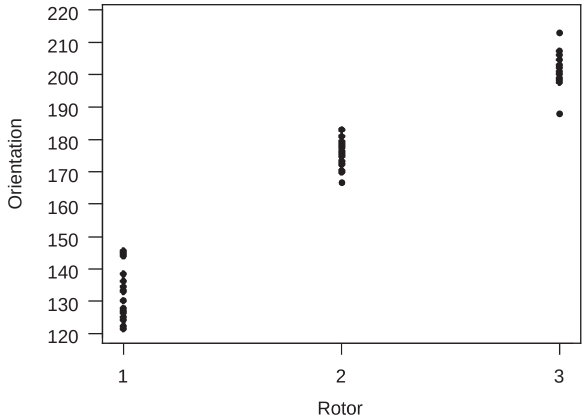


There is clear evidence of a drift over time. Note that since the pistons were chosen to span the normal range of diameters, the amount of drift seen is substantial.

b. The gage was recalibrated a number of times during the investigation. However, looking at the run chart, the regular calibration appears to have no effect.

7.6 Consider the brake rotor balance example described in the case studies. In the measurement investigation, three rotors were specially selected: one well balanced, another poorly balanced, and the final rotor requiring weight near the specification limit of 0.5. The three rotors were measured twice by each of the three gages on three separate days. There is no operator effect since the gages are automated. The 54 measurements are given in *brake rotor balance measurement*. The analysis given in the case study focuses on the measurement of the weight needed to balance the rotor. However, the location (or orientation) of the weight needed to move the rotor's center of gravity is also important. Can the measurement system consistently determine the orientation of the required balance weight? From the baseline investigation, the orientation of the weight was roughly uniform from 0° to 360°.

With an output like orientation measured in degrees we must be careful when analyzing the data. The data are circular since 0° is the same as 360°. Plotting the data for orientation by the three rotors we see that for this data we can ignore the circular nature of the output since none of the measured orientations are near 0°.



A one-way ANOVA analysis gives:

```
Analysis of Variance for orientation
Source     DF        SS        MS        F        P
rotor       2   43340.3   21670.1   589.70    0.000
Error      51    1874.1      36.7
Total      53   45214.4
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ---------+---------+---------+-------
1          18    132.57      8.11   (*)
2          18    175.47      4.08                    (*)
3          18    201.26      5.27                              (-*)
                                   ---------+---------+---------+-------
Pooled StDev =     6.06                   150       175       200
```

The measurement variation is estimated as roughly 6°, the pooled standard deviation. The team decided that the measurement system was adequate, since the measurement variation was small relative to the full extent of variation.

7.7 If necessary, measurement variation can be reduced by applying the Statistical Engineering algorithm. Describe how each of the seven variation reduction approaches might be used to improve a measurement system.

Fix the Obvious

If there are substantial operator-to-operator or gage-to-gage differences, we may reduce measurement variation by using fewer operators or gages or by training the operators.

Desensitization

With desensitization, we need to change some input in the measurement system that makes it less sensitive to variation in a known dominant cause of measurement variation. Desensitization was the chosen approach to improve eddy current measurements of rotor hardness. The team explored changing a number of settings to see if it was possible to make the measurement system less sensitive to variation in the amount of dirt on the casting, the dominant cause of measurement variation. Ultimately, in this example, the approach failed (see the discussion in Chapter 16).

Feedforward Control

The use of feedforward control to improve a measurement system is difficult to imagine. One idea is to use different gages depending on the class of part we are to measure. This would be effective in reducing measurement variation if each measurement system performed better for one of the classes of parts. For instance, suppose we wish to measure the thickness of foam. Say the current measurement system (hand calipers) worked well for narrow foam, but not very well for thick foam. Then, to implement feedforward, when foam of different grades required measurement we would choose the best gage for the job.

Feedback Control

We can use a feedback control scheme if a measurement system drifts. To set up the controller, we select a reference part and measure it repeatedly in the short term to determine the average and the standard deviation. Then to implement, we measure the same part on a periodic basis and adjust the measurement system if the characteristic value is materially different from the average. We use the short-term variation to determine if the change is material.

Robustness

Maintenance on a measurement system can be thought of as the Robustness approach. We hope to reduce the measurement variation without knowledge of the dominant cause. Another idea to reduce measurement variation that uses the robustness idea is to take two or more measurements on each part and average the obtained results. This will reduce the measurement variation, at the expense of increased measurement costs, if the repeated measurements are independent.

Inspection

Inspection is not an option to improve a measurement process. We have no way to determine which measurements should be rejected.

Move the Process Center

Moving the process center is trivial for a measurement process. We just add some value to all measured values. The Move Process Center approach is appropriate to eliminate a known and persistent measurement bias.

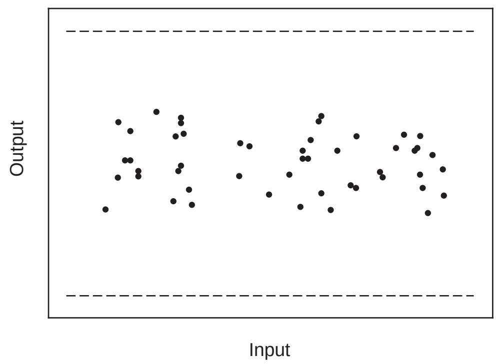
CHAPTER 8—NO EXERCISES

CHAPTER 9

9.1 Think of a process and problem you know well. Define various families of causes.

The solution depends on the process and problem chosen.

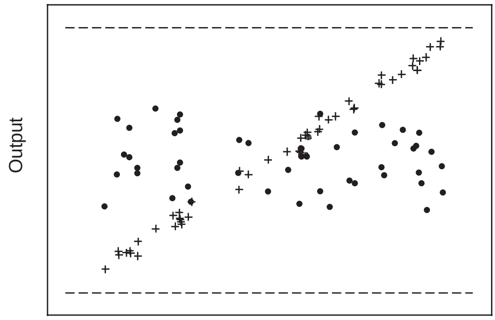
9.2 The following plot shows the results of a process investigation aimed at finding a dominant cause. The dashed lines give the full extent of variation in the output as defined by the problem baseline. Can the input be ruled out as a dominant cause of variation?



The simple answer is no, because we have not seen the full extent of variation. However,

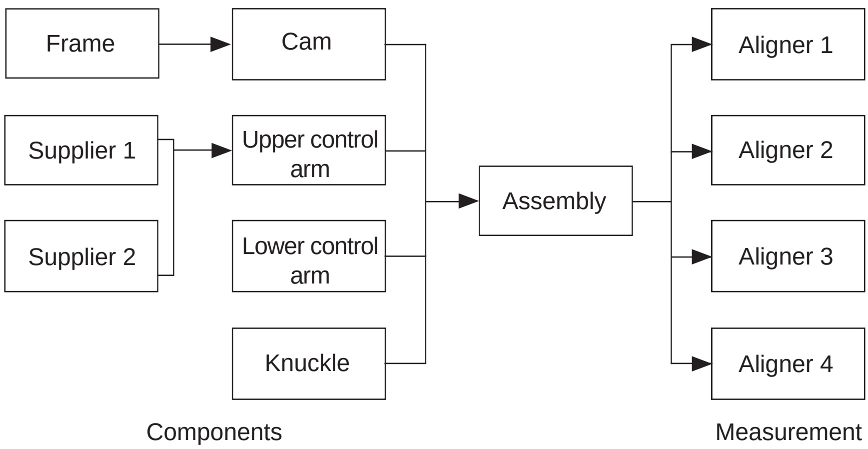
- If the full range of variation in the input is bigger than that observed, the input may be a dominant cause.
- If the full range of variation in the input has been observed, the input is not a dominant cause on its own but may be dominant in combination with another input (one that has not varied over its full range of variation in the investigation).

For instance, the input/output relationship may look like the plot that follows, where different plotting symbols are used to distinguish between two possible values for another input. In the original investigation, the level of the other input was restricted to the values given by solid circles.

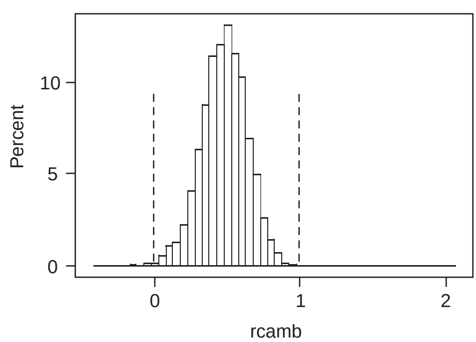


CHAPTER 10

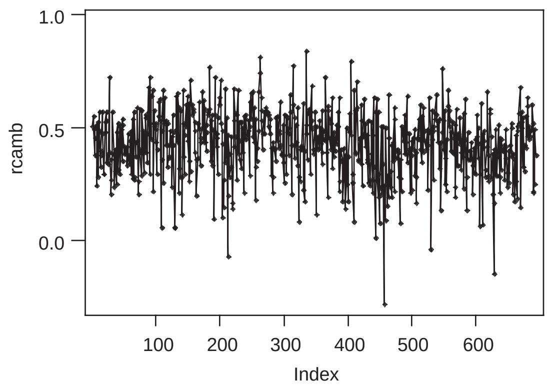
10.1 The flow chart that follows shows the major steps in an assembly process to set the wheel alignment of a truck.



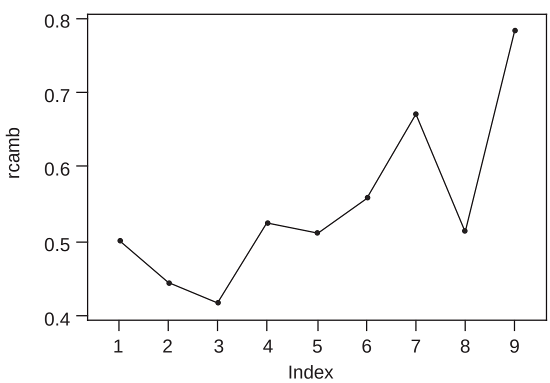
The characteristic of interest is right camber with specification 0.5 ± 0.5°. Camber is measured on every truck by one of the four gages (aligners). The process performance for right camber is shown as follows based on about 6200 consecutive trucks.



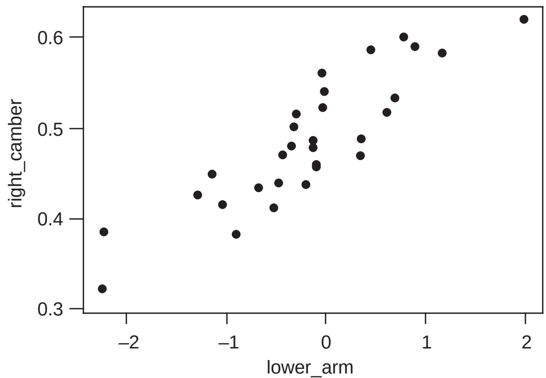
a. **Based on this histogram, can the measurement system be eliminated as a dominant cause of the camber variation?**
b. **What data could you collect to demonstrate that a dominant cause does not act in the measurement system?**
c. **How could you rule out the assembly operation as the home of a dominant cause?**
d. **How could you eliminate differences in the suppliers of the upper control arm as the home of a dominant cause?**
e. **The plot that follows shows the process behavior over three shifts. What family of causes can be eliminated based on these data?**



f. **The following plot shows the camber variation for the first nine trucks in the data set. What families can be ruled out as the home of a dominant cause using these data?**



g. **In a special study, one key characteristic of the lower control arm was measured for 30 trucks. The other components were specially selected to ensure that they were well within specification. Based on the plot that follows, is the lower control arm characteristic a dominant cause of right camber variation? Explain.**



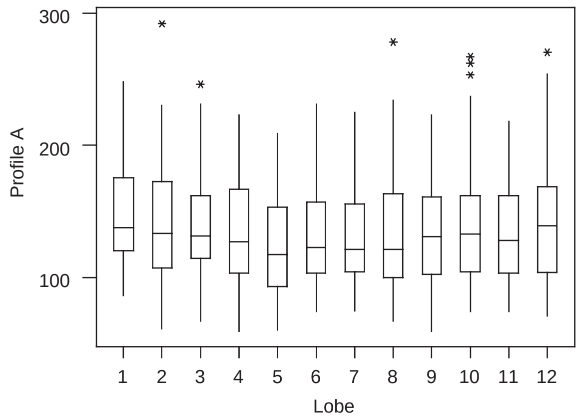
a. No. The measurement variation contributes to the range of values in the histogram but there is no way to tell from this investigation how much.
b. We need to conduct an investigation of the measurement system where right camber is measured more than once on a number of trucks.
c. If we are able to disassemble and reassemble the alignment components repeatedly without damage, we can assess the assembly operation. If disassembly/reassembly does not change the right camber values much, we eliminate the assembly operation as the home of a dominant cause. Repeated disassembly and reassembly using the production assembly process is difficult in this application.
d. If we were able to track which trucks contained the upper control arms from the different suppliers, we could stratify the right camber values by control arm supplier. If the control arm supplier was a dominant cause, the average right camber values in the two groups would be quite different, relative to the full extent of variation.
e. We are seeing roughly the full extent of variation in the output (right camber) within a short time frame. With this evidence we eliminate the family of causes that acts over the long term.
f. We have not observed the full extent of variation in right camber. It is premature to draw conclusions about the home of the dominant cause.
g. While the relationship between the lower control arm characteristic and right camber appears strong, we have not observed the full extent of variation in the right camber. We may ask if the range of lower control arm characteristic values observed in the investigation is typical of the process.

10.2 Consider again the camshaft lobe runout problem introduced in Chapter 1. Each camshaft has 12 lobes with a number of characteristics of interest. In a search for a dominant cause, we may compare the lobe-to-lobe and camshaft-to-camshaft families of variation. Using the problem baseline data given in the file *camshaft lobe runout baseline,* explore the relative sizes of the two families for the following characteristics and decide which family, if any, can be eliminated.

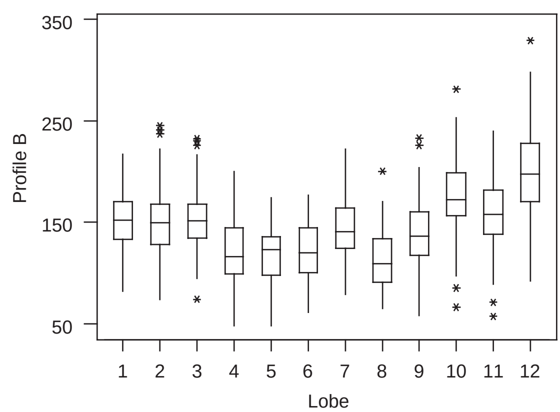
a. Profile *A*
b. Profile *B*
c. Profile *C*

We can graphically compare the lobe-to-lobe and camshaft-to-camshaft families of causes by stratifying the baseline data by lobes using box plots. The observations for each lobe show the effect of the camshaft-to-camshaft family, while comparing the box plots for different lobes shows the effect of the lobe-to-lobe family.

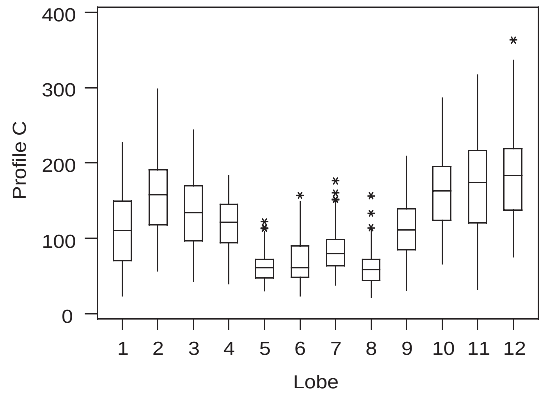
a. For profile *A,* the camshaft-to-camshaft family is the home of a dominant cause. There is very little difference from lobe to lobe in either the average or variation of profile *A*.



b. For profile *B,* we conclude the dominant cause acts in the camshaft-to-camshaft family. There is, however, more variation (though still small) attributable to the lobe-to-lobe family than for profile *A*.



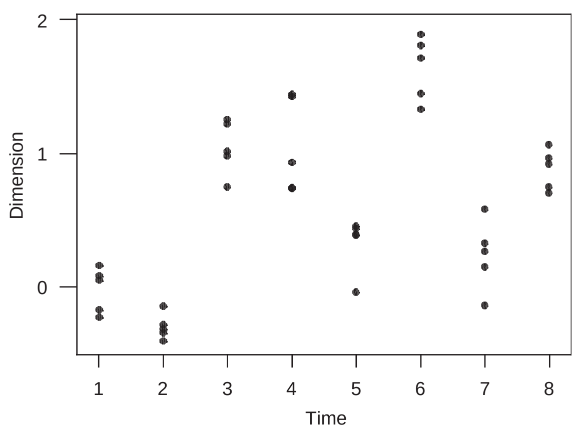
c. For profile *C,* the middle lobes exhibits less variation (and lower average) than for lobes on the end. The dominant cause is an interaction between lobe number and some, as yet, unidentified cause that acts in the lobe-to-lobe family. Note that ANOVA results corresponding to this plot will not reflect the differences in variation by lobe. The ANOVA analysis focuses on differences between the profile *C* averages across lobes.



10.3 In the manufacture of an injection molded part, a key crossbar dimension exhibited excess variation. The problem baseline estimated the standard deviation of the crossbar dimension as 0.46 with full extent of variation –0.3 to 2.0. The goal was to reduce the standard deviation to less than 0.25. An investigation showed the measurement system to be highly capable.

Next the team conducted a multivari investigation where five consecutive parts were sampled every 30 minutes for four hours. Analyze the data given in *crossbar dimension multivari*. Which family of variation can be eliminated as the home of the dominant cause?

From the plot of crossbar dimension by time (given as follows) we see the time-to-time family contains the dominant cause of variation. The dominant cause is acting over the 30-minute time periods, but not over consecutive parts, that is, not in the part-to-part family. We can eliminate the part-to-part family.



10.4 As described in Chapter 7, in a process that placed labels on bottles, the team searched for an acceptable measurement system. The file *label height measurement* contains the data from an investigation in which three operators using a hand feeler gage measured three specially chosen bottles three times on two

different days. The bottles were chosen to roughly cover the range of label height values seen in the process. From a baseline investigation an estimate of the overall standard deviation was 0.022. The results of a one-way ANOVA are:

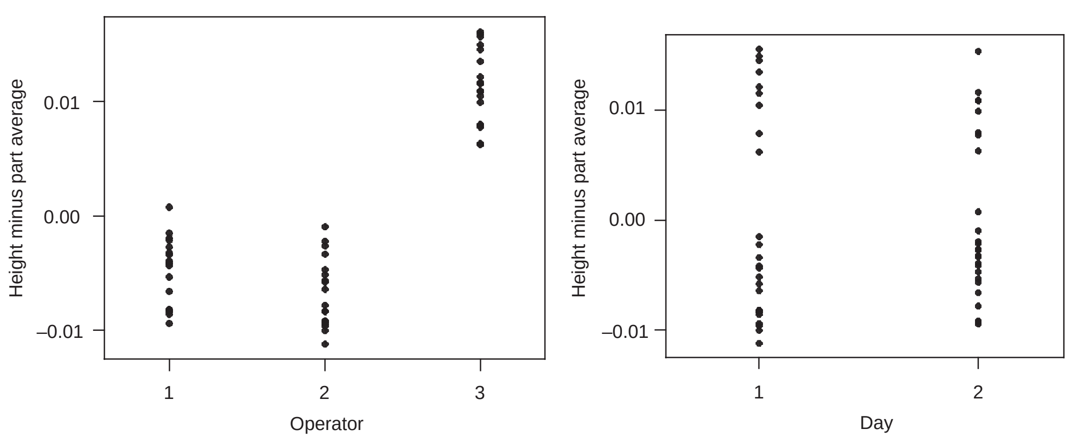
```
Analysis of Variance for height
Source      DF        SS        MS        F        P
part         2 0.0413263 0.0206631   263.10    0.000
Error       51 0.0040054 0.0000785
Total       53 0.0453317
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N      Mean     StDev ----+---------+---------+---------+--
1           18   0.06966   0.00965 (-*-)
2           18   0.10950   0.00930                (-*)
3           18   0.13705   0.00749                           (-*)
                                   ----+---------+---------+---------+--
Pooled StDev =   0.00886                0.075     0.100     0.125     0.150
```

We have stdev(due to measurement) = 0.00886, and thus

$$stdev(\text{due to process}) = 0.0204\left(\sqrt{(.022)^2 - (.00886)^2}\right)$$

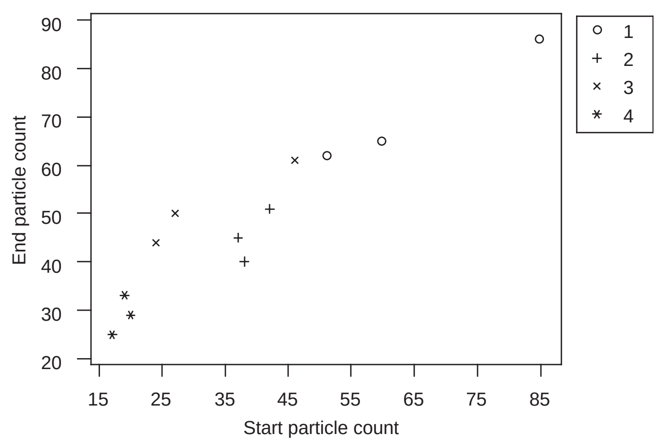
and an estimated measurement discrimination ratio of 2.3. The team decided to improve the measurement system before addressing the original label height variation problem. Reanalyze the measurement investigation results to eliminate families of possible dominant causes of measurement variation.

We start our analysis by calculating the heights minus the part average since we want to look for sources of variation other than part. Plotting height minus part average versus operator and day we get the plots given as follows:



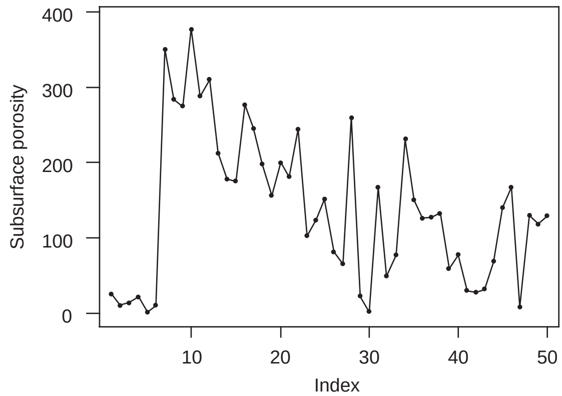
We see a clear difference in average height between operators and no difference between days. Operator three consistently gives larger values than the other two operators. This relative bias was found to be due to differences in procedure used by the three operators. With additional training the relative bias between operators was eliminated, and the measurement discrimination ratio was increased to around seven.

10.5 A process improvement problem was initiated to reduce the number of updings on a roof panel. Updings are small outward dents in the metal surface caused by contamination. The team discovered that the dominant cause was contamination before the forming process step. In an investigation, the team measured the particle count on coils directly after the arrival from steel supplier and again after blanking and stamping (before the forming process). They measured at the tail, middle, and head of four different coils. The data are given in the file *roof panel updings variation transmission*. What does the following scatter plot tell us about the dominant cause? The plotting symbols correspond to the four different coils.



Assuming the investigation has seen the full extent of variation in the output (which we are not given in the exercise description), we conclude the dominant cause acts in the raw blanks. The blanking, stamping, and forming process steps all transmit but do not add variation.

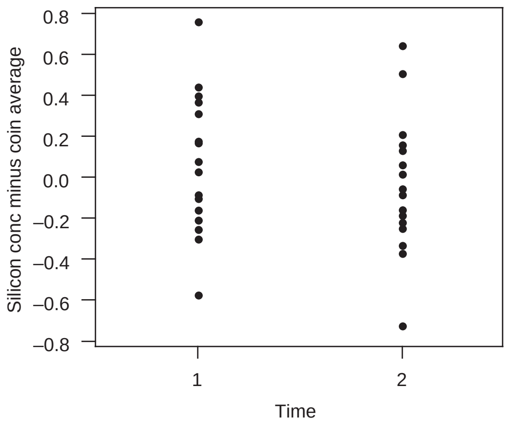
10.6 In the engine block porosity example discussed in Chapter 10, the team found the occurrence of high porosity coincided with production directly after breaks. To explore this clue further, they conducted another investigation in which the porosity of 50 consecutive blocks was measured. The first six blocks were taken from directly before the lunch break, and the next 44 blocks were the first ones produced after the break. The data are given in the file *engine block porosity run chart*. What does the observed pattern in the run chart tell us about the dominant cause?

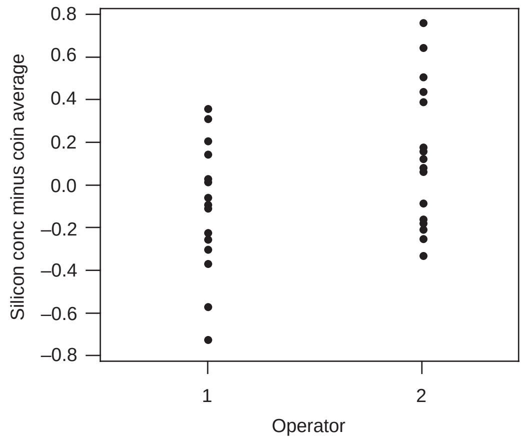


All six blocks from before lunch have extremely low porosity. Immediately following the break, the first six blocks have the highest porosity. The porosity gradually trends downward. This pattern implies the dominant cause is some input that changes abruptly during the breaks and then gradually returns to the original level. The return time is greater than 44 blocks. The team concluded that pouring temperature was a suspect dominant cause. The temperature of the iron decreased during breaks because there was no external heat source in the pouring ladles. The pouring temperature gradually increased after a break since the pouring ladles are frequently replenished with hot iron.

10.7 High silicon concentration in cast iron is undesirable as it was found be a dominant cause of fluidity variation. However, measuring the silicon level can be difficult. The measurement process consisted of sampling the molten iron by pouring sample coins for testing. The coins are then machined and polished before being spectrochemically analyzed. The full extent of variation in percent silicon as measured in the current process was 1 to 4%. The measurement system was investigated by repeatedly measuring three different coins that roughly covered full extent of variation in the observed percent silicon. Two operators measured each of the three coins three times on each of two days. The data are given in the file *iron silicon concentration measurement*. Analysis of the measurement results estimated the measurement standard deviation as 0.33. The corresponding estimate of the process standard deviation was 0.5; thus the discrimination is too small at around 1.5. The team needs to improve the measurement system. Using the existing measurement investigation data, are there any clues about the dominant cause of the measurement variation?

We start our analysis by calculating the silicon concentration minus the average for each coin since we want to look for sources of measurement variation other than differences from coin to coin. Some plots of the data are given as follows.



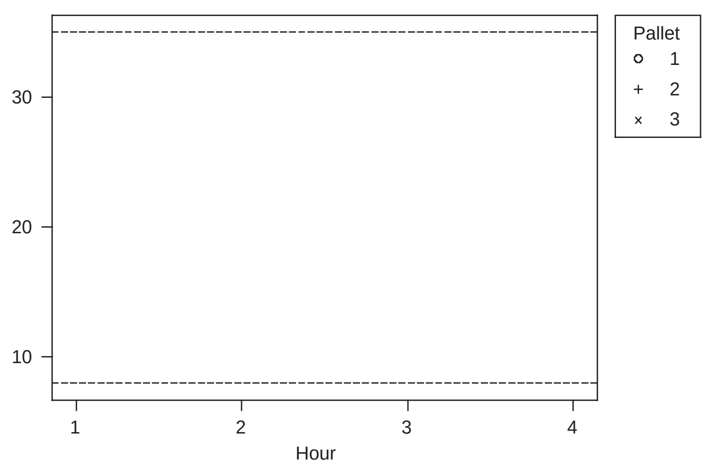


We see a small difference between the two operators, but no dominant cause. The dominant cause is acting over the short term. Using other investigations, the team was unable to determine the dominant cause of measurement variation. In Chapter 19, we describe a robustness investigation used to look for changes in this measurement process that would result in less variation.

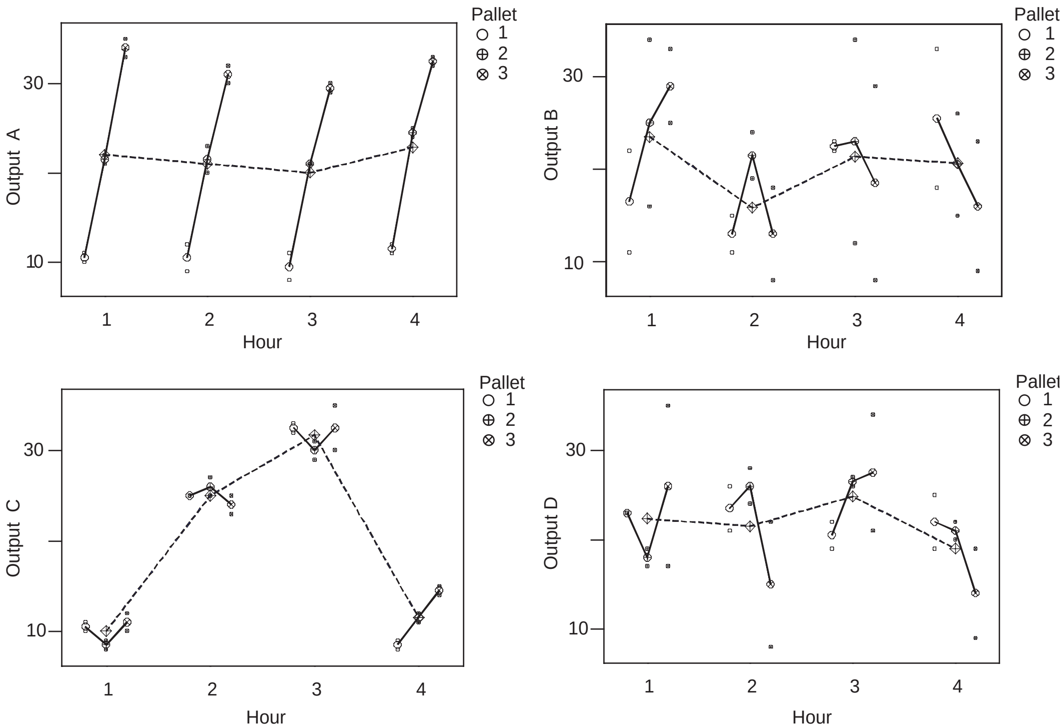
CHAPTER 11

11.1 In a multivari investigation, two consecutive pieces are selected from each of three pallets once per hour. Sketch the appearance of the multivari chart that shows all three families at the same time if a dominant cause lies in the following family. Use the following multivari chart template in which the dashed lines indicate the full extent of variation.

a. Pallet-to-pallet family
b. Part-to-part family
c. Hour-to-hour family
d. An interaction between the part-to-part and pallet-to-pallet families



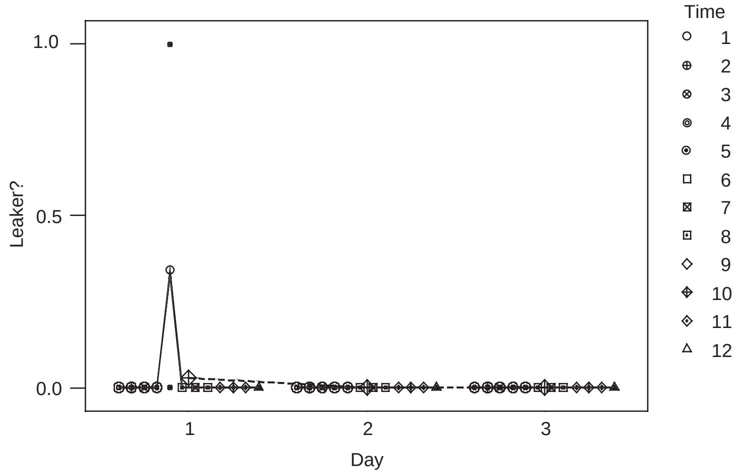
We give examples of the possible multivari charts, where the output name matches the question part. We suppose the full extent of variation is 8 to 35 units.



Note in the multivari chart for output *D,* the output has the most variation for pallet 3.

11.2 In the engine block leakers example, introduced in Chapter 1, the baseline defect rate was 2–3%. The team conducted a multivari investigation where three consecutive blocks were taken at twelve different times throughout the day. The investigation continued for three production days giving a total of 108 castings. Each block was tested for leaks. The data are given in the file *engine block leaks multivari*. What can you conclude?

The resulting multivari chart is:



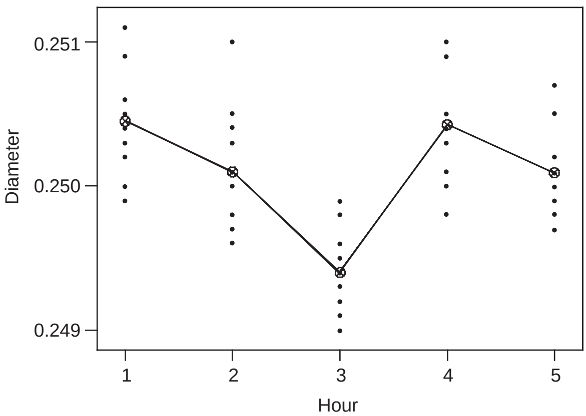
There is only a single leaker in the 108 blocks. This was not surprising considering the low baseline leak rate. This plot does not provide much information. In general, multivari investigations are not effective for binary output characteristics unless large sample sizes are used and the proportion defective is plotted.

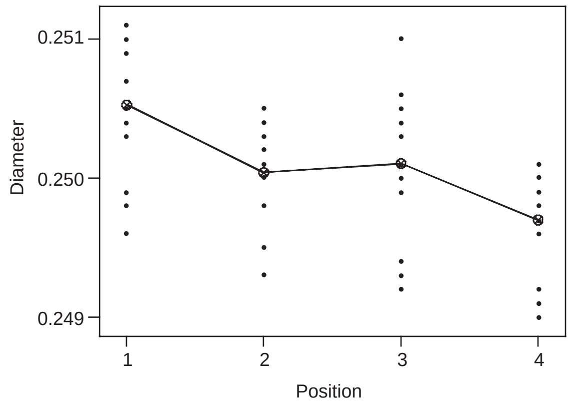
11.3 At an intermediate operation the team planned a multivari investigation in which three consecutive parts were taken from each of two machines operating in parallel once every hour for two days. Consider two different processes. In the first process, the order of the parts coming from upstream is preserved, while in the second process the order is jumbled. When interpreting the resulting multivari chart (think specifically about the part-to-part family), what difference does it make which process we are observing?

In the first process, the part-to-part family includes inputs that act from part-to-part upstream and in the intermediate operation. In the second process, the part-to-part family includes inputs that act part to part in the intermediate operation and all other upstream causes. For example, causes that acted in the time-to-time family upstream appear to act in the part-to-part family because of the jumbling.

11.4 In a multivari investigation, the diameter of a transmission shaft was measured at four positions (left and right side at two different orientations) for three consecutively sampled shafts each hour. The data are available in the file *transmission shaft diameter multivari*.

a. What conclusion can you draw from the multivari charts that follow?

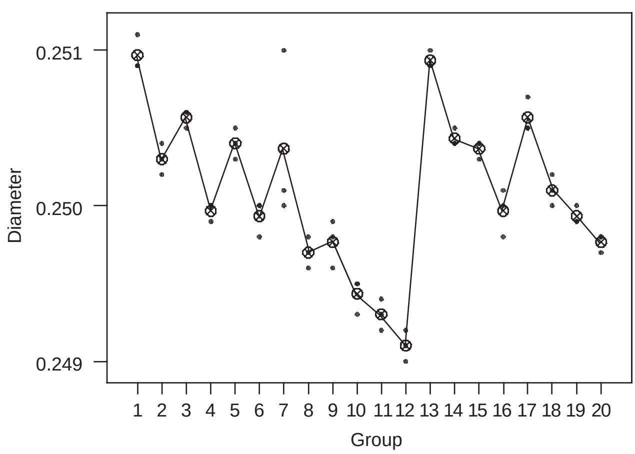




b. Using the data assess whether the dominant cause acts in the shaft-to-shaft family.

a. From the two multivari charts, we see that a dominant cause acts in the hour-to-hour family.

b. To assess the shaft-to-shaft family we define a new variate group to uniquely identify the 20 groups of three shafts that are consecutive. The multivari chart using group (=[hour – 1] ¥ 4 + position) is:

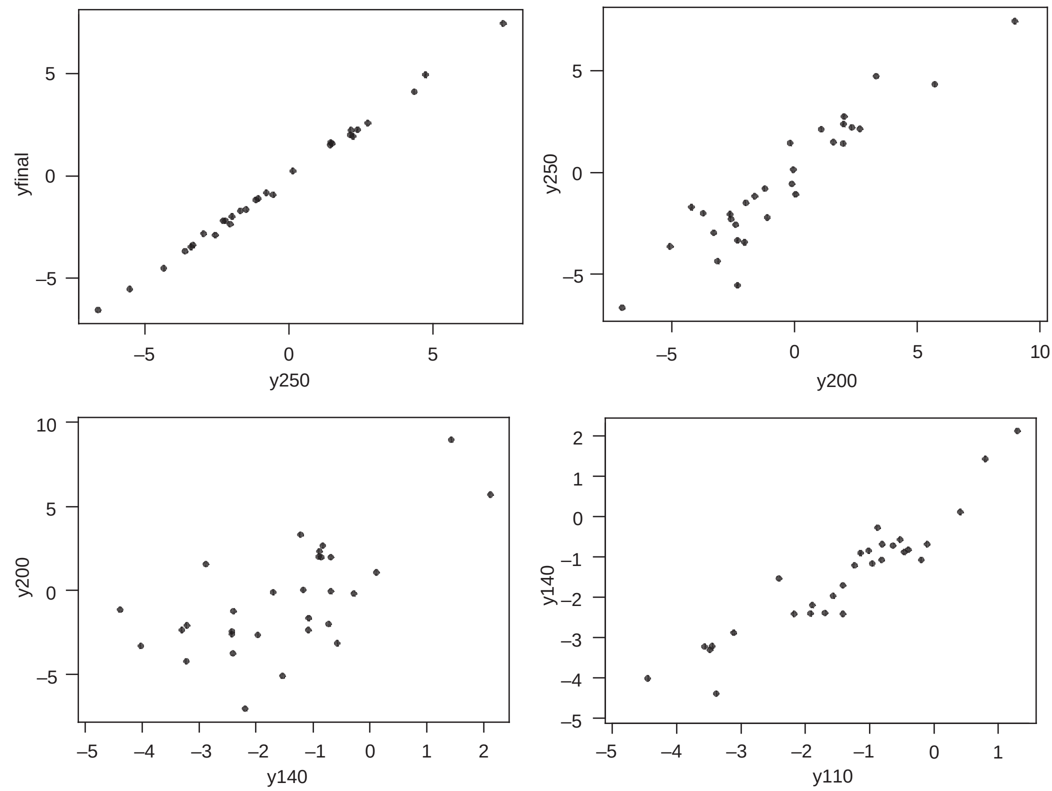


The variation from shaft to shaft inside each group is small, so the shaft-to-shaft family does not contain a dominant cause.

11.5 In the production of engine blocks, bore diameters are key characteristics. Bore diameter is measured at three heights and two orientations in each of the eight bores in each block. The team used Statistical Engineering to address a problem

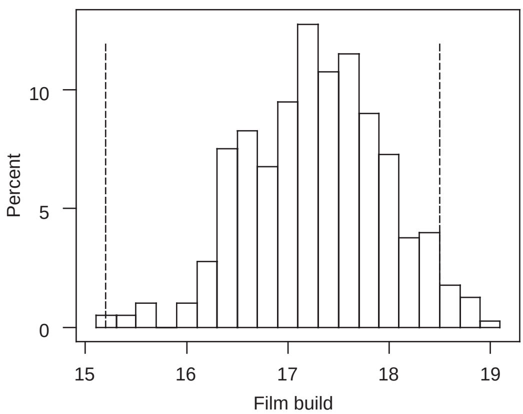
of excess bore diameter variation. The baseline investigation found a standard deviation of 3.04 and the full extent of variation of –9 to 9 as measured from nominal in microns. There were no strong differences between the different bores, heights, or positions. Another investigation concluded that the measurement process was adequate. To isolate the processing step where the dominant cause acts, the team selected 30 engine blocks haphazardly from a day's production. In the investigation the bore diameter (measured from nominal at that processing step) in the first bore at the top position and first orientation was measured at each of five processing steps in the machining part of the process. The data are given in the file *block bore diameter variation transmission*. Which processing step is home to the dominant cause?

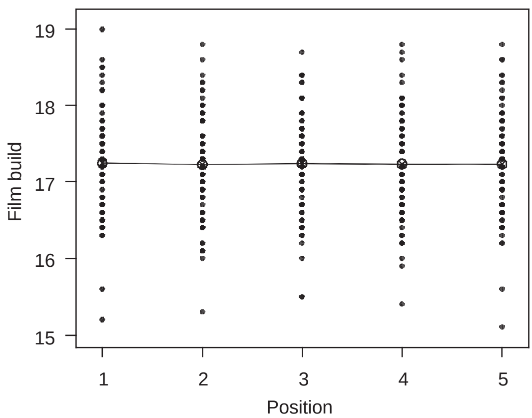
The plots showing incoming and outgoing diameters (from nominal) are given as follows. The output is labeled based on the operation after which it is measured.

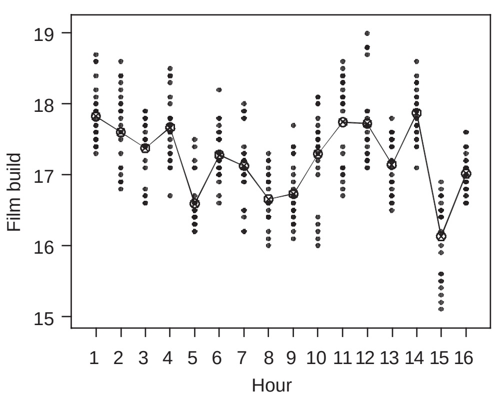


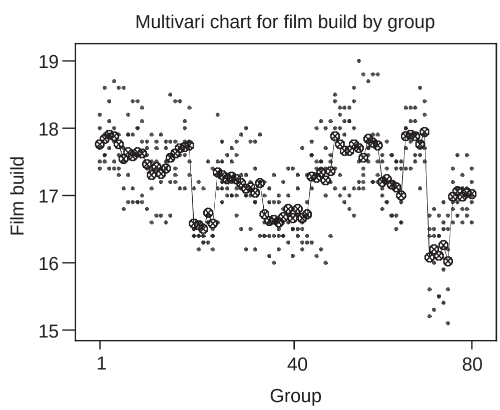
Starting from the end of the process, we see that little variation is added between operation 250 and the final diameter measurement, and also little variation is added between operations 200 and 250. Based on the plot in the lower left hand side, the dominant source of variation acts between operations 140 and 200 (including possibly operation 200 itself).

11.6 In the paint film build example described in Chapter 3, a baseline investigation found the standard deviation was 0.315, with an average of 16.2 thousandths of an inch. The full extent of variation was 15.2 to 18.5. To search for a dominant cause, the team conducted a multivari investigation where they measured the film build at five positions on five cars in a row every hour for two shifts (16 hours). This resulted in a total of 400 film build measurements. The data are given in the file *paint film build multivari*. Based on the plots that follow, what conclusions can you draw? We define group as (hour – 1) ¥ 5 + position.







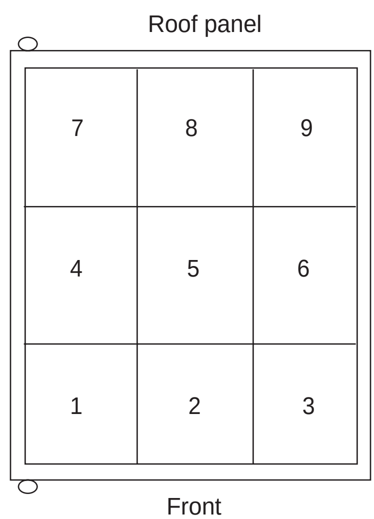


The histogram of the multivari data covers more than the full extent of variation, so we know the dominant cause has acted. From the multivari charts, we conclude that the dominant cause acts in the hour-to-hour family. We can use a one-way ANOVA to estimate the within-hour standard deviation as 0.47, much less than the baseline standard deviation 0.68. In other words, if we could make the average thickness the same at each hour, we could reduce the variation substantially.

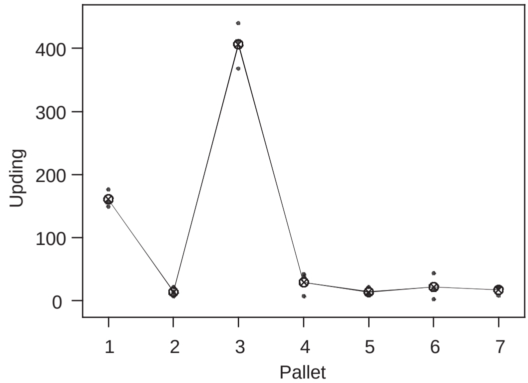
11.7 A team wanted to reduce the number of updings on a roof panel. Updings are small outward dents in the metal surface caused by contamination. A baseline

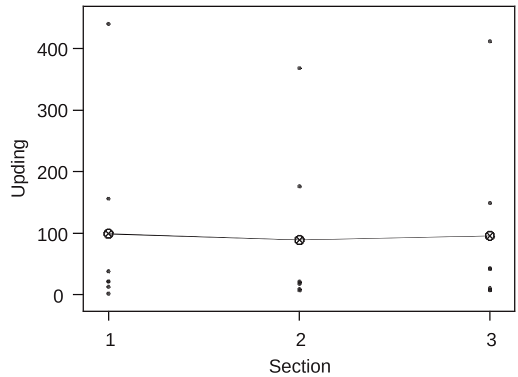
investigation found that the total number of updings in 20 consecutive panels ranged between 5 and 438. To search for a dominant cause the team conducted a multivari investigation where the number of updings was counted for 20 consecutive roof panels from three sections of seven different pallets of steel sheets. Originally, the plan was to repeat this data collection over two separate days. However, the team found the full extent of variation from the baseline was observed on the first day so they stopped collecting data. The data are given in the file *roof panel updings multivari*.

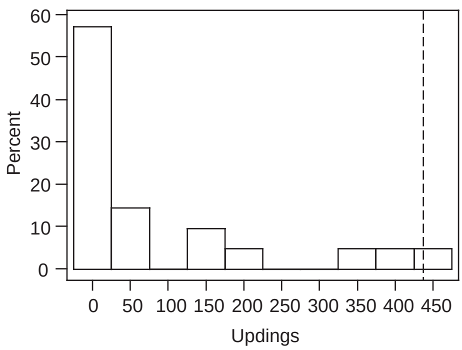
a. Analyze the data using multivari charts and draw conclusions.
b. When the number of updings was counted they were classified into one of the nine locations as numbered in the schematic that follows. Analyze the multivari data using a concentration diagram based on the given schematic.

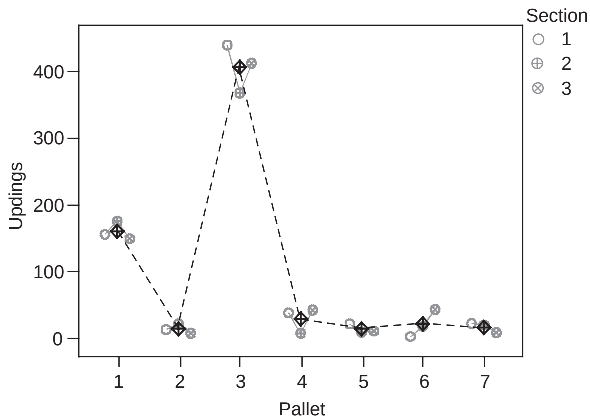


a. The multivari data covers the full extent of updings variation. The multivari charts suggest a dominant cause acts in the pallet-to-pallet family of causes. There is no section-to-section effect and no evidence of an interaction between the pallet and section families.

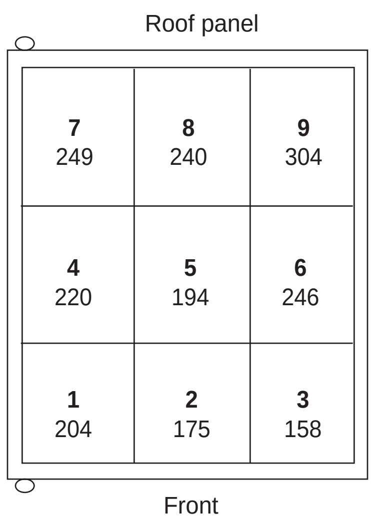








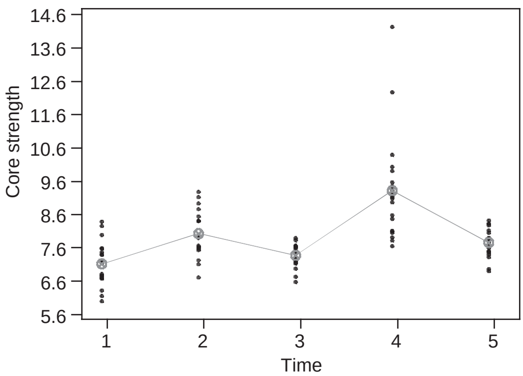
b. Plotting the number of updings by location using the concentration diagram does not show any clear patterns. We can eliminate the location-to-location family.

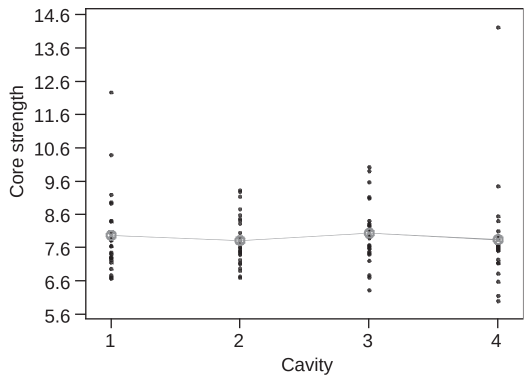


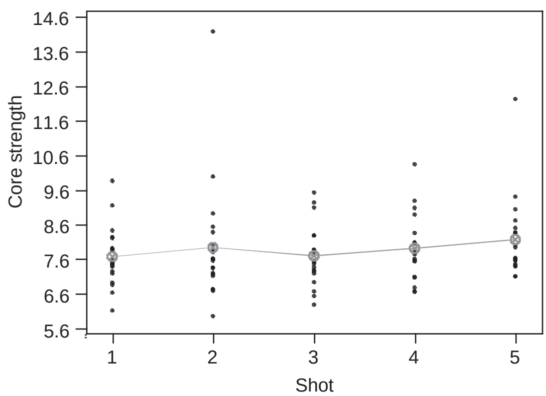
11.8 The baseline investigation for the sand core example discussed in Chapter 1 involved taking five samples over a single day of five consecutive shots of four

cavities each. The data are given in *sand core strength baseline*. What conclusions can you draw?

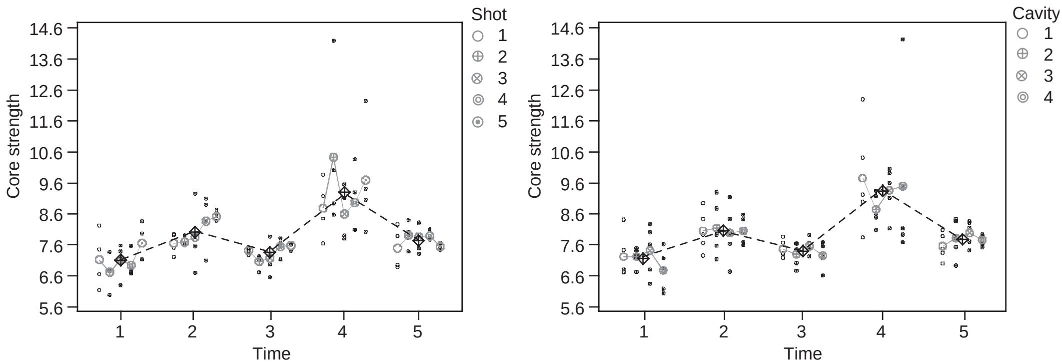
Multivari charts using one input at a time follow. Since we are analyzing the baseline data, the core strength values by definition cover the full extent of variation. We see no strong cavity, time, or shot effects. However, there is much greater variation at time 4 than the other times. The core-to-core family is a large source of variation only at time 4.







Exploring the data further, we make multivari plots by time together with shot and cavity in turn. As there is no clear pattern, the shot and cavity families can be eliminated.

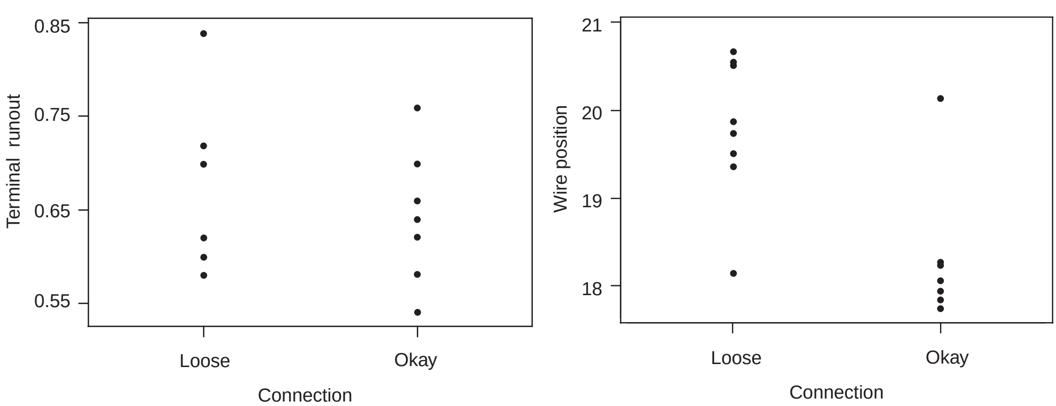


We need to look for a dominant cause consistent with the extra core-to-core variation at time 4.

CHAPTER 12

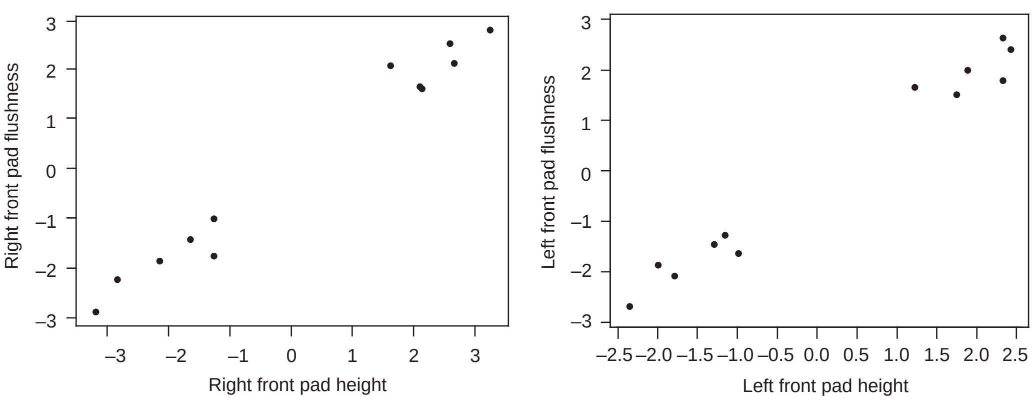
12.1 Vehicle plant and warranty complaints for loose spark plug wires at the spark plug end prompted an improvement project. As a result of several investigations, the family of causes related to push forces on the wires was the home of a dominant cause. A further investigation then compared eight loose and eight good connections. For each of the 16 connections, the team measured the terminal position of wires and terminal runout of the spark plug in millimeters. The data are given in the file *spark plug connection comparison*. What do the data tell us about the dominant cause?

Plotting the terminal runout and wire position by connection quality, as given, suggests wire position is a dominant cause. Note that there is one value in each group where wire position does not explain the connection quality. It is possible there is also a secondary cause. Before proceeding the team should verify that wire position is a dominant cause.



12.2 A sunroof installation process suffered from a 90% rework rate due to a lack of flushness. Flushness is defined as the difference in height between the sunroof seal and the metal roof. It is measured using digital calipers at six points (three at the front and three at the back). A baseline investigation showed that flushness problems were most pronounced at the two front corners with full extent of variation –3.5 to 4 mm and standard deviation 1.25 millimeters. A goal of reducing the front corner flushness variation to 0.5 and a range of –2 to 1 millimeters was established. Based on engineering knowledge, the team felt that only two types of characteristics could lead to flushness variation, namely roof crown height and attachment pad height. When the roof is adapted to allow installation of a sunroof, six installation pads are added. Based on this knowledge, the team selected six vehicles with large positive flushness and six vehicles with large negative flushness on both front corners. The sunroof modules were removed and the six attachment pad heights and roof crown height were measured at the front and back. The data are given in the file *sunroof flushness input-output*. What conclusions can you draw?

To start the analysis, we fit a multiple regression for both left front flushness and right front flushness with all the eight inputs. The residual variation for the two regression models was 0.35 (left) and 0.32 (right). Since both the residual variations are less than the target of reducing flushness variation to 0.5, the results appear promising. Plotting the data shows that a dominant cause of the right front flushness variation is the right front pad height. Similarly, a dominant cause of the left front flushness variation is the left front pad height.

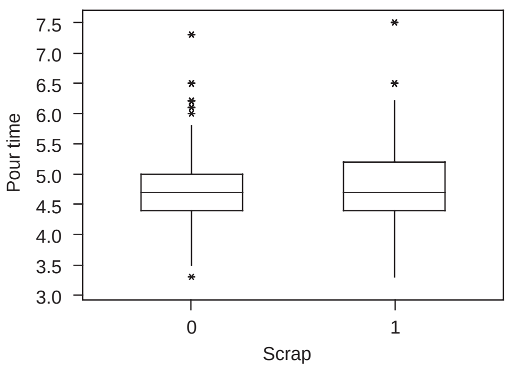


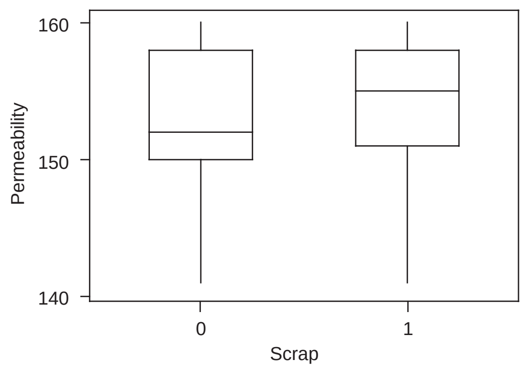
The regression results given for left front flushness show that by eliminating the dominant cause we could reduce the variation in flushness variation to 0.30 for the left side and 0.35 for the right side.

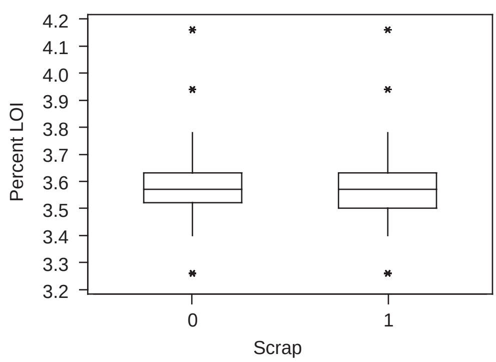
```
The regression equation is
left front flushness =  -0.117 + 1.05 left front pad height

Predictor        Coef     SE Coef          T       P
Constant     -0.11745     0.08689      -1.35   0.206
left front      1.05208     0.04677      22.49   0.000

S = 0.2996      R-Sq = 98.1%     R-Sq(adj) = 97.9%
```

12.3 An example related to sand defects in manifolds was discussed in Chapter 12. Before the problem mentioned in Chapter 12, the team carefully monitored the process for a shift. Each of 970 manifolds was classified as either scrap (due to sand issues) or not scrap. In addition many inputs relating to each manifold, including some discrete inputs such as mold number and continuous inputs such as pour time were recorded. In some cases the linkages were difficult to establish, and the team did the best they could. The data are given in the file *manifold sand scrap comparison*. What conclusions can you draw?

There are no clear differences in the distributions of any of the input values when we stratify by scrap or not scrap. The box plots that follow are typical.







The team concluded that linking input values to individual manifolds was too difficult. Many of the inputs are only measured a few times each shift and there is great uncertainty about the time linkages between the condition of each manifold and the measured input values. The team next conducted the investigation described in Chapter 12.

CHAPTER 13

13.1 In a verification experiment there were two suspects at two levels. The low and high levels for each suspect were chosen based on the extremes from historical variation. The results of the first three runs of the experiment are shown in the following table.

Input *A*	Input *B*	Order	Output
Low	Low	2	13
Low	High	3	16
High	Low	1	17
High	High	4	?

Given that the full extent of output variation is 12 to 30, what conclusions can you draw about the dominant cause?

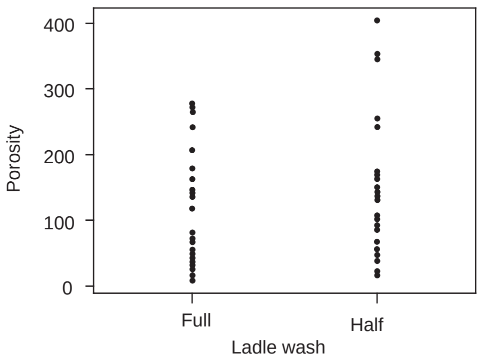
Suspects *A* and *B* alone are not dominant causes. However, without making the last run, we cannot tell if there is a dominant cause that involves both inputs. If the run with both *A* and *B* at the high level gives an output value near 30, the dominant cause involves both inputs. If the last run gives an output value near 17, then neither suspect is a dominant cause.

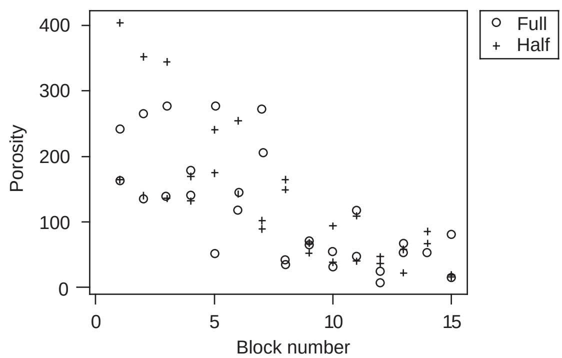
13.2 In the engine block porosity example discussed in the text and exercises of Chapter 10, a dominant cause of porosity acted immediately following scheduled breaks in production. Based on this evidence, the team identified two suspects: iron pouring temperature and the addition of ladle wash. During work stoppages, iron that remained in the six pouring ladles cooled off because there was no external heat source. At the start of the break, ladle wash was added to the ladles to protect the refractory (surface). The team could not easily manipulate the pouring temperature, but they could change the amount of ladle wash. They conducted a verification experiment in which they added the normal amount of wash to ladles 1, 3, and 5 and half the normal amount to the other three ladles over two lunch breaks. At each break, they measured the porosity of the first 30 blocks poured (five from each ladle). The data are given in the file *engine block porosity verification*.

a. What have we learned about the identity of the dominant cause of porosity?
b. Explain how the effects of ladle number and the presence or absence of ladle wash are confounded. Does this matter?

c. **Is it a problem that in this verification experiment we have not observed the behavior of the process before lunch breaks?**

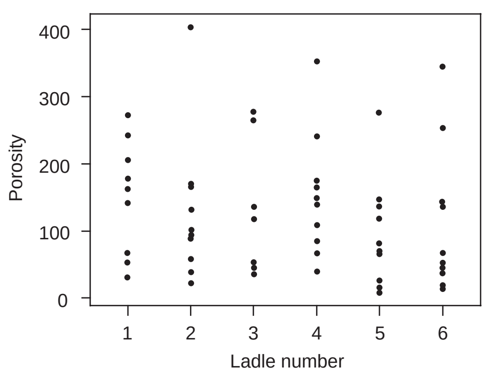
a. We use a box plot to stratify the porosity by the amount of ladle wash. We also plot porosity against block number for the two levels of ladle wash.





Ladle wash is not a dominant cause of porosity. Because the porosity decreased as the block number increased, the team concluded that pouring temperature was a dominant cause. They accepted the risk that some other unknown cause, confounded with pouring temperature, was actually responsible for the change in porosity.

b. In the experiment, the effects of ladle number and the presence or absence of ladle wash are confounded, since ladles 1, 3, and 5 always had full ladle wash and ladles 2, 4, and 6 always had half ladle wash. As shown in the following plot, the porosity scores were consistent across the odd- and even-numbered ladles. Based on engineering knowledge, the team believed there was no other cause that matched the change in the amount of ladle wash. Thus, in this case the confounding does not limit the conclusions in any meaningful way.



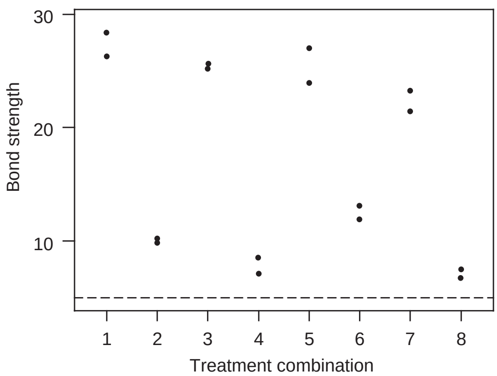
c. Observing the process before breaks was not necessary because the team was trying to determine if wash or temperature was the dominant cause of the porosity. These suspects changed at, or after, the break.

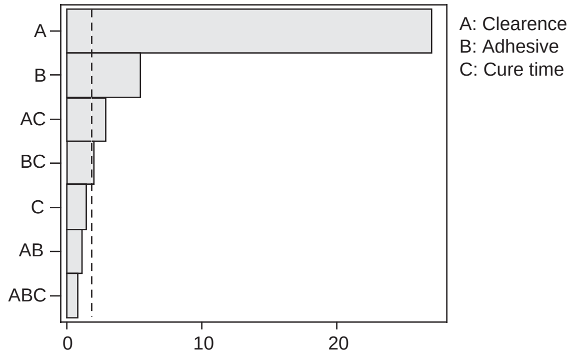
13.3 The manufacture of a tube assembly required a protective nylon sleeve to be positioned and bonded to a tube. The bond strength of this tube assembly was occasionally tested using a destructive test where the sleeve was subject to increased tensile shear load until failure. In the current process, the average pull-off force was around 15 pounds, but roughly 8% of assemblies tested had a pull-off force less than the desired minimum of five pounds. The team decided to try to solve the problem by reducing the variation in pull-off force rather than by increasing the average pull-off force. A number of investigations were conducted to find the dominant cause. A multivari investigation suggested that the dominant family of causes was tube-to-tube. At this point, the team decided to conduct an experiment to search for a dominant cause using the limited process information they had gathered. They planned a factorial experiment with three suspects—clearance between the sleeve and tube, amount of adhesive, and cure time—all consistent with the tube-to-tube family clue. The team chose the low and high levels of each suspect to roughly match their range in regular production. The levels of clearance were achieved by sorting sleeves and tubes. There were two replicates of each treatment, and the run order was randomized. The data are given in the file *nylon bond strength verification* and summarized in the following table:

Treatment	Order	Clearance	Adhesive	Cure time	Bond strength
1	6, 7	Low	Low	Low	26, 28
2	14, 8	High	Low	Low	10, 10
3	2, 3	Low	High	Low	25, 26
4	12, 5	High	High	Low	7, 9
5	16, 4	Low	Low	High	24, 27
6	13, 10	High	Low	High	12, 13
7	15, 11	Low	High	High	23, 21
8	1, 9	High	High	High	7, 7

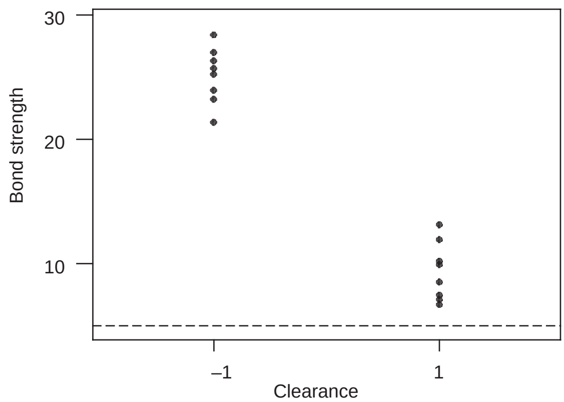
What do the results tell us about the dominant cause?

Plotting bond strength by treatment and creating a Pareto plot of the effects based on a full model gives the following plots.





We see that input *A* (clearance) has by far the largest effect. A plot of bond strength by clearance (input *A*), given as follows, clearly shows that clearance is a dominant cause of bond strength variation, and that low clearance gives higher average bond strength.



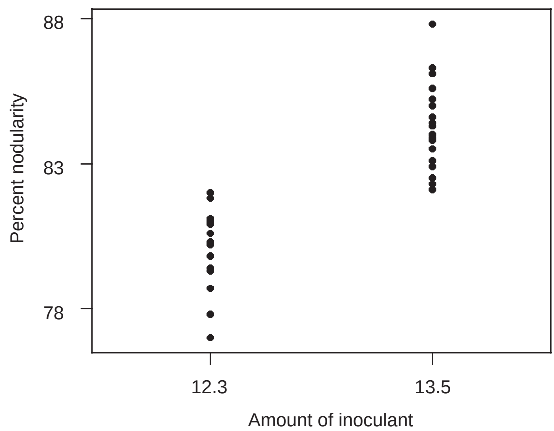
13.4 Steering knuckles are produced in a gray iron casting process. Around 2% of castings were scrapped because the percent nodularity was too small. In this example the team did not clearly establish a problem baseline. The team thought the cause must be related to the inoculation of the molten iron using a silicon-based alloy. The inoculant was added as the iron was poured to increase nodularity (and thus casting strength). Based on observing the process, the team noticed that the amount of inoculant added by the automated delivery system seemed to vary. The desired amount of inoculant was obtained by slowly shaking the inoculant onto a plate. The plate was designed to tip automatically when the required weight of inoculant was present. The team saw that the location on the plate where the inoculant fell varied, and they thought that this might influence when the plate tipped and thus how much inoculant was delivered. The team decided to verify inoculant amount as the dominant cause of nodularity variation. In the verification experiment, they produced a total of 20 castings at each of two levels of inoculant amount, 12.3 and 13.5 grams. For the

experiment the inoculant was carefully weighed and added by hand. The experiment consisted of eight runs of five castings each. The order of the eight runs (four at each level) was randomized. For each of the 40 castings the percent nodularity was determined. The data are given in the file *steering knuckle strength verification* and are summarized in the table that follows:

Run	Inoculant amount	Order	Percent nodularity
1	12.3	2	81.8, 79.4, 80.3, 80.6, 79.3
2	12.3	3	79.8, 77.0, 77.8, 79.3, 78.7
3	12.3	8	80.9, 82.0, 80.6, 80.6, 81.1
4	12.3	4	81.0, 79.4, 77.0, 80.6, 80.2
5	13.5	5	82.5, 86.1, 82.3, 83.5, 85.2
6	13.5	7	82.1, 84.6, 83.9, 85.0, 85.6
7	13.5	6	85.0, 87.8, 83.1, 84.0, 84.4
8	13.5	1	85.0, 84.3, 86.3, 83.8, 82.9

a. What considerations should the team have used in choosing the two levels for inoculant?
b. Why was randomizing the order of the runs important?
c. Has the team verified the dominant cause of nodularity variation?

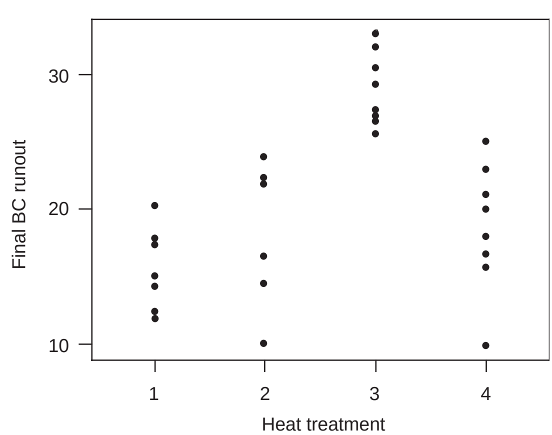
a. The team chose the levels to cover the range of inoculant values seen in regular production. Over a single shift, the team determined the range 12.3 and 13.5 grams by weighing the amount of inoculant obtained using the automatic delivery system on a separate scale.
b. The team chose an experiment with four replicates of each level of the suspect and five repeats for each run. Here, since the team had no information about the time family that contains the dominant cause, the randomization is extremely important. There is a danger that some unknown input may change in the same way as the suspect. Four replicates may not be sufficient. The repeats do not help.
c. Plotting the percent nodularity by amount of inoculant gives:



Because there was no baseline, we cannot tell if we have seen the full extent of nodularity variation in the experiment. However, since all of the castings with percent nodularity less than 80 would be scrapped, the team concluded that the amount of inoculant was the dominant cause. An obvious fix was to replace the plate with a cone. With the plate, not all the inoculant collected at the center. Using a cone, all the inoculant collects at the bottom, and the cone does not tip too early.

CHAPTER 14

14.1 In the camshaft lobe runout example, the team searched for a dominant cause of variation. As discussed in Chapter 10, they conducted a variation transmission investigation where runout was measured directly before heat treatment and after the final step of the process, on the same 32 parts selected over the course of one day. In the investigation the grinder (one of eight) and heat treatment spindles (one of four) used were also recorded. The data are given in the file *camshaft lobe runout variation transmission*. They found that a dominant cause of variation was the BC runout just after heat treatment and, more specifically, as shown in the plot that follows, that heat treatment spindle was a dominant cause.



In this example, the team decided not to reformulate the problem but to look for a more specific cause.

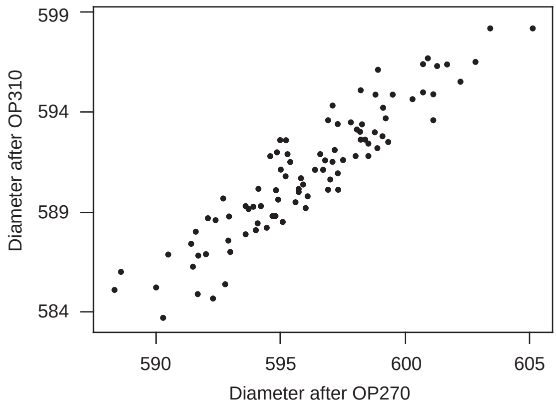
a. Discuss the advantages and disadvantages of the decision not to reformulate.
b. Suppose the team had reformulated the problem based on heat treatment spindle and that the original goal was to reduce the final runout standard deviation to less than 4.5. Using the results from a one-way ANOVA model based on heat treatment spindles, derive a goal for the new problem based on differences among spindle averages.

a. The advantages of not reformulating the problem are mainly that a new baseline for the runout after heat treatment (that is, the dominant cause, or the output in the new problem) does not need to be established, and we need not determine a goal for the new problem. The main disadvantage of not reformulating is that we still need to measure the final runout in future investigations. Had we reformulated, we would use the runout after heat treatment. In this case, it was not cheaper or easier to measure runout after heat treatment.
b. We use a one-way ANOVA model to assess the possible reduction in runout if we could align all the heat treatment spindles perfectly. The appropriate ANOVA results are:

```
Analysis of Variance for final BC
Source           DF        SS        MS        F        P
heat treatment    3     894.7     298.2    17.53    0.000
Error            28     476.3      17.0
Total            31    1371.0
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ---------+---------+---------+------
1           8    15.538     2.819   (----*----)
2           8    16.688     5.491     (----*----)
3           8    28.900     2.751                       (----*----)
4           8    18.675     4.729        (----*----)
                                   ---------+---------+---------+------
Pooled StDev =    4.124                   18.0      24.0      30.0
```

By perfectly aligning the heat treatment averages, we predict the standard deviation could be reduced to 4.1. Hence, to meet the goal we need to ensure that the heat treatment averages are close to equal.

14.2 In Chapter 11, the team found that the piston diameter directly after operation 270 was a dominant cause of final V6 piston diameter variation. The relationship is illustrated in the scatter plot that follows. The data are given in the file *V6 piston diameter variation transmission*.



The team decided to look further for a more specific dominant cause. Suppose, however, they had wanted to reformulate the problem in terms of the operation 270 diameter. Determine an appropriate goal for the reformulated problem. Recall that the goal for the original problem was to reduce the final diameter standard deviation to less than 2.0.

Fitting a linear model to the variation transmission investigation results gives:

```
The regression equation is
diameter after OP310 = 64.3 + 0.884 diameter after OP270

Predictor        Coef     SE Coef          T        P
Constant        64.27       22.29       2.88    0.005
diameter      0.88358     0.03739      23.63    0.000

S = 1.224       R-Sq = 85.6%     R-Sq(adj) = 85.4%
```

Using the regression results, we estimate that

$$stdev(\text{final diameter}) = \sqrt{0.884^2\, stdev(\text{OP270 diameter})^2 + 1.224^2}$$

Thus, to meet the goal of reducing the final diameter variation to less than 2.0, we solve to get stdev(OP270 diameter) < 1.58. Thus, a reasonable goal for the reformulated problem is to reduce the diameter variation at operation 270 to less than 1.6. Note that in the variation transmission investigation, the diameter variation at OP270 was 3.4, so the goal requires more than a 50% reduction in diameter variation.

CHAPTER 15

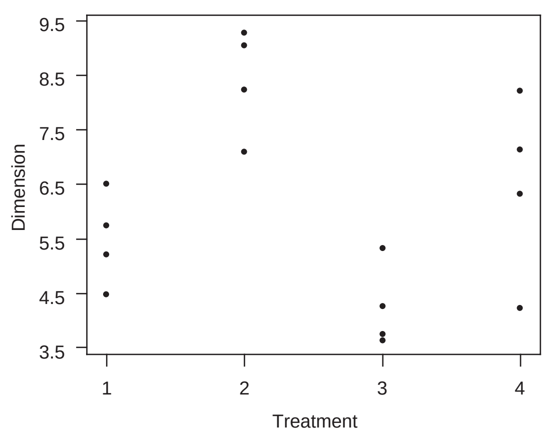
15.1 **Based on customer complaints about assembly difficulty, a team investigated fascia dimension variation. A baseline investigation found that some fascias were too large. The team felt that reducing the average size of the fascias could**

solve the problem (that is, they adopted the Move the Center approach). They planned a (full) factorial experiment with two candidates, cycle time and cure time, each at two levels to look for an adjuster. They chose the levels for each candidate based on engineering judgment. The results of the experiment are given in the file *fascia dimension move center* and in the following table. For each treatment, the team conducted four runs producing 10 fascias for each run. The order of the 16 runs was randomized over a day. In the data, we give only the average fascia dimension from each run and not the individual values.

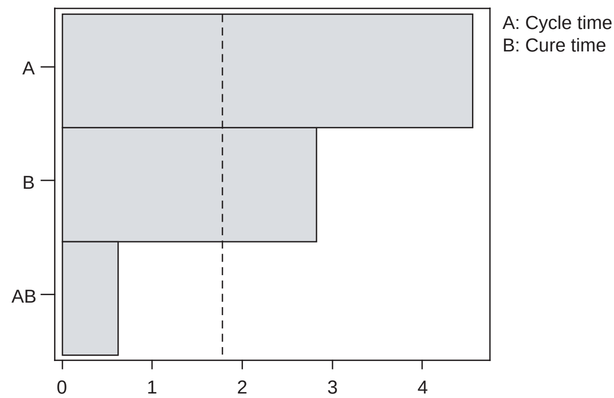
Treatment	Run order	Cycle time (minutes)	Cure time (hours)	Average fascia size (from nominal)
1	8, 10, 1, 14	85	5	4.50, 5.23, 5.75, 6.51
2	6, 15, 12, 2	113	5	7.12, 8.25, 9.06, 9.28
3	11, 3 , 9, 16	85	19	3.65, 3.75, 4.27, 5.34
4	13, 5, 4, 7	113	19	4.24, 6.31, 7.15, 8.22

a. Can cycle time or cure time be used as an adjuster?
b. Suppose the goal was to reduce the average fascia size to 3.0. What do you recommend?
c. What is the advantage of looking at the dimensions for all the fascias within a run, rather than the averages?

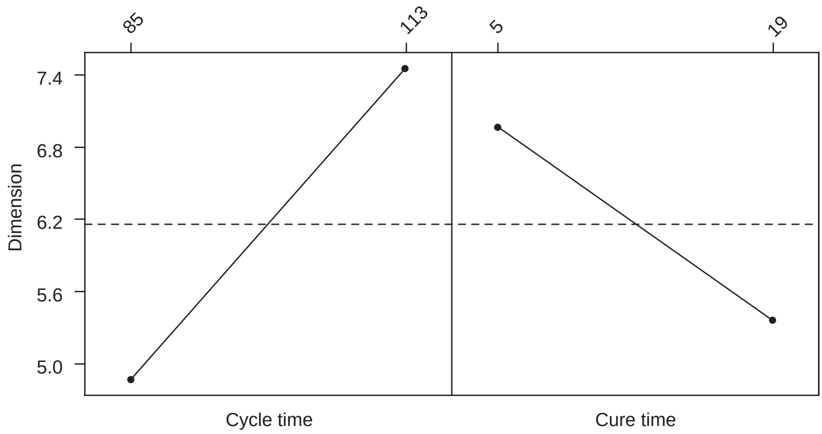
a. We first plot dimension by treatment. We see that there are differences between the average dimensions for different treatments. In particular, treatment 3, with low cycle time and high cure time, leads to smaller average dimensions.



Next, we fit a full model and look at the Pareto plot of the effects.



Both cycle time and cure time have large effects; hence, both are potential adjusters. The interaction between these two candidates is small. From the following main effect plots, we see that lowering cycle time and increasing cure time will reduce the average dimension.



b. None of the treatments used in the experiment gave an average dimension as small as 3. Shorter cycle time and longer cure result in smaller dimensions. The team should consider another investigation to calibrate the effects of changing both adjusters.
c. With the individual dimensions for all 160 fascias, we could look at the variation within each run. This would allow us to determine whether changing the candidate settings changed the short-term fascia-to-fascia variation.

15.2 An experiment was carried out to investigate four candidates to search for an adjuster of the formability safety margin of galvanized sheet metal trunk lids. The purpose was to increase the average safety margin from the baseline value

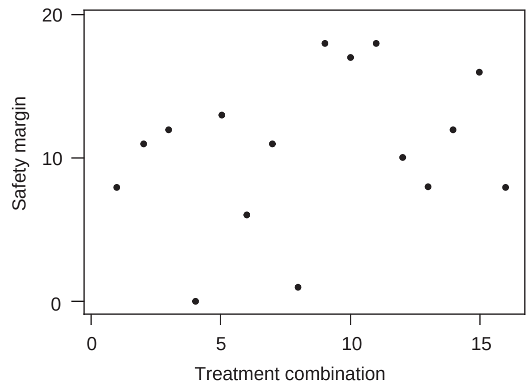
of 10.7. In the experiment, each candidate was tested at two levels, selected to be near the edge of what was physically possible—see the table that follows. Note that none of the treatments corresponded to the existing process settings.

Treatment	Run order	Tonnage	Lubrication	Blank size	Prebending	Safety margin
1	6	310	Unlubricated	949__1494_m m	No	8
2	16	375	Unlubricated	949__1494_mm	No	11
3	3	310	Lubricated	949__1494_mm	No	12
4	11	375	Lubricated	949__1494_mm	No	0
5	7	310	Unlubricated	965__1500_mm	No	13
6	15	375	Unlubricated	965__1500_mm	No	6
7	4	310	Lubricated	965__1500_mm	No	11
8	13	375	Lubricated	965__1500_mm	No	1
9	1	310	Unlubricated	949__1494_mm	Yes	18
10	12	375	Unlubricated	949__1494_mm	Yes	17
11	5	310	Lubricated	949__1494_mm	Yes	18
12	14	375	Lubricated	949__1494_mm	Yes	10
13	2	310	Unlubricated	965__1500_mm	Yes	8
14	10	375	Unlubricated	965__1500_mm	Yes	12
15	8	310	Lubricated	965__1500_mm	Yes	16
16	9	375	Lubricated	965__1500_mm	Yes	8

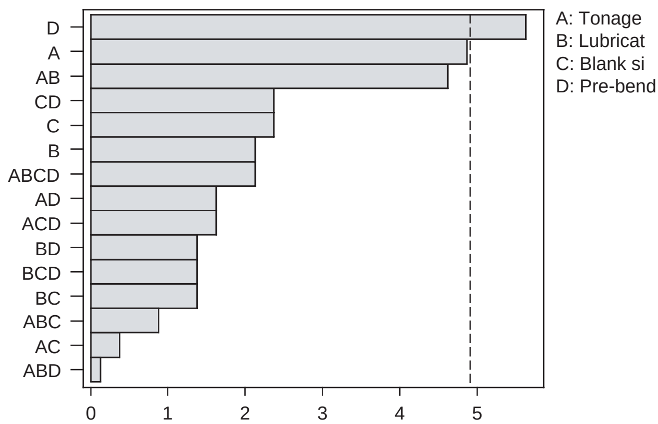
Press tonnage was very difficult to change so all eight runs with low press tonnage were carried out first. Within each group of eight runs, the order was randomized. The data are given in the file *sheet metal move center*.

a. Analyze the experimental data to see if any of the candidates is an adjuster.
b. Does the restriction on randomization required for this experiment make any difference to the conclusions we can draw?

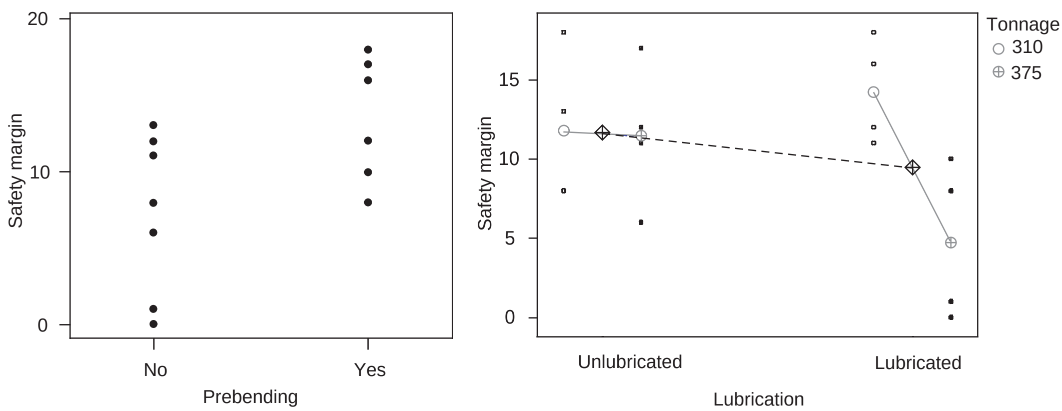
a. We start the analysis by looking at a plot of the safety margin by treatment. In this case, since there is only a single output for each treatment, it is difficult to draw conclusions from the plot. We do see that there are treatments with very different values for the safety margin.



Next, we fit a full model and examine a Pareto plot of the effects.



D has a large main effect and *AB* a large interaction effect. We follow up with plots of safety margin by prebending (input *D*) and an interaction plot of tonnage by lubrication.



We see that prebending leads to higher safety margin on average and that the safety margin is much more sensitive to the presence or absence of lubrication at the high level of tonnage than at the low level. The average safety margin for the two runs with prebending, low press tonnage, and lubrication was (18 + 16)/2 = 17, considerably higher than the baseline average 10.7. This finding needs to be confirmed. See the discussion in part b.

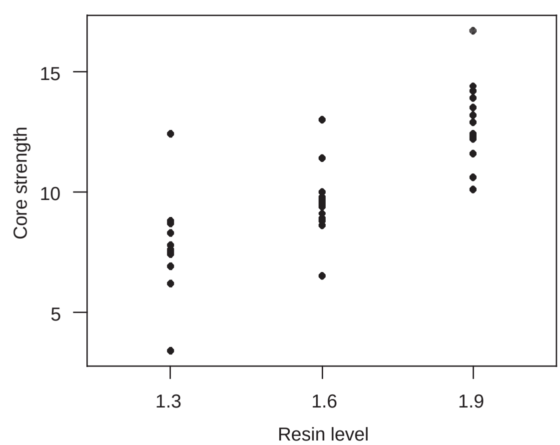
b. Without the randomization it is more likely that the effect attributed to press tonnage is due to some other (unknown) input. This could occur if the level of the unknown cause happened to change at (or near) the moment in time when press tonnage was changed. When interpreting the experimental results we cannot tell if this occurred, or not.

15.3 In the sand core strength example introduced in Chapter 1, too many cores were breaking during handling. A suggested solution was to increase the core strength (and thereby reduce core breakage) by increasing the resin concentration. It was known that increasing the resin would result in a stronger core. However, the precise nature of the relationship—that is, how much the core strength increases for a given change in resin concentration—was not known. An experimental investigation was planned to quantify the relationship. Three levels of resin concentration (1.3, 1.6, 1.9% by weight) were chosen based on engineering judgment. In the experiment, 40 cores for each level of resin were produced; 15 were measured for strength (using a destructive test) and the remaining 25 were processed to look for casting problems. The experiment consisted of three runs with 15 repeats. The order of the runs was not randomized. The data are given in the file *sand core strength move center*.

a. What can you conclude about the relationship between resin concentration and core strength?

b. The team used only three runs with 15 repeats for each run. Discuss the advantages and disadvantages of this plan compared with using five replicates for each treatment with three repeats each.

a. We plot strength by resin level as follows:



Because the increase in average strength is roughly linear as a function of the resin concentration, to quantify the relationship, we fit a regression model:

```
The regression equation is
core strength =  -3.76 + 8.61 resin level

Predictor        Coef     SE Coef          T        P
Constant       -3.762       1.609      -2.34    0.024
resin          8.6111      0.9943       8.66    0.000

S = 1.634       R-Sq = 63.6%     R-Sq(adj) = 62.7%
```

We expect, on average, a 0.86 unit increase in core strength for each 0.1 increase in the percent resin. The team had quantified the effect of resin concentration. At the 2% level of resin, one of the 25 cores led to a defective casting, so the team knew they should not raise the resin concentration to this level. As a result, although the team had now quantified how to move the process center, the approach was abandoned because of the fear of increased scrap due to core-related defects.

b. Using five replicates for each treatment would allow randomization of the order. This would help protect against some unknown input changing in the same way as resin in the experiment. A disadvantage is that it may be more difficult and expensive to change the resin level more often.

CHAPTER 16

16.1 In a sonic welding operation, problems arose due to poor weld strength, measured as pull-off force. The goal was to reduce the variation and increase the average pull-off force. The second goal is not addressed here. From the baseline, the full extent of variation for pull-off force was 0.9 to 3.0. The team discovered that the dominant cause acted in the time-to-time family. While they could not be more specific, the team felt that the dominant cause was related to material hardness, which was outside their control. They decided to try to desensitize the process to variation in the dominant cause.

The team planned a fractional factorial experiment with four candidates at two levels each in eight treatments. Using the results of regular process monitoring, they identified three time periods when weld strength was low, medium, and high relative to the baseline. In each period, they randomized the order and then produced a part with each of the eight treatments. The pull-off force data and plan are given in the file *sonic weld desensitization* and the table that follows. The three values in the columns Order and Pull-off force correspond to the three different time periods. The original settings of the candidates correspond to treatment 2.

Treatment	Order	A	B	C	D	Pull-off force
1	7, 2, 6	–1	–1	–1	–1	1.8, 2.1, 2.3
2	4, 6, 1	–1	–1	1	1	0.9, 2.0, 2.7
3	6, 3, 4	–1	1	–1	1	2.0, 2.4, 2.1
4	8, 1, 2	–1	1	1	–1	0.6, 1.5, 3.0
5	3, 7, 8	1	–1	–1	1	2.8, 2.9, 3.0
6	2, 4, 3	1	–1	1	–1	2.1, 3.0, 3.7
7	1, 8, 7	1	1	–1	–1	2.9, 3.2, 3.1
8	5, 5, 5	1	1	1	1	2.3, 2.5, 4.0

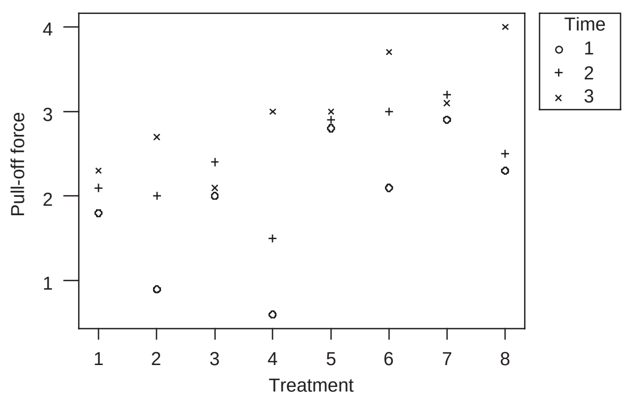
a. Explain why the team believed the dominant cause acted over the three runs for each treatment.

b. What levels of the candidates do you recommend to reduce the variation in pull-off force?

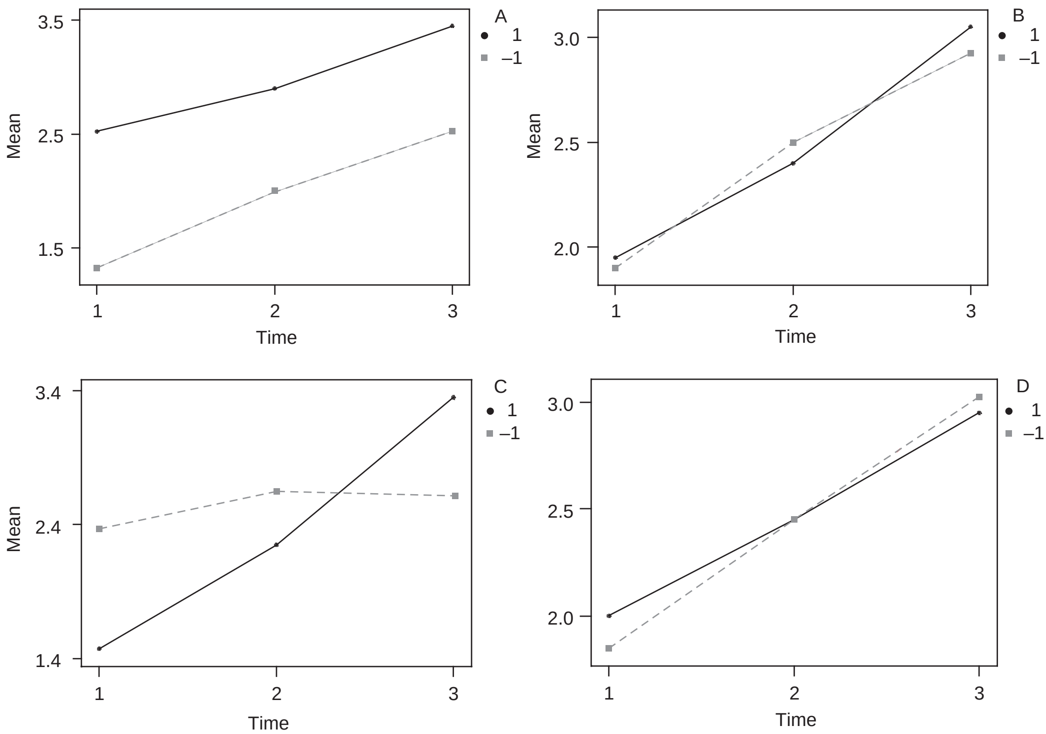
c. Another way to assess the results of this experiment is to summarize the output across each treatment using log standard deviation. Using this performance measure, do your conclusions differ from part b?

a. The team knew that the dominant cause acted time to time. Since they selected periods with low, medium, and high weld strength under the current conditions, they were confident that the dominant cause acted over the three periods. In the experiment, this assumption was verified since the range of pull-off force values for treatment 2, the current process settings, covered the full extent of variation.

b. To start the analysis we look at a plot of the pull-off force by treatment.

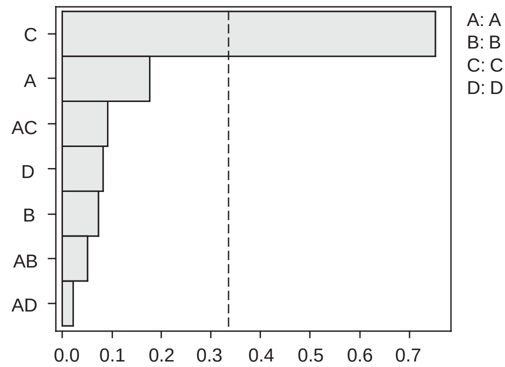


There are several promising treatments with little variation in pull-off force. To desensitize the process we are looking for a special kind of interaction between time (a surrogate for the dominant cause) and the candidates. The interaction plots for time versus the candidates (using the average pull-off force as the response) are

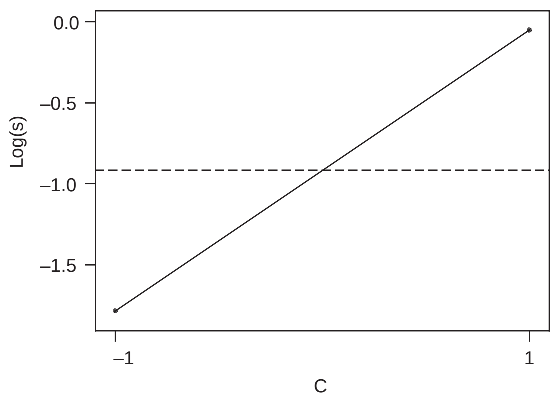


We see that changing the level of candidate *C* flattens the relationship between the pull-off force and time. More specifically, with the low level of candidate *C*, the time-to-time variation in pull-off force is much reduced. Changing the other candidates does not help. We recommend changing to the low level of candidate *C*, and leaving the remaining candidates at their current levels.

c. Fitting a full model in the four candidates and analyzing the results using the performance measure log(s) gives the Pareto effects plot that follows. The plot indicates that candidate *C* has the largest effect.



From the main effect plot that follows, we see that the low level of *C* reduces the variation in pull-off force. This is the same conclusion as in part b. In most cases the conclusions obtained with the two analysis methods will be the same. We prefer the analysis that looks directly at the candidate by cause interaction plots rather than the analysis based on log(s).



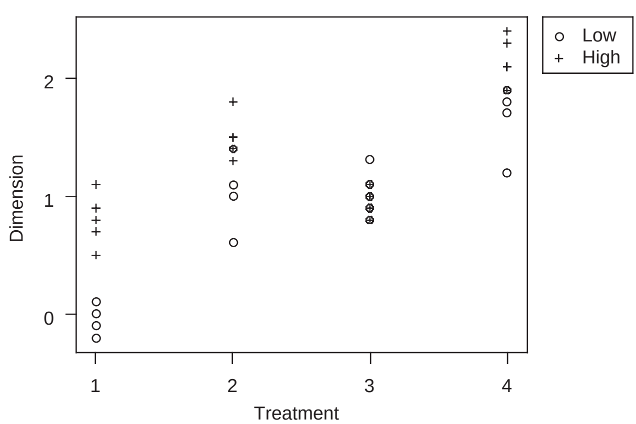
16.2 In the crossbar dimension example discussed in Chapter 12, the team found that the dominant cause of dimension variation was barrel temperature. Because it was hard to control in regular operation, the team decided to try to make the process less sensitive to barrel temperature variation. In the current process, barrel temperatures ranged over roughly 4°C. The team planned a half fraction factorial experiment with three candidates—target barrel temperature, injection pressure, and material—at two levels each, as shown in the following table. The current injection pressure and target barrel temperature were 1000 and 75, respectively. Note that although the variation in barrel temperature was the dominant cause, the target barrel temperature is a fixed input. Five crossbars were produced and measured in each run. For each treatment, there were two runs, one at the target barrel temperature plus 2°C and the other at the target barrel temperature minus 2°C. The data are given in the file *crossbar dimension desensitization* and in the table as follows.

Treatment	Target barrel temperature	Injection pressure	Material	Dimensions at barrel temperature –2°C	Dimensions at barrel temperature +2°C
1	75	1000	Old	0, –0.1, 0.1, –0.1, –0.2	0.5, 1.1, 0.8, 0.9, 0.7
2	75	1200	New	1.1, 0.6, 1.0, 1.4, 1.1	1.5, 1.8, 1.5, 1.4, 1.3
3	79	1000	New	1.1, 1.0, 1.3, 0.9, 0.8	1.0, 1.1, 0.8, 0.9, 1.0
4	79	1200	Old	1.2, 1.8, 1.8, 1.7, 1.9	2.3, 2.1, 2.4, 2.1, 1.9

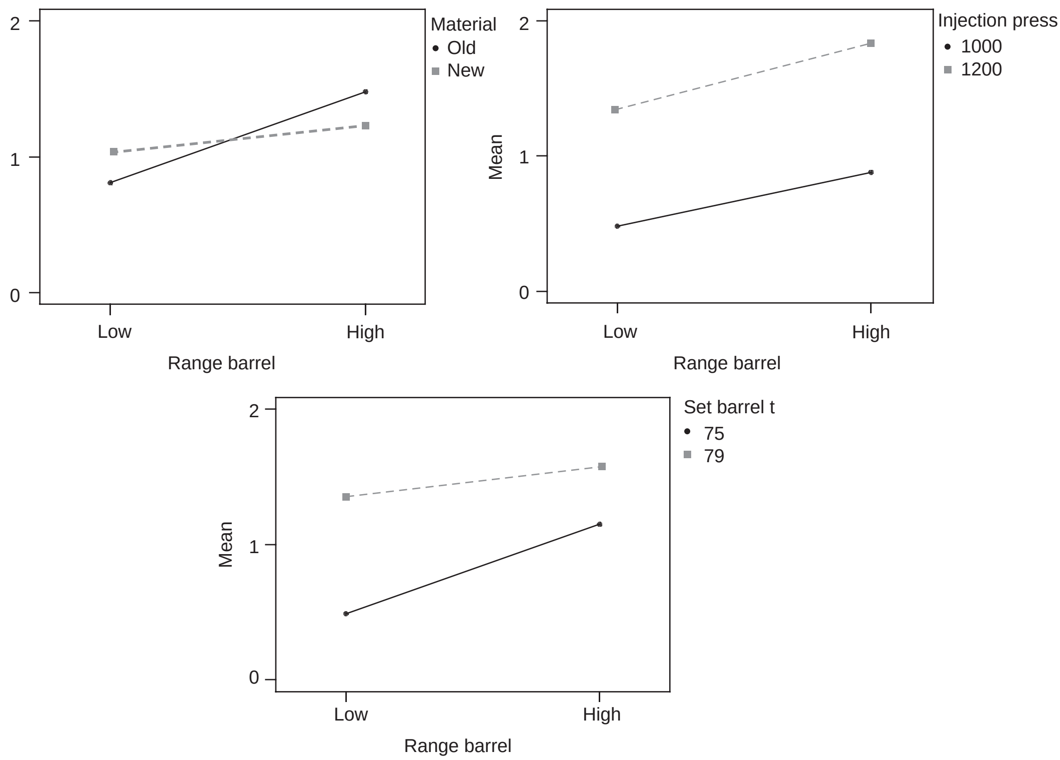
Since the average dimension can be easily adjusted, we focus the analysis on finding a way to make the process less sensitive to barrel temperature variation.

a. What levels of the candidates do you recommend?
b. Injection pressure and material were chosen as candidates based on engineering judgment. Looking again at the results presented in Chapter 12, what motivates the choice of target barrel temperature as a possible candidate?

a. We begin the analysis by plotting dimension by treatment with a different plotting symbol for each of the two levels of the dominant cause. Treatment 3 looks most promising. Changing to the higher target barrel temperature, using the new material, and keeping the injection pressure the same is expected to make the process less sensitive to variation in the barrel temperature.



We examine the nature of the interaction between the candidates and the dominant cause with the following plots (where the response is the average dimension):



If we increase the target barrel temperature and, to a lesser degree, use the new material, we see that the relationship between crossbar dimension and barrel temperature is flatter. In other words, dimension is less sensitive to barrel temperature variation at the higher target barrel temperature. Changing the injection pressure has no effect. The benefits of changing both the target barrel temperature and the material were later validated with another investigation. Note that the observed variation in treatment 3 has likely overestimated the variation when actually running the process at these settings, since we have forced barrel temperature to near its expected extremes with the new target temperature.

b. The nonlinear relationship between barrel temperature and dimension is shown in the left panel of Figure 12.3. In the plot, it appears there is less dimension variation for high than for low barrel temperatures.

16.3 In Chapter 16, we describe a desensitization experiment for the refrigerator frost buildup example where each refrigerator is subjected to only two extreme levels of environmental causes. Here we consider a hypothetical experiment in which each refrigerator is exposed to a number of environmental conditions to ensure that any chosen new design works well under any conditions, not just extreme conditions.

The experimental design for the four candidates—D1, D2, D3, and D4—is the same as in Chapter 16. Here we plan to test each of the eight refrigerators (treatments) under all eight possible combinations of the usage or environmental inputs as given in the following table:

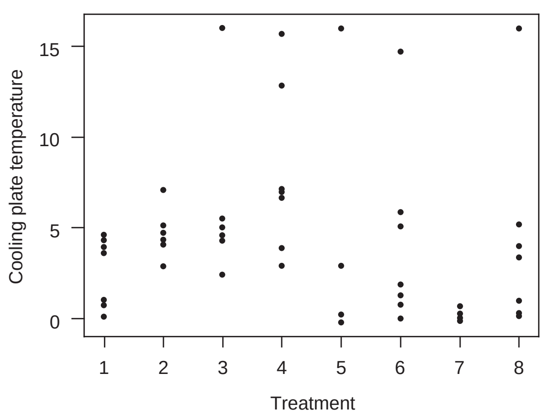
	Cause combination							
Varying input	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Ambient temperature (°C)	26	26	26	26	32	32	32	32
Relative humidity (%)	70	70	90	90	70	70	90	90
Door openings per hour	4	8	4	8	4	8	4	8

The experimental plan had 64 runs. To conduct the experiment, all eight refrigerators were simultaneously placed in a test chamber and exposed to each cause combination in the given order. The cooling plate temperatures are given in the file *refrigerator frost buildup desensitization2* and in the following table:

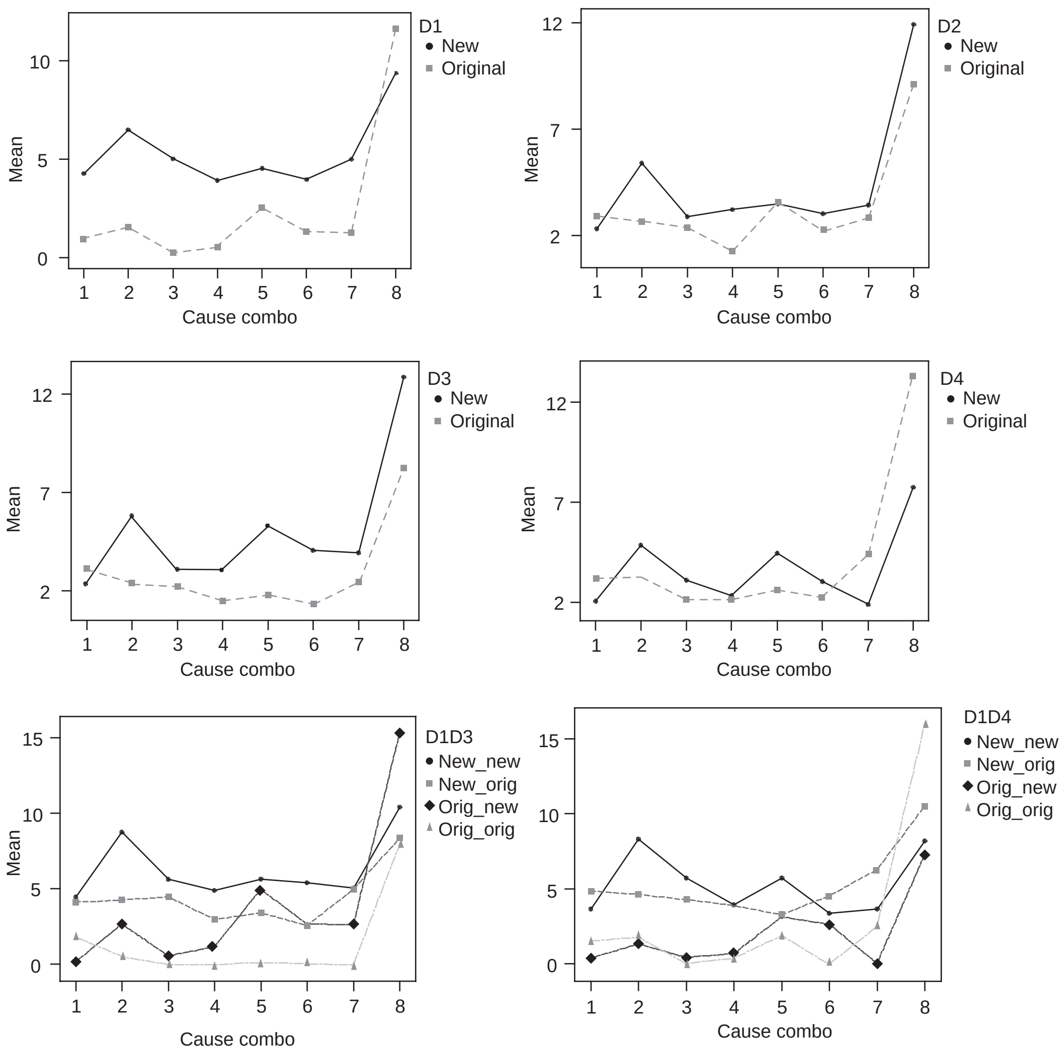
	Candidates				Cooling plate temperatures (in cause combination)							
Treatment	**D1**	**D2**	**D3**	**D4**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	N	O	O	N	3.6	3.9	4.6	1.0	4.4	0.1	4.4	0.7
2	N	O	N	O	5.1	4.7	4.3	2.9	4.2	4.1	7.1	5.1
3	N	N	O	O	4.6	4.6	4.3	4.9	2.4	5.0	5.5	16.0
4	N	N	N	N	3.8	12.8	6.9	6.9	7.1	6.7	3.0	15.7
5	O	O	O	O	2.9	0.2	–0.2	–0.2	–0.2	–0.2	–0.2	16.0
6	O	O	N	N	0.1	1.9	0.8	1.3	5.9	5.1	0.0	14.7
7	O	N	O	N	0.7	0.8	0.1	0.1	0.4	0.2	0.1	–0.1
8	O	N	N	O	0.2	3.4	0.3	1.0	4.0	0.2	5.2	16.0

In the table, we have coded the new and original settings for the candidates as N and O, respectively. What conclusions can you draw? Remember, the goal is to desensitize cooling plate temperature to changes in the environmental conditions.

We start the analysis by plotting the cooling plate temperatures for all of the eight environmental conditions (cause combinations) by treatment.



Treatment 7 gives consistently low cooling plate temperatures over all combinations of the environmental inputs. Treatment 7 corresponds to using the new levels for D2 and D4 and leaving D1 and D3 at their original settings. To examine how changing each candidate flattens the relationship between the changing environmental conditions and cooling plate temperature, we examine the appropriate interaction plots. To create an interaction plot of the environmental conditions versus D1 and D3, for example, we create a new column D1D3 with the four levels for the pair D1 and D3.



Not all possible interaction plots are given. The other plots give similar conclusions. From the interaction plots, we see that changing a single candidate or a pair of candidates does not reduce the variation in average cooling plate temperature. In particular, all combinations of the candidates have a relatively large cooling plate temperature for cause combination 8 (high temperature, humidity, and number of door openings per hour). The good properties of treatment 7 are due to a high-order interaction between the candidates that we cannot explore due to the confounding in the design.

16.4 There were excessive failures in the accelerated life testing of electric motors. Using a group comparison investigation, the team found that unevenness in the commutator shaft surface was a dominant cause of these failures. The team next reformulated the problem to one of reducing the unevenness in the commutator shaft. The surface unevenness is measured on a scale of 1 (smooth) to 10 (rough). With further investigation, the team determined that the dominant cause of the variation in the (final) smoothness was the shaft profile before machining. The team adopted the Desensitization approach. They decided to conduct a fractional factorial experiment with eight treatments using four candidates. For each of the eight treatments there were two runs, one that used a shaft with a premachined smooth profile, and a second that used a rough profile. The experimental plan and data are given in the file *electric motor failure desensitization* and the table that follows. The order of the runs was randomized.

						Smoothness	
Treatment	Depth	Grind time	Rotational speed	Feed rate	Order	Smooth profile	Rough profile
1	Shallow	Short	1800	Slow	4, 5	2	7
2	Deep	Short	1800	Fast	6, 11	3	8
3	Shallow	Long	1800	Fast	1, 14	1	9
4	Deep	Long	1800	Slow	16, 12	2	8
5	Shallow	Short	2400	Fast	13, 9	3	2
6	Deep	Short	2400	Slow	10, 8	1	4
7	Shallow	Long	2400	Slow	3, 7	2	3
8	Deep	Long	2400	Fast	15, 2	3	5

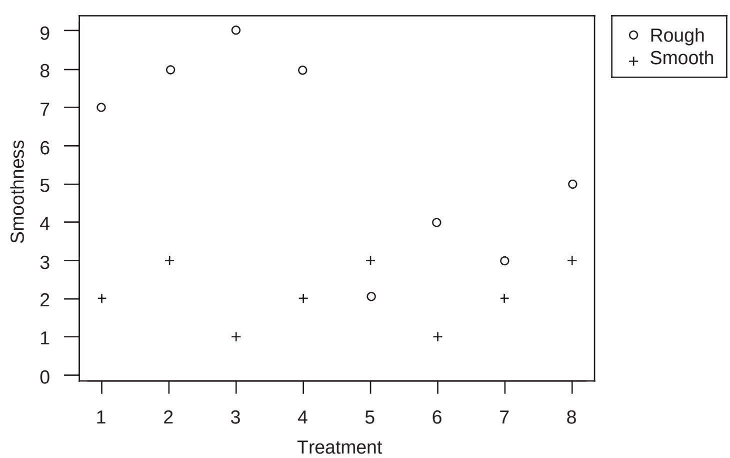
a. What is the confounding structure of the design? What limitations does this introduce?
b. What conclusions can you draw?
c. What would be the advantages and disadvantages of measuring the time to failure using the accelerated life test for each run rather than judging the smoothness of the commutator surface after machining?

a. The confounding structure including two- and three-input interactions in candidates and the dominant cause (profile) is given by MINITAB as

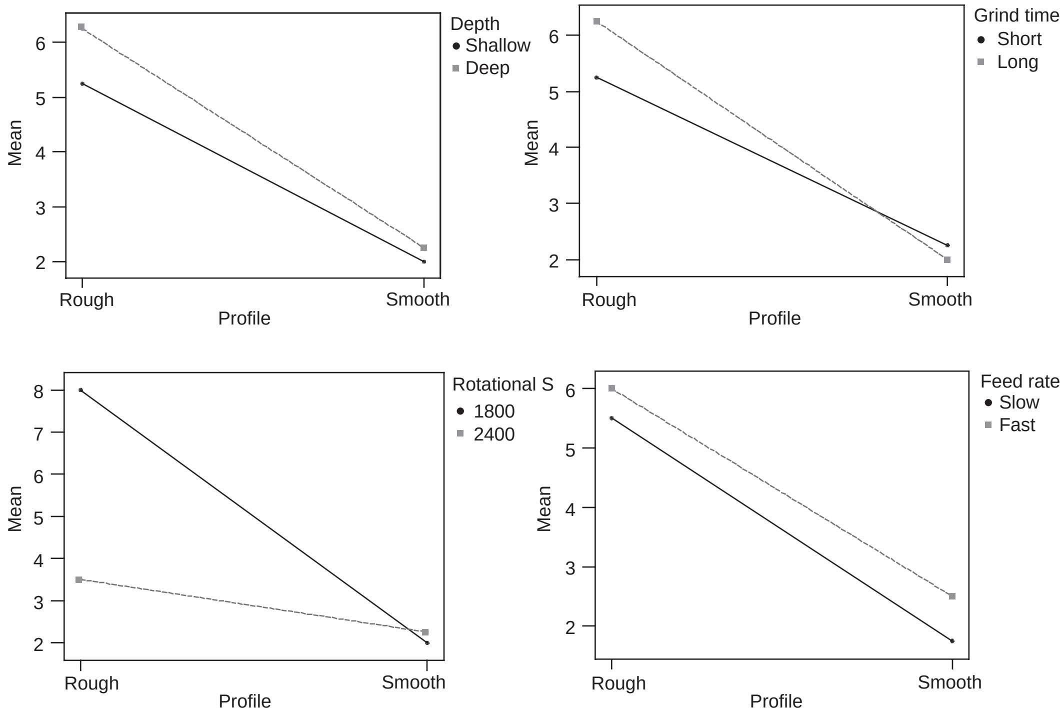
```
Alias Structure (up to order 3)

I
Depth + Grind*Rotation*Feed
Grind + Depth*Rotation*Feed
Rotation + Depth*Grind*Feed
Feed + Depth*Grind*Rotation
Profile
Depth*Grind + Rotation*Feed
Depth*Rotation + Grind*Feed
Depth*Feed + Grind*Rotation
Depth*Profile
Grind*Profile
Rotation*Profile
Feed*Profile
Depth*Grind*Profile + Rotation*Feed*Profile
Depth*Rotation*Profile + Grind*Feed*Profile
Depth*Feed*Profile + Grind*Rotation*Profile
```

The design is resolution IV in the candidates; hence, in the crossed design we can separately estimate the interactions between each candidate and the premachined profile (the dominant cause). For example, in the alias structure we see the interaction depth by profile is in a row by itself.

b. We start the analysis by plotting smoothness by treatment with a different plotting symbol for each profile. Treatments 5 to 8, all with high rotational speed, look promising.



The interaction plots for profile versus each of the four candidates follow. Only rotational speed flattens the relationship between smoothness and the initial shaft profile.



We conclude that setting the rotational speed to 2400 will desensitize smoothness to changes in the initial shaft profile.

c. With the existing experiment there is a risk that the treatment that results in a smoother commutator surface may not translate into better performance in the accelerated test. However, the chance of this occurring is small, since the team verified that the commutator surface unevenness was a dominant cause of electric motor failures. Using the accelerated test for the experiment would have been more expensive.

CHAPTER 17

17.1 In an investigation, 100 trucks were selected from regular production over two weeks. The frame geometry as given by four summaries (left and right front, left and right rear) and the alignment outputs left and right camber and caster were determined for all 100 trucks. The data are given in the file *truck pull feedforward*. In Chapter 17 an analysis determined that feedforward control based on frame geometry was feasible for left caster. Repeat the analysis for the other outputs: right caster, left camber, and right camber.

```
A numerical summary of the three outputs is
Variable          N       Mean      Median     TrMean      StDev    SE Mean
right caster    100     4.0106      3.9645     4.0213     0.8298     0.0830
left camber     100     0.3922      0.4130     0.3970     0.5077     0.0508
right camber    100     0.4811      0.4160     0.4836     0.4146     0.0415

Variable         Minimum    Maximum         Q1         Q3
right caster      1.7820     5.8710     3.5450     4.5588
left camber      -0.8490     1.4890     0.0078     0.7065
right camber     -0.5860     1.3650     0.2232     0.7630
```

The results from fitting a linear regression model for each of the three outputs with all four possible inputs are:

```
The regression equation is
right caster =  -19.0 + 0.557 left front + 1.07 right front + 0.218 left rear +
  0.399 right rear

Predictor        Coef     SE Coef          T        P
Constant     -19.0241      0.6798     -27.99    0.000
left front    0.55657     0.04478      12.43    0.000
right front   1.06849     0.03994      26.75    0.000
left rear     0.21845     0.03079       7.09    0.000
right rear    0.39874     0.04638       8.60    0.000

S = 0.1962      R-Sq = 94.6%     R-Sq(adj) = 94.4%

The regression equation is
left camber =  -12.3 + 0.609 left front + 0.360 right front + 0.172 left rear +
  0.113 right rear

Predictor        Coef     SE Coef          T        P
Constant     -12.3225      0.4321     -28.52    0.000
left front    0.60906     0.02846      21.40    0.000
right front   0.36025     0.02539      14.19    0.000
left rear     0.17192     0.01957       8.78    0.000
right rear    0.11300     0.02948       3.83    0.000

S = 0.1247      R-Sq = 94.2%     R-Sq(adj) = 94.0%
```

```
The regression equation is
right camber =  -10.2 + 0.344 left front + 0.556 right front + 0.0136 left rear
  + 0.121 right rear
```

Predictor	Coef	SE Coef	T	P
Constant	-10.2119	0.3540	-28.84	0.000
left front	0.34438	0.02332	14.77	0.000
right front	0.55571	0.02080	26.71	0.000
left rear	0.01364	0.01604	0.85	0.397
right rear	0.12114	0.02416	5.01	0.000

```
S = 0.1022      R-Sq = 94.2%     R-Sq(adj) = 93.9%
```

In each case, the residual variation (given by *s* in the regression results) is much smaller than the variation for the given output in the 100 trucks. We assume this closely matches the baseline variation since the data for the feedforward investigation were collected over a relatively long time. Feedforward control using the frame geometry has the potential to greatly reduce pull variation. As described in Chapter 17, the team was able to use a model to predict camber and caster based on the truck frame geometry inputs and compensate for the effect if necessary.

17.2 Engine assembly problems occurred due to a poor fit between the pistons and the engine bore. The dominant cause of poor fit was found to be variation in the clearance, the difference between the (minimum) bore diameter and the (maximum) piston diameter. To solve this problem, the team thought about using the feedforward (selective fitting) approach. The idea was to measure each piston diameter and place them into bins of similar diameter. Then, after each bore diameter was measured, a piston would be selected from the appropriate bin. To assess this proposal the diameter measurements for 469 pistons and bores, as measured from nominal, are given in the file *block bore diameter feedforward*. Quantify the expected reduction in clearance variation when using one (that is, no selective fitting), two, three, or four bins of pistons. A suggestion is to define the bins by dividing the range in piston and bore diameters (roughly –10 to 10 microns) into equal widths.

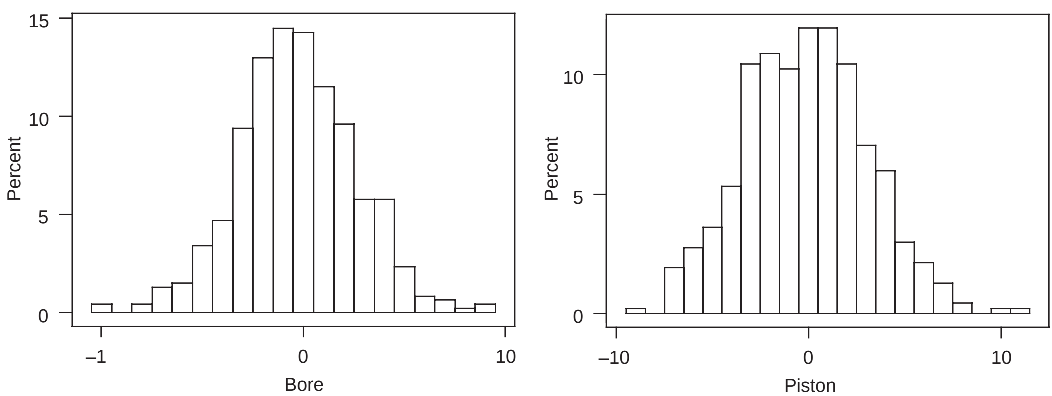
A numerical summary of the data is:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
piston	469	-0.124	0.000	-0.147	3.219	0.149
bore	469	-0.366	-0.500	-0.366	2.899	0.134
clearance	469	-0.242	0.000	-0.190	4.316	0.199

Variable	Minimum	Maximum	Q1	Q3
piston	-9.000	11.000	-2.000	2.000
bore	-10.100	9.200	-2.200	1.500
clearance	-15.100	12.800	-2.750	2.500

The range in piston diameters roughly matches –10 to 10, and the variation in bore and piston diameters is very similar. The standard deviation of the clearance for the given random pairing between piston and bore is 4.3. This provides an estimate of the process performance without selective fitting.

To quantify the expected variation in clearance using selective fitting we use simulation. We have two options: we can sample from the existing data to simulate the effect of selective fitting or we can model the piston and bore diameter variation and sample from the model. In this case, since both the piston and bore diameters have a bell-curve shape, as shown by the histograms that follow, they are well modeled by Gaussian random variables.



To implement the first simulation option where we sample from the existing data, we wrote a MINITAB macro that randomly selected a number of bores and corresponding pistons assuming a given number of bins, and calculated the clearance. See Appendix A for more information on writing MINITAB macros. The MINITAB macro that assessed the effect of using two bins, defined as pistons and bores with negative or positive diameter for this example, follows.

```
MACRO
#with the data file open call the MACRO from the command line as below
# %'macro2piston.txt' 'piston' 'bore'

macro2piston piston bore

mconstant I I2 cpiston borenow pnow test clearan
mcolumn piston bore borelist allvals temp

let cpiston=1
let  I2=1
let temp=1
let allvals=0
```

```
#repeat 50 times to get good estimate of the standard deviation
while I2<=50
Sample 200 bore borelist    #get a list of 200 bore diameters at random
let I=1
while I<=200
let borenow=borelist(I)   #look at the next bore

if borenow>0   #try to find a piston that is also bigger than nominal
  let test=0
  while test<=0
    let pnow=piston(cpiston)   #look at the next piston
    let cpiston=cpiston+1
    if cpiston>469  #end of list, start again
      let cpiston=1
      endif
    if pnow>0
      let test=1  #found appropriate match
      endif
    endwhile
  let clearan=borenow-pnow
  endif
if borenow<=0   #try to find a piston that is also smaller than nominal
  let test=0
  while test<=0
    let pnow=piston(cpiston)   #look at the next piston
    let cpiston=cpiston+1
    if cpiston>469   #end of list, start again
      let cpiston=1
      endif
    if pnow<=0
      let test=1   #found appropriate match
      endif
    endwhile
  let clearan=borenow-pnow
  endif
  if (I2=1) AND (I=1)
         let allvals=clearan
       else
    Stack allvals clearan allvals.   #store all the clearance values
  endif

  let I=I+1
  endwhile
let I2=I2+1
endwhile

let test = MEAN(allvals)
print test
Let test = STDEV(allvals)
print test
ENDMACRO
```

Alternatively, we can simulate choosing random bores and finding the matching pistons from the Gaussian models in some programming language. We next give Matlab (see http://www.mathworks.com) code that allows us to explore the effect of any number of bins.

```
function []=selfitkbins(k,num)
%simulate the effect of using k bins for the piston/bore selective %fitting
  example
%num equals the number of simulated bore/piston combinations

%the estimated distribution of piston and bore diameters are Gaussian %with
  mean and standard deviation
pm=-0.12; ps=3.2;
bm=-.37; bs=2.9;

%we use k bins of pistons, divide the natural range -10 to 10 for the
  %pistons into intervals of equal width
% a better choice would be to use bins of roughly equal frequency
%we use the same bins definition for both pistons and bores
%this works because here the distributions are close to the same

%bin intervals
bins=[-inf,-10+(20/k):(20/k):10-(20/k),inf];

bores=normrnd(bm,bs,num,1);   %generate a random sample of bore diameters
%generate more pistons so we have enough in each bin
pistons=normrnd(pm,ps,num*2,1);
%match the bores with the appropriate pistons
allclear=[];  %keep track of all the clearance values

for i=1:k,  %find the bores in each group
   bind=find(bores<bins(k+1));
   bind2=find(bores(bind)>=bins(k));  %all the bores that match the group
   pind=find(pistons<bins(k+1));
   pind2=find(pistons(pind)>=bins(k));  %all pistons in appropriate bin
   len=length(bind2);  %number of bores to match
   if len>0,  %there are some bores in the bin
      %find some pistons that would match
      temp=pistons(pind2);  %pistons that match
      clearance=bores(bind2)-temp(1:len);
      allclear=[allclear;clearance]; %keep track of all clearance values
   end;
end;
hist(allclear)
mean(allclear)
std(allclear)
```

Conducting the simulation using either method, the expected reduction in clearance variation is:

Number of bins	Expected clearance standard deviation
1	4.3
2	2.6
3	1.8
4	1.6

The benefits of selective fitting are large even for only two bins. The team needed to decide whether the substantial costs and logistical difficulties of implementing feedforward control warranted adopting the approach. To implement selective fitting, all piston and bore diameters must be measured. In addition, pistons need to be placed in the appropriate bin. Logistical problems can occur if one of the bins runs out. In this example, the team decided that while feedforward was technically feasible other approaches should be considered due to concerns about the high cost of implementing selective fitting.

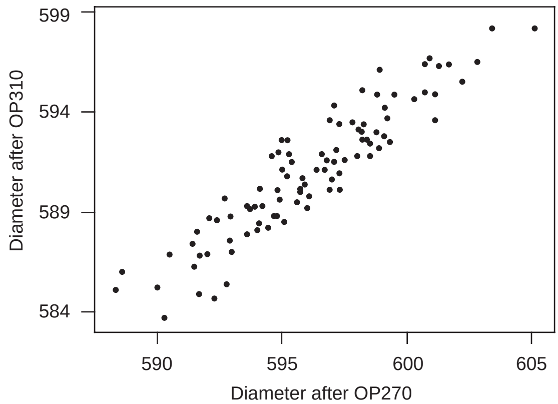
17.3 In the V6 piston diameter example discussed in Chapter 11, the team found that piston diameter after Operation 270 was a dominant cause of the final diameter. The data are given in the file *V6 piston diameter variation transmission*. This suggested that feedforward control might be a feasible approach.

a. What are the requirements for feedforward to be feasible in this context?
b. If feedforward were feasible, assess the potential benefit using the results of the variation transmission investigation.
c. Could the team also use the diameter after Operation 200, rather than the diameter after Operation 270, as the input to a feedforward controller?

a. For feedforward to be effective, there must be a way to quickly and cheaply measure the diameter after Operation 270 and then to adjust the process to compensate for large deviations between the predicted final diameter and the target. The prediction comes from the regression equation found in the analysis of the variation transmission investigation. Since there was no way to make the compensating adjustment, feedforward was not feasible in this example.
b. Using the results of the variation transmission investigation, we fit a regression equation relating the final diameter (after Operation 310) and the diameter after Operation 270. We get:

```
The regression equation is
diameter after OP310 = 64.3 + 0.884 diameter after OP270

Predictor        Coef     SE Coef          T        P
Constant        64.27       22.29       2.88    0.005
diameter      0.88358     0.03739      23.63    0.000

S = 1.224       R-Sq = 85.6%     R-Sq(adj) = 85.4%
```

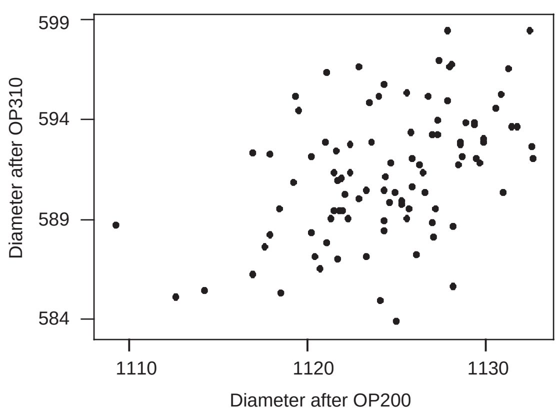
The strong relationship is also shown in the plot that follows:



Based on the regression model, using feedforward control and assuming perfect compensation would reduce the standard deviation in the final diameter to 1.2. This is a large reduction from the baseline standard deviation 3.32.

c. The diameter after Operation 200 is not a dominant cause of the final diameter variation. We can see this clearly in the regression results and plot given as follows:

```
The regression equation is
diameter after OP310 = 225 + 0.325 diameter after OP200

Predictor        Coef      SE Coef          T        P
Constant       225.12        72.72       3.10    0.003
diameter      0.32542      0.06467       5.03    0.000

S = 2.862       R-Sq = 21.2%     R-Sq(adj) = 20.4%
```



Feedforward control based on the diameter after Operation 200 would not work. The Operation 200 diameter does not provide a good prediction of the final diameter. The regression results suggest that if we implement perfect feedforward control based on the Operation 200 diameter, the final diameter variation could be reduced to 2.86 from 3.32 in the baseline. One advantage over using the Operation 270 diameter is that there is more opportunity to find an adjuster.

CHAPTER 18

18.1 The bias of the system used to measure camshaft journal diameters tended to increase over time. The cause of this increase was not determined. Instead, the team introduced a feedback controller. At the start of each shift, a master journal with known diameter was measured. If the measured diameter deviated from the known value by a large amount, the measurement system was recalibrated.

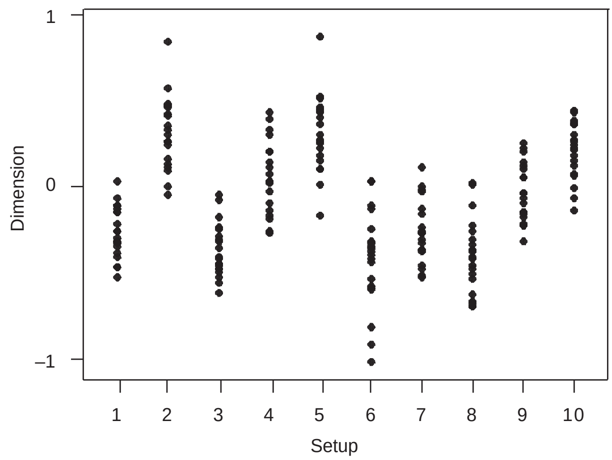
a. How could we sensibly define a large deviation in this context?
b. What would happen to the measurement variation if the measurement device were recalibrated every time the master journal was measured, rather than only when the deviation from the known dimension was large?

a. We define large by comparing the measurement error to the short-term measurement variation. For instance, we may decide to recalibrate if the difference between the measured value and the known value differ (in absolute value) by more two times the standard deviation of the measurement system over the short term.
b. If the change in measurement bias were slow relative to the short-term measurement variation, adjusting for any deviation from the known value would increase diameter variation. This is tampering as defined by Deming (1986).

18.2 In a machining process, the dominant cause of dimension variation acted in the setup family. That is, the dimension variation within a particular setup was small relative to the variation from one setup to the next. The existing control plan called for a complete process adjustment back to the target based on the first observation after each setup. There were more than 200 parts machined between setups. The baseline dimension standard deviation was 0.31. The team decided to explore a new feedback control scheme based on the average for the first five observations after each setup. In an offline investigation, they carried out 10 setups and produced 20 parts after each setup without any adjustment. The dimension data, scaled so that the target is zero, are given in the file *machining dimension feedback*.

a. Use a one-way ANOVA to estimate the standard deviation if the process could be adjusted so that the dimension averages across all setups were equal.
b. Use simulation to compare the performance of the existing feedback controller with the proposed controller that makes a complete adjustment based on the average for the first five observations after each setup.
c. In general, we may design a feedback controller by averaging the output from the first *n* observations after each setup. What considerations help you decide how many observations should be used to estimate the process average after each setup?

a. Plotting dimension by setup we get:



From the plot we can see the setup-to-setup variation. Using a one-way ANOVA we can partition the overall variation into the within-setup component and the variation between setups. The edited results of the ANOVA from MINITAB are:

```
One-way ANOVA: dimension versus setup

Analysis of Variance for dimension
Source     DF        SS        MS        F        P
setup       9   15.7998    1.7555    43.56    0.000
Error     190    7.6573    0.0403
Total     199   23.4572

Pooled StDev =   0.2008
```

The pooled standard deviation 0.2008 estimates the within-setup variation, the dimension standard deviation if we could adjust the process to keep the setup averages equal. This is substantially lower than the baseline value 0.31.

b. We can simulate the performance of the two adjustment schemes by:

Scheme 1: For each setup, subtract the first observation from the remaining 19 to model the adjustment. Find the standard deviation of the 20 ¥ 19 adjusted values to estimate the process performance.

Scheme 2: For each setup, calculate the average of the first five observations and subtract the average from the remaining 15 to model the adjustment. Find the standard deviation of the 20 ¥ 15 adjusted values to estimate the process performance.

We get:

Adjustment scheme	Estimated process standard deviation
None	0.34
First observation	0.29
Average of first five	0.22

By using the average of five parts, we can nearly reach the estimated minimum value 0.20. These comparisons are valid if the number of parts produced between setups is much greater than 20 pieces. If the number of parts was only 20 pieces, we should include the unadjusted parts (first one or first five parts) in the estimated standard deviation.

c. We need to ask questions like:

How much variation is there within each setup? If there is little variation, we need fewer observations.

How many parts are machined between setups? If there are only a small number of parts (or if the cost of poor parts is high), we may wish to make an initial adjustment based on a small number of observations and then adjust again once a few more observations are available.

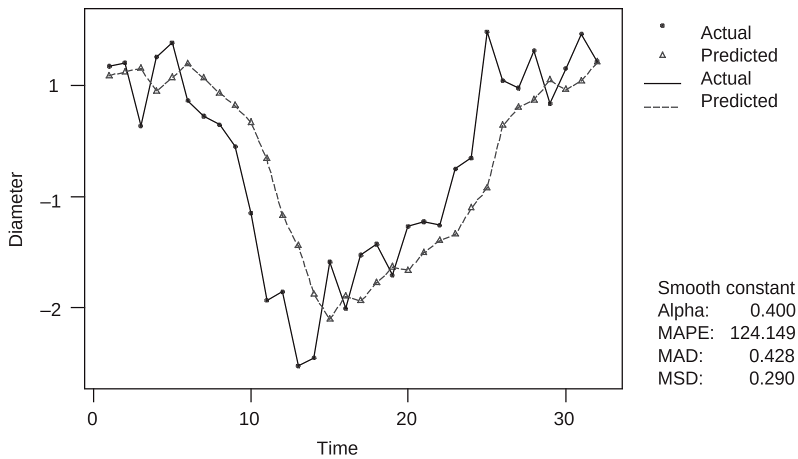
What are the adjustment costs? If the adjustment cost is large, we may decide to adjust only if the deviation from target is large.

18.3 In a machining process, there was excess variation in the diameter of a precision ground shaft. The shaft diameter was measured for all shafts using a complex automated gage (that also measured other outputs). Upon investigation, the team discovered that the dominant cause acted in the measurement family. In particular, the measurement bias changed from day to day, consistent with the pattern observed in the baseline. To explore this bias change further the team planned an investigation where the diameter of the same shaft was measured each hour for four days. A total of 32 diameter measurements were made. The data are given in the file *precision shaft diameter feedback*, with the output being the diameter measured from nominal. The results show a gradual drift. The team speculated that the drift was caused by changes in some (unidentified) environmental conditions. They decided to reduce the measurement variation using a feedback controller.

a. What type of feedback controller (that is, what prediction equation and what adjustment rule) would you recommend in this application?

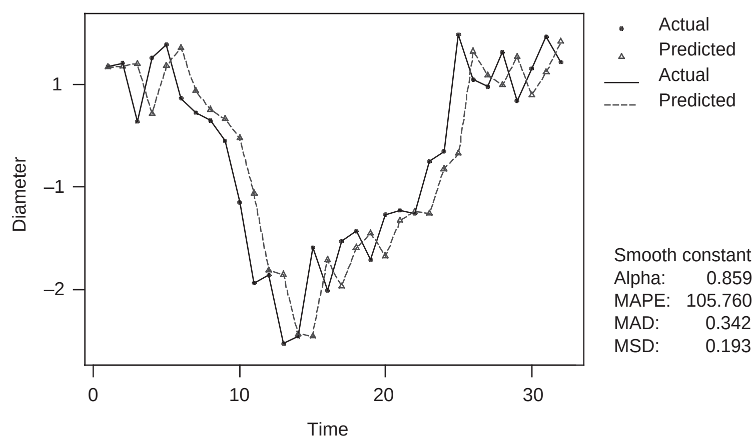
b. Suppose the team decided to use a feedback controller based on EWMA forecasts with the smoothing parameter alpha equal to 0.4. What kind of a reduction in the measurement variation could they expect?

a. The change in the diameter is gradual (that is, does not have sudden shifts). As a result, a feedback controller based on the EWMA forecasts is reasonable. The adjustments should be made so that if the one-step-ahead forecast is correct and the adjustment is perfect, the next value will be zero (that is, diameter equals the nominal value).
b. Using simple exponential smoothing with alpha equal to 0.4 gives:



From these results, we predict that with the feedback controller, the measurement standard deviation would be reduced to roughly 0.54 ($\sqrt{0.29}$). Note that given the small sample size, we are not confident of this prediction.

If instead we try to optimize the value of alpha, MINITAB gives:



However, because of the small sample size, we are not confident that alpha = 0.86 would be substantially better than alpha equal to 0.4 when applied to the actual process.

CHAPTER 19

19.1 In the paint film build example introduced in Chapter 3, the baseline standard deviation in film build (paint thickness) was 0.67 thousandths of an inch. With this variation, to ensure a minimum film build of 15-thousandths of an inch, the process was centered at 17. The goal was to reduce the standard deviation to 0.35, thereby allowing for a reduction in the average film build.

The dominant cause of film build variation was found using a multivari investigation to act in the car-to-car family. Despite further effort, the dominant cause was not found. The team decided to adopt the process robustness approach. Based on process experience, candidates and their corresponding levels were chosen as follows:

Candidate	Low level	High level
Anode dimension	3.1	3.9
Conductivity of paint	Low	High
Temperature	30	50
Zone X voltage	450	475
Zone Y voltage	500	525

The team selected a fractional factorial resolution V experiment with the 16 treatments given as follows. To reduce the cost of the experiment, panels were used rather than cars. With this choice there was a risk of study error.

Treatment	Anode dimension	Conductivity of paint	Temperature	X voltage	Z voltage
1	3.1	Low	30	450	500
2	3.9	Low	30	450	525
3	3.1	High	30	450	525
4	3.9	High	30	450	500
5	3.1	Low	50	450	525
6	3.9	Low	50	450	500
7	3.1	High	50	450	500
8	3.9	High	50	450	525
9	3.1	Low	30	475	525
10	3.9	Low	30	475	500
11	3.1	High	30	475	500
12	3.9	High	30	475	525
13	3.1	Low	50	475	500
14	3.9	Low	50	475	525
15	3.1	High	50	475	525
16	3.9	High	50	475	500

For each run, five panels were painted. The order of the treatments was randomized. Since the dominant cause acted car to car, the team believed the unknown dominant cause would act within each run. Film build was measured at five locations on each panel. The data for one location are given in the file *paint film build robustness* and in the table that follows.

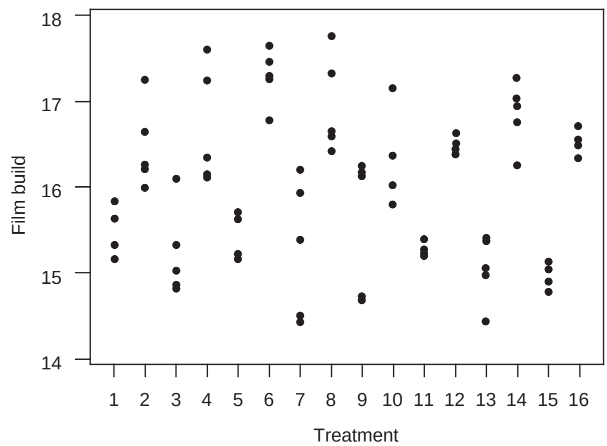
Treatment	Order	Film build	Average	Log(s)
1	14	15.6, 15.3, 15.9, 15.2, 15.8	15.56	–0.51
2	5	16.0, 16.3, 17.3, 16.2, 16.6	16.47	–0.31
3	6	15.0, 14.8, 14.9, 15.3, 16.1	15.22	–0.28
4	2	16.1, 17.6, 17.2, 16.3, 16.1	16.69	–0.16
5	9	15.7, 15.6, 15.2, 15.2, 15.7	15.49	–0.57
6	12	17.3, 17.6, 16.8, 17.5, 17.3	17.28	–0.49
7	13	16.2, 14.4, 15.4, 14.5, 15.9	15.30	–0.09
8	4	17.3, 16.6, 16.6, 16.4, 17.8	16.94	–0.25
9	7	16.1, 14.7, 16.2, 14.7, 16.2	15.59	–0.09
10	16	17.2, 15.8, 16.4, 16.0, 15.8	16.23	–0.24
11	15	15.4, 15.2, 15.4, 15.3, 15.2	15.29	–1.06
12	1	16.6, 16.4, 16.4, 16.5, 16.4	16.48	–1.00
13	3	15.1, 15.4, 15.4, 15.0, 14.4	15.05	–0.41
14	10	16.8, 16.9, 17.0, 17.3, 16.3	16.89	–0.42
15	11	15.0, 15.1, 15.0, 14.9, 14.8	14.97	–0.86
16	8	16.6, 16.7, 16.3, 16.5, 16.3	16.48	–0.79

a. Analyze the data using the standard deviation of film build over the five consecutive panels to measure performance. Is it possible to make the process robust to noise variation? What levels of the candidates do you suggest?

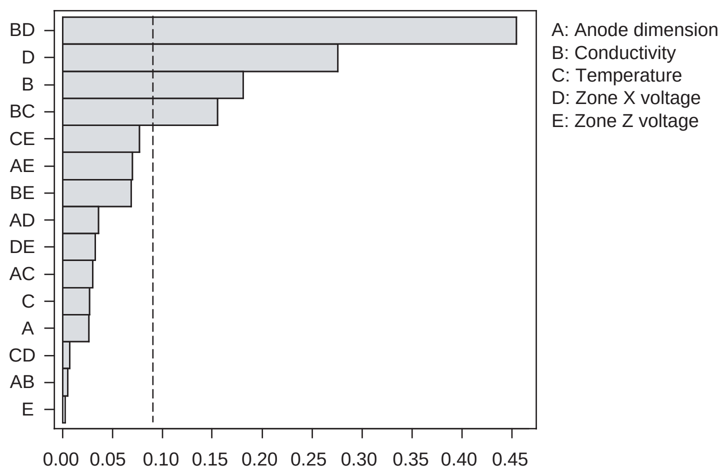
b. The team had a way to adjust the process center. However, we can also use the robustness experiment to look for an adjuster. Analyze the data using the average film build over the five consecutive panels to measure performance. Are any of the candidates adjusters?

c. In the experiment, the film build at a particular location on five consecutive cars (panels) was used to define a run. Suppose, instead, that the five observations came from five fixed locations on a single door. What, if any, changes are needed in the analysis presented in part a?

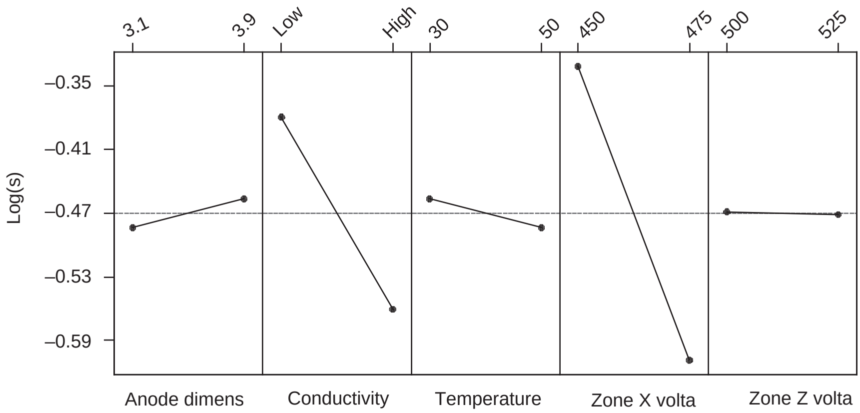
a. Plotting the film build by treatment suggests that treatments 11, 12, 15, and 16 are promising.

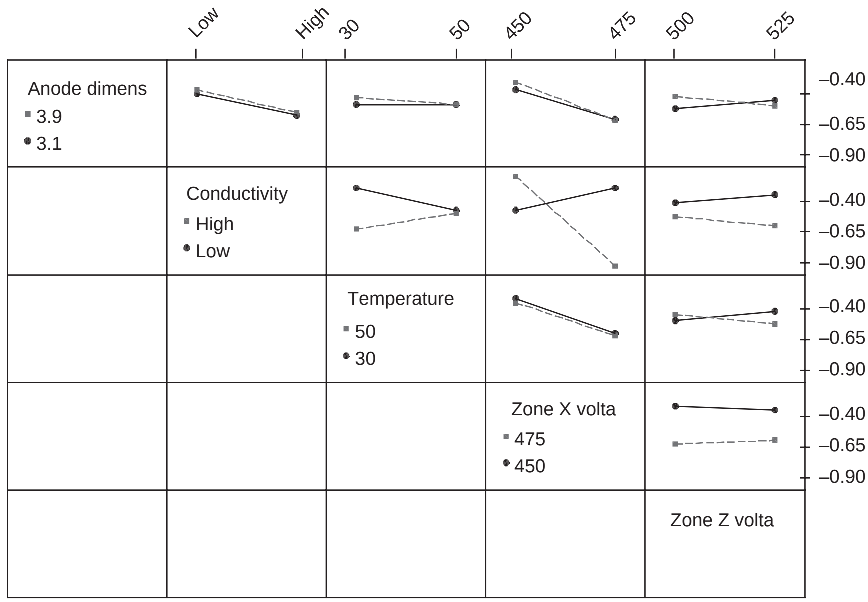


Using log standard deviation as the response, in MINITAB we get a Pareto chart of the effects, given as follows, that shows there are large main effects due to conductivity, zone *X* voltage, and large interactions between conductivity and zone *X* voltage and between conductivity and temperature.



To draw conclusions we also look at the following main effects and interaction plots.

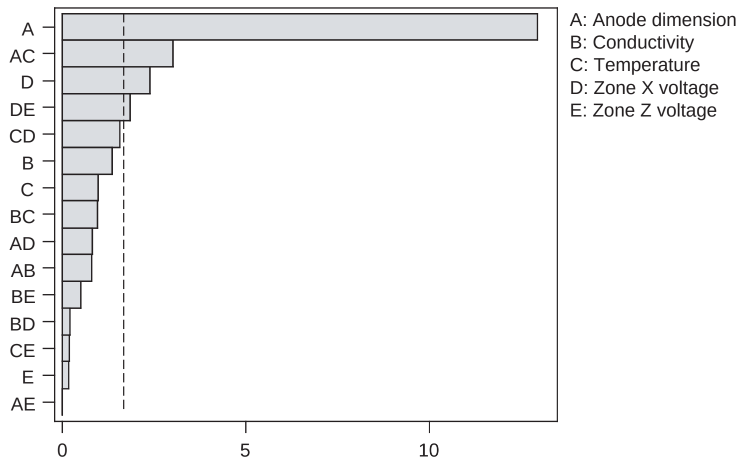




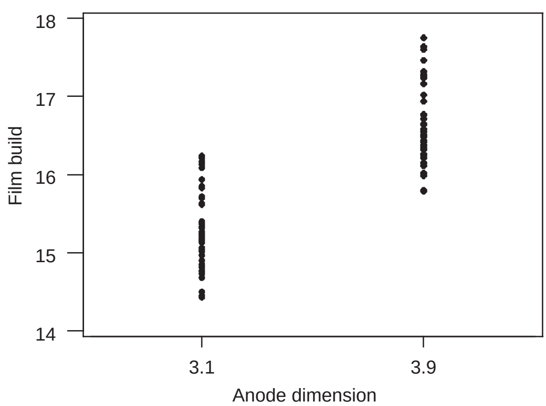
Recall that smaller log(s) is better. From the interaction plots, the combination of high zone X voltage, high conductivity, and low temperature is best. We are fortunate that high conductivity is best in both the large interactions. The high level of zone X voltage and conductivity, and low level for temperature corresponds to treatments 11 and 12. Using the new process settings is expected to reduce the baseline standard deviation of 0.67 to 0.37, since the average log(s) for treatments 11 and 12 is –1.03 and $e^{-1.03} = 0.37$.

With the new process settings, the average film build is about 15.9. The team had a way to adjust the film build average. The reduction in variation suggested a reduction in the average film build to around 16.25 from the current 17. This would translate into a 4% reduction in paint volume used. After implementing this solution the dominant cause acted within a door. There is ongoing effort to address this source of variation.

b. Looking for an adjuster, we analyze the experimental results using film build (or average film build in each run) as the output. The Pareto plot of the effects is:



Input *A* (anode dimension) is an adjuster. From the plot of film build by anode dimension, given as follows, we see that lowering the average dimension lowers the average film build.



c. When measuring the same five locations on each door, we may expect a systematic difference between the door locations. In the analysis we would want to keep track of the location and use an analysis as in Chapter 16. Using a performance measure like log(s) is not advised.

19.2 In a trim plant, customer complaints about seat appearance prompted management to assign a team the task of reducing shirring variation. The team proceeded without using Statistical Engineering and made a number of mistakes. Seat cover shirring was scored on a scale of 1 to 6 using boundary samples by how much cloth was gathered by the parallel stitching. Shirring scores of 1 to 4 were acceptable with 1 being the best. Scores of 5 or 6 resulted from either too much or too little shirring. A review of historical data suggested that the observed shirring score over a week covered all six possible values. Next, the team informally checked the measurement system. They found the measurement system added little variation. The team decided not to look for a dominant cause. Rather they moved directly to assessing the feasibility of making the process robust. They used brainstorming to select six candidates with two levels each as follows:

Candidate	Low level	High level
Leather thickness	0.8	1.2
Leather toughness	Pliable (soft)	Stiff (tough)
Seam width	9 mm	11 mm
Material feed	Top up	Bottom up
Steam to skin bun	Used	Not used
Bun thickness	+5 mm	–5 mm

The team planned a resolution III fractional factorial experiment with 16 runs (one for each treatment) as follows:

Treatment	Leather thickness	Seam width	Leather toughness	Machine feed	Steam	Bun thickness
1	High	Low	Tough	Top up	Yes	High
2	High	Low	Soft	Bottom up	No	Low
3	Low	High	Tough	Top up	No	Low
4	Low	High	Soft	Bottom up	Yes	High
5	High	High	Tough	Bottom up	Yes	Low
6	High	High	Soft	Top up	No	High
7	Low	Low	Tough	Bottom up	No	High
8	Low	Low	Soft	Top up	Yes	Low
9	High	High	Tough	Bottom up	No	Low
10	High	High	Soft	Top up	Yes	High
11	Low	Low	Tough	Bottom up	Yes	High
12	Low	Low	Soft	Top up	No	Low
13	High	Low	Tough	Top up	No	High
14	High	Low	Soft	Bottom up	Yes	Low
15	Low	High	Tough	Top up	Yes	Low
16	Low	High	Soft	Bottom up	No	High

Each run consisted of three seats (repeats). The runs were conducted in the treatment order given in the table. The data are given in the file *seat cover shirring robustness* and reproduced as follows:

Treatment	Order	Seat 1	Seat 2	Seat 3	Average score
1	13	3	1	2	2.0
2	16	1	2	1	1.3
3	7	2	2	2	2.0
4	6	2	2	2	2.0
5	10	2	1	1	1.3
6	1	3	1	3	2.3
7	11	4	2	1	2.3
8	15	2	2	4	2.7
9	5	1	2	2	1.7
10	3	4	5	2	3.7
11	14	3	3	2	2.7
12	9	2	3	3	2.7
13	8	1	2	2	1.7
14	2	2	2	3	2.3
15	4	1	4	2	2.3
16	12	2	3	1	2.0

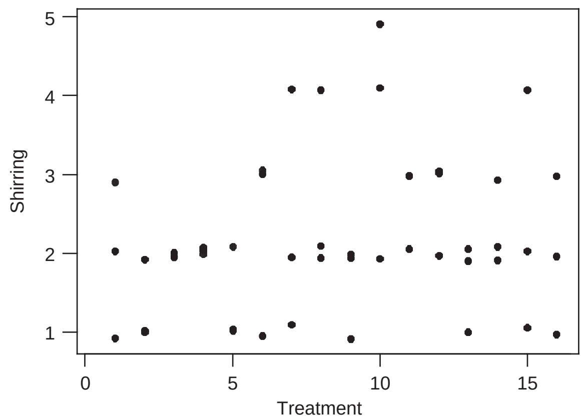
a. Explain why choosing the process output as a measure of variation (that is, high scores come from either too much or too little shirring) was a poor one.
b. The goal is to find process settings that lower the average shirring score. Can we use any of the candidates to achieve the goal?
c. Each run consisted of three seats. Discuss this choice in the context of a robustness experiment.

For the last two parts of this question, suppose the first three candidates (leather thickness, leather toughness, and seam width) used in the robustness experiment were normally varying rather than fixed inputs.

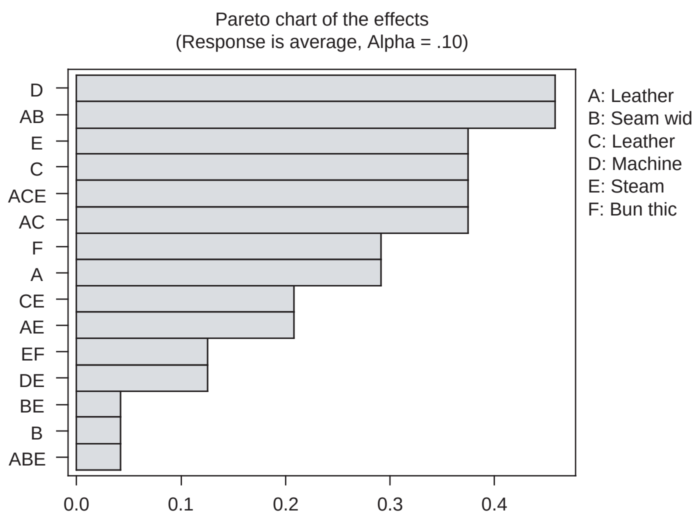
d. How should the levels of the first three inputs have been chosen?
e. Discuss changes you would make to the analysis you conducted in part b.

a. We should avoid defining the output in terms of a measure of variation if possible. With the shirring score as defined, it is difficult to find a dominant cause. Low or high values of the dominant cause both lead to high scores. The team was forced into the Make the Process Robust approach. A better scoring system would have given low scores for too little shirring and high scores for too much.

b. We first plot the shirring values versus the treatment number. In the plot that follows, we have added some jitter in the vertical direction (see Appendix C) to eliminate the problem of overplotting.



There are some promising treatments that result in a low average shirring score. Next, we fit a model with all possible terms and present the results using a Pareto plot of the effects.



There are no large effects. We may be tempted to choose one of the treatments that resulted in a low average shirring score as the new process settings. However, consider the answer to part c.

As an aside the particular fractional factorial design selected has the confounding structure given as follows. The inputs are labeled *A, B,* and so on, using the same order as in the preceding tables. The design is resolution III. This is a poor choice, since with 16 runs in six candidates a resolution IV design is possible.

```
Alias Structure (up to order 3)

I - A*D*F - B*C*F
A - D*F + B*C*D
B - C*F + A*C*D
C - B*F + A*B*D
D - A*F + A*B*C
E
F - A*D - B*C
A*B + C*D - A*C*F - B*D*F
A*C + B*D - A*B*F - C*D*F
A*E - D*E*F
B*E - C*E*F
C*E - B*E*F
D*E - A*E*F
E*F - A*D*E - B*C*E
A*B*E + C*D*E
A*C*E + B*D*E
```

c. In a robustness experiment, we do not know the dominant cause, and we define a run over sufficient time so that the unknown dominant cause has time to act. In this example, the team did not know the time family of the dominant cause. It seems unlikely that three repeats (seats) for each run is sufficient. This means that the dominant cause has likely not acted in (most of) the treatments. With this choice of run, the experiment will not be able to identify process settings that are robust. The experiment was doomed to failure because of poor planning.

d. The levels for each suspect should be selected at the extremes of the normal range of values. This requires measuring the value of the three suspects over a sufficient time span to see the full range before planning the robustness (desensitization) experiment.

e. The goal of the experiment would change to finding settings of the last three inputs that result in a lower average shirring score. It no longer makes sense to find the best settings of the suspects, since they are varying in the normal process. We would refer to only the last three inputs as candidates. The appropriate performance measure would be the average shirring score across all runs (and repeats) with a given combination of the three candidates. There are now only eight treatments in terms of the candidates. For example, we would average the shirring scores across runs 1 and 10. There is still no guarantee that the experiment will provide useful results. We do not know if the three suspects include important, let alone dominant, causes. This sort of an experiment, where both candidates and suspects are mixed together, is a common mistake when teams know about designed experiments but do not follow a structured problem-solving algorithm.

19.3 Torsional rigidity of the weather stripping was the dominant cause of door assembly problems. Management set a goal of reducing standard deviation in torsional rigidity to 0.3. A baseline investigation found the variation in torsional rigidity was roughly 0.55 mm and that the dominant cause acted over the short

term and certainly within any half hour. The team looked briefly for a dominant cause of rigidity variation without success. Next, they planned a robustness experiment with four candidates at two levels each, chosen based on engineering judgment. The candidates and levels are:

Candidate	Low level (–1)	High level (+1)
Heat (pre)	100	700
Extruder RPM	22	26
Tension (pre)	1	5
Water flow	2	6

The team planned a full factorial experiment with 16 runs, one for each treatment. The correspondence between treatments and candidate levels is given in the table that follows.

Treatment	Heat	Extruder RPM	Tension	Water flow
1	–1	–1	–1	–1
2	–1	–1	–1	1
3	–1	–1	1	–1
4	–1	–1	1	1
5	–1	1	–1	–1
6	–1	1	–1	1
7	–1	1	1	–1
8	–1	1	1	1
9	1	–1	–1	–1
10	1	–1	–1	1
11	1	–1	1	–1
12	1	–1	1	1
13	1	1	–1	–1
14	1	1	–1	1
15	1	1	1	–1
16	1	1	1	1

Each run consisted of running the process for half an hour after the candidate levels had been reached. Within each run, 10 weather-strip samples were

selected spread out over the half hour. The order of the runs was randomized. The torsion rigidity of each of the 10 weather-strip samples for each treatment is given in columns s1 to s10 of the table that follows and in the file *weatherstrip torsional rigidity robustness*.

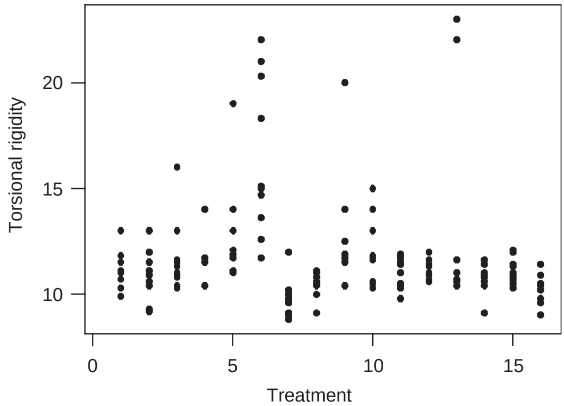
Treatment	Order	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
1	13	10.3	13.0	11.5	11.8	10.7	9.9	10.7	11.5	11.0	11.1
2	6	11.5	13.0	10.4	11.1	10.9	10.6	12.0	9.3	9.2	9.3
3	9	11.6	13.0	10.4	16.0	10.3	10.8	11.5	11.0	11.3	10.9
4	1	11.5	11.7	10.4	11.7	14.0	11.7	10.4	11.7	10.4	10.4
5	3	14.0	11.7	11.7	19.0	11.9	11.7	12.1	13.0	11.1	11.0
6	11	22.0	15.0	18.3	11.7	20.3	21.0	12.6	13.6	14.7	15.1
7	5	9.1	9.6	10.2	9.8	9.0	9.7	10.0	12.0	9.0	8.8
8	14	10.0	9.1	10.6	10.4	10.8	11.0	11.1	10.8	10.5	10.8
9	2	11.7	12.5	11.9	11.7	20.0	14.0	10.4	11.5	11.7	20.0
10	10	10.3	11.6	10.5	10.6	13.0	14.0	11.7	10.3	15.0	11.8
11	7	10.3	10.5	11.0	11.4	9.8	10.4	11.7	11.8	11.5	11.9
12	15	11.6	11.0	11.4	11.3	12.0	10.6	10.9	10.7	10.7	10.7
13	16	10.6	10.7	11.6	10.6	10.7	22.0	11.0	10.4	10.4	23.0
14	8	9.1	10.4	10.6	11.4	10.9	10.4	10.8	10.9	11.0	11.6
15	12	10.3	11.0	12.0	12.1	10.5	10.7	11.3	11.4	10.8	10.9
16	4	10.4	10.4	10.4	10.5	10.9	11.4	9.0	9.6	9.8	10.2

a. To analyze the results of this robustness experiment, what performance measure(s) do you recommend and why?

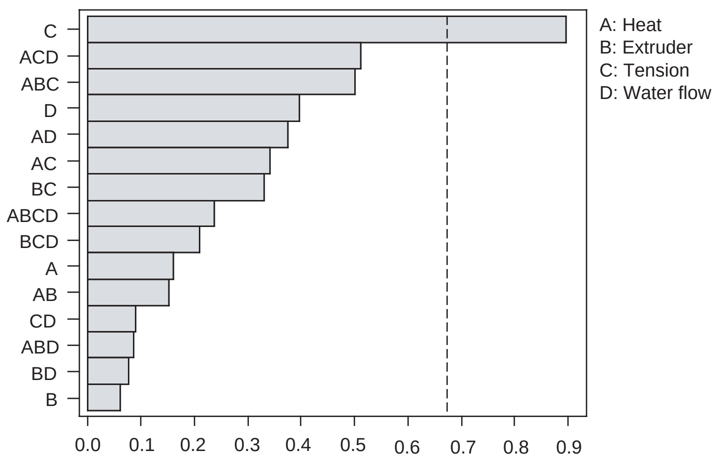
b. Analyze the experimental results using your chosen performance measure(s). What can you conclude?

a. The goal of the experiment is to find process settings that results in less variation in torsional rigidity (as the unknown dominant cause acts). The performance measure log(s), where s is the standard deviation of the rigidity values across the 10 repeats within each run, is appropriate. A secondary performance measure is the average rigidity across each run. If we find process settings that result in less rigidity variation, we may need to adjust the process center back to the target.

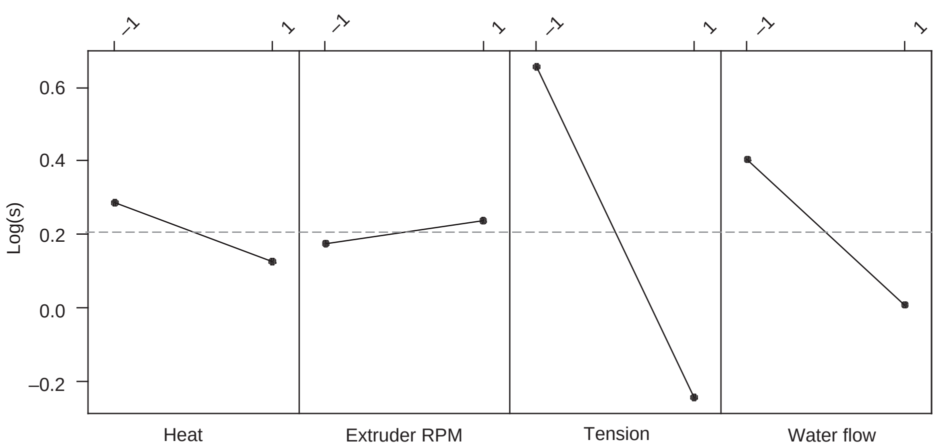
b. Plotting the torsional rigidity values by treatment, we get:



There are some treatments that have much less rigidity variation than others. Fitting a complete model (using the log(s) performance measure) a Pareto plot of the effects gives:



Candidate *C* (tension) has a large effect. Looking at the main effects plots suggests that the high level of tension gives substantially lower rigidity variation.



However, the average torsional rigidity variation at the high level of tension is 0.8 ($e^{-0.2}$), much higher than the baseline variation of 0.55. In fact, even the best treatment, number 12, has a variation of 0.47, only marginally better than the baseline. The experiment has identified a candidate (tension) that can be changed to *increase* rigidity variation. The experiment failed to find more robust process settings. The team was curious about this result and decided to investigate even higher tension levels. This investigation also failed since high tension led to other negative side effects and little reduction in rigidity variation. The approach was abandoned.

CHAPTER 20—NO EXERCISES

CHAPTER 21

21.1 Discuss whether lessons learned are properly maintained in corporate memory in your organization. What could be done to improve the situation?

The answer depends on the organization.

21.2 In the paint film build example described in Chapter 19, the team found new process settings that resulted in reduced car-to-car variation in film build. To validate the proposed solution, 80 cars were painted over one day with the settings given in the following table. These were the best settings found in the robustness investigation. The film build values from five specific positions on one door for each of the cars are available in the file *paint film build validation*.

Candidate	Setting
Anode dimension	3.5 (midpoint)
Conductivity of paint	High
Temperature	30
Zone X voltage	475
Zone Y voltage	500

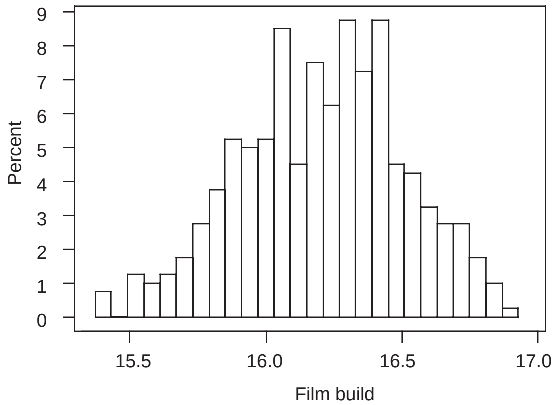
a. **The baseline film build standard deviation was 0.68. The problem goal was to reduce the standard deviation to 0.35, and the robustness experiment results suggested that changing settings would reduce the standard deviation to about 0.37. Has the solution been validated?**

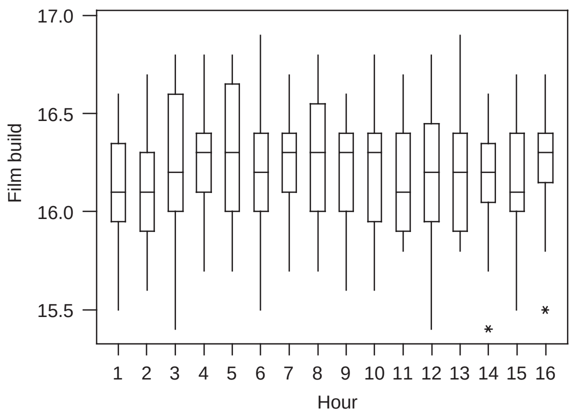
b. **What, if anything, do the validation results tell us about the home of the dominant cause in the remaining variation?**

a. We analyze the data as in the baseline investigation with the following numerical and graphical summaries.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
film build	400	16.208	16.222	16.212	0.302	0.015

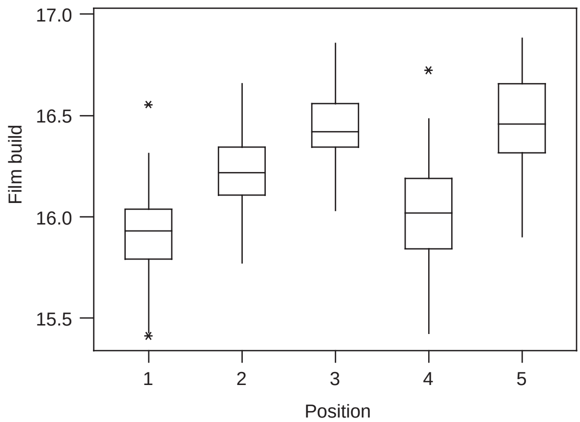
Variable	Minimum	Maximum	Q1	Q3
film build	15.411	16.880	15.987	16.422





The overall standard deviation has been reduced to 0.302, a substantial reduction from the baseline standard deviation 0.68, and exceeding both the problem goal of 0.35 and the expected improvement. The average film build is 16.2. Further efforts were made to bring the average to 16, a little more than 3 ¥ 0.30 above the minimum acceptable film build of 15 units. With these changes, the ultimate paint cost savings were about 6%.

b. As suggested by the plot that follows, the team could now look for further improvement by reducing the position-to-position variation in film build.



Looking at the results from a one-way ANOVA, we estimate that if we could eliminate the position-to-position variation, the film build variation would be further reduced to 0.21.

```
Analysis of Variance for film build
Source     DF         SS         MS        F        P
position    4    19.7020     4.9255   113.79    0.000
Error     395    17.0980     0.0433
Total     399    36.8000

Pooled StDev =    0.208
```

21.3 **In the truck pull example described in Chapter 17 and Exercise 17.1, a feedforward controller was implemented to compensate for the effect of truck-frame geometry on pull. After the feedforward system had been operating successfully for some time, management decided to review its operation. The four frame geometry measurements and left and right caster and camber were recorded for roughly a month of production consisting of over 6600 trucks. The data are given in the file *truck pull validation*.**

a. The standard deviations for caster and camber before implementation of the feedforward controller can be estimated from the 100-truck investigation described in Chapter 17. From the same investigation, the team predicted the possible reduction in standard deviation using a feedforward controller. A summary is given in the following table.

Characteristic	Baseline standard deviation	Predicted reduction in standard deviation
Left caster	0.90	0.18
Right caster	0.83	0.20
Left camber	0.51	0.13
Right camber	0.41	0.10

Do the results of the investigation validate the reduction in left and right caster variation due to the feedforward controller?

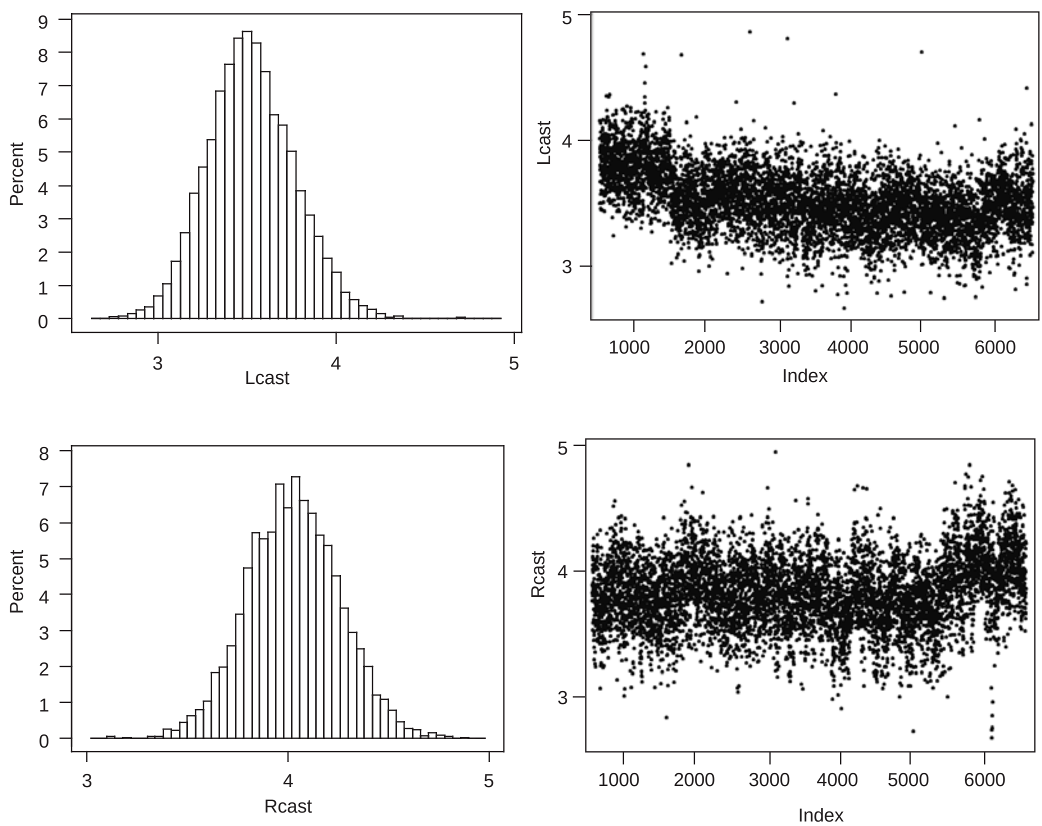
b. For each of the two caster characteristics, conduct a regression analysis to see if the feedforward controller can be improved. Recall that the feedforward controller should be compensating for variation in the frame geometry.

c. Repeat the analysis in parts a and b for left and right camber.

a. To see if we can validate the process improvement, we summarize the caster characteristics from the validation investigation numerically and graphically.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
lcast	6632	3.5225	3.5120	3.5193	0.2461	0.0030
rcast	6632	4.0273	4.0260	4.0269	0.2314	0.0028

Variable	Minimum	Maximum	Q1	Q3
lcast	2.6630	4.8560	3.3570	3.6830
rcast	3.0650	4.9640	3.8680	4.1820



The standard deviations for left and right caster (0.25 and 0.23, respectively) are much smaller than they were before implementation of the feedforward controller.

The reduction in standard deviations did not completely meet the predictions based on the overly optimistic assumption that perfect compensation was possible. The feedforward controller was a great success.

b. To check the operation of the feedforward controller for the caster characteristics, we fit a regression model for left caster (and right caster) as a function of the truck frame characteristics—left front, right front, left rear, and right rear. If the feedforward controller is working properly, there will be at most a weak relationship between caster and the frame geometry characteristics.

```
The regression equation is
lcast = 3.86 + 0.00623 left front + 0.0732 right front - 0.0213 left rear
           - 0.0978 right rear

Predictor        Coef     SE Coef          T        P
Constant      3.85986     0.09292      41.54    0.000
left fro     0.006230    0.004943       1.26    0.208
right fr     0.073226    0.004031      18.16    0.000
left rea    -0.021255    0.004744      -4.48    0.000
right re    -0.097787    0.005149     -18.99    0.000

S = 0.2351      R-Sq = 8.8%      R-Sq(adj) = 8.7%

The regression equation is
rcast = 3.76 - 0.0718 left front + 0.0840 right front + 0.00547 left rear
           + 0.00966 right rear

Predictor        Coef     SE Coef          T        P
Constant      3.75977     0.08599      43.73    0.000
left fro    -0.071805    0.004574     -15.70    0.000
right fr     0.084034    0.003731      22.53    0.000
left rea     0.005466    0.004390       1.25    0.213
right re     0.009664    0.004765       2.03    0.043

S = 0.2176      R-Sq = 11.7%     R-Sq(adj) = 11.6%
```

From the solution for part a, the standard deviations for left and right caster are 0.2461 and 0.2314 when using the feedforward controller. The corresponding residual standard deviations in the regression analysis are 0.2351 and 0.2176. The residual standard deviations estimate the process standard deviation if we could hold the frame geometry fixed or totally compensate for its effect. Because the residual standard deviations are so close to the caster standard deviations with the existing feedforward controller, the effects of frame geometry on the outputs are very small. In other words, allowing for the inevitable small measurement and adjustment errors, the existing feedforward controller cannot be improved further using the frame geometry characteristics.

c. The summary of the current process performance, in terms of camber, is

```
Variable        N       Mean     Median     TrMean      StDev    SE Mean
lcmb         6632    0.49288    0.49000    0.49239    0.16324    0.00200
rcmb         6632    0.47708    0.48200    0.47883    0.16222    0.00199

Variable        Minimum    Maximum         Q1         Q3
lcmb           -0.42600    1.45600    0.38200    0.60100
rcmb           -0.39300    2.05400    0.37600    0.58400
```

To check the operation of the feedforward controller for camber, we fit a separate regression model for left camber and right camber, as a function of the truck frame characteristics.

```
The regression equation is
lcmb = - 1.69 + 0.0761 left front + 0.00141 right front + 0.0727 left rear +
  0.0669 right rear

Predictor        Coef     SE Coef          T        P
Constant     -1.69356     0.05600     -30.24    0.000
left fro     0.076120    0.002979      25.55    0.000
right fr     0.001407    0.002430       0.58    0.563
left rea     0.072739    0.002859      25.44    0.000
right re     0.066909    0.003103      21.56    0.000

S = 0.1417      R-Sq = 24.7%     R-Sq(adj) = 24.7%

The regression equation is
rcmb = 0.761 - 0.0482 left front + 0.0192 right front - 0.0242 left rear
           + 0.0275 right rear

Predictor        Coef     SE Coef          T        P
Constant      0.76091     0.05994      12.69    0.000
left fro    -0.048213    0.003189     -15.12    0.000
right fr     0.019163    0.002601       7.37    0.000
left rea    -0.024170    0.003060      -7.90    0.000
right re     0.027540    0.003321       8.29    0.000

S = 0.1517      R-Sq = 12.7%     R-Sq(adj) = 12.6%
```

The current standard deviations for left and right camber are 0.1632 and 0.1622, respectively. The corresponding residual standard deviations from the regression analysis are 0.1417 and 0.1517. The team was surprised by the left camber results and found a small error in the mathematical model used to predict left camber from the frame geometry. This error was corrected; however, because camber has a small effect on pull, there was no noticeable reduction in the pull standard deviation.

Chapter Supplements

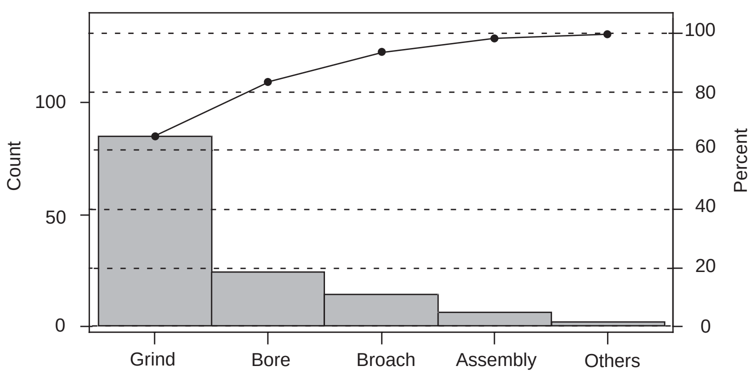
Chapter 2 Supplement

Describing Processes

S2.1 PARETO PRINCIPLE

The Pareto principle was introduced by Vilfedo Pareto (1848–1923) to describe the distribution of wealth. Pareto found that in most societies, the majority of the wealth was controlled by a small number of people. Over time, the principle found application in many areas. J. M. Juran was the first to apply the idea to manufacturing processes. He coined the terms *vital few* and *trivial many* to describe problems (Juran et al., 1979). The Pareto idea is sometimes summarized as the *80/20 rule,* meaning, for example in a manufacturing context, that 80% of the process problems are attributable to 20% of the process issues.

Pareto analysis is best presented in graphical form using a bar graph to illustrate the relative importance of the various categories. For example, Figure S2.1 shows a Pareto chart of one month's records of rod scrap data (discussed further in Chapter 6). The Pareto plot shows that 62% of the scrap was found at a grinding operation. This suggests focusing on grind scrap.



Defect	Grind	Bore	Broach	Assembly	Others
Count	85	24	14	6	2
Percent	64.9	18.3	10.7	4.6	1.5
Cum%	64.9	83.2	93.9	98.5	100.0

Figure S2.1 Pareto chart for scrap by operation for the rod line.

We use the Pareto principle extensively to choose projects and focus the problem. The Pareto principle implies that, for any problem, there are one or two dominant causes.

S2.2 DEFINING DOMINANT CAUSE

To define dominant cause we use the well-known conditional variance formula

$$Var(Y) = Var_x(E(Y|X)) + E(Var(Y|X))$$

See, for example, Kalbfleisch (1985). Replacing Y with the output and X with the input, we have

$$\mathrm{sd}(output) = \sqrt{\mathrm{sd}\left(E\left(output|input\right)\right)^2 + E\left(\mathrm{sd}\left(output|input\right)^2\right)}$$

The first term under the square root sign can be thought of as the variation in the output explained by the input. The second term is the residual variation that would remain if we could hold the input constant. The input is a dominant cause if the first term is large compared to the second.

Formally, we call a cause *dominant* if it explains more than half the variation. This implies the residual standard deviation must be less than $1/\sqrt{2} = 0.71$, or roughly 70% of the overall standard deviation, sd(*output*).

A formal derivation of dominant cause when the output is discrete or binary is difficult. For a binary output, we use an informal definition and say an input is dominant if for some level of the input the proportion defective is substantially reduced.

S2.3 DOMINANT CAUSE INVOLVING TWO (OR MORE) INPUTS

For many problems we can find a dominant cause that involves a single input. However, for some problems the situation is more complex. Possibilities include:

1. The dominant cause involves a single input.
2. There are two (or more) large causes (that is, no dominant cause).
3. The dominant cause involves two (or more) inputs.

Later, we see that it is important to understand the difference between these three cases, especially when considering possible solutions. We illustrate using a simple situation where the output varies between 0 and 10, and there are two inputs that we examine at two levels each (low and high, coded –1 and +1). Figure S2.2 shows the four values as circles and also plots the average of the two values at each the low and high levels of input 1 using a diamond.

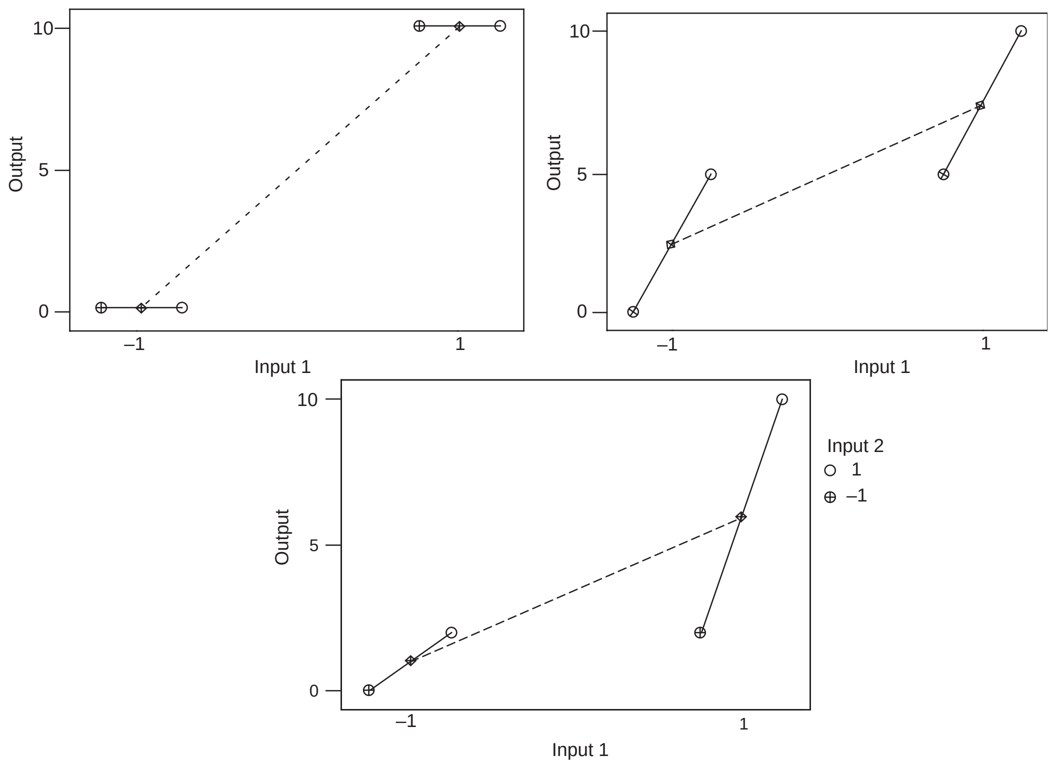


Figure S2.2 Plots illustrating the three cases for a dominant cause.

Case one is illustrated in the top left panel of Figure S2.2, where there is a single input (input 1) that explains the variation in the output. Here the second input has no effect on the output. In case two, as given by the top right panel of Figure S2.2, there are two large causes (inputs), but the effect of each input does not depend on the level of the other input (we say the effects of the two inputs on the output are additive). A dominant cause involving two inputs is presented in the bottom panel of Figure S2.2. This last case is the most complicated. We say a dominant cause involves two inputs if, for *both* inputs, we can observe close to the full extent of output values as the input changes for *some value of the other input*. In the plot, for the high level of input 1, changing the level of input 2 results in a large difference in the output, while similarly, for the high level of input 2, changing the level of input 1 results in a large difference in the output. There are gray areas among the three cases.

A dominant cause involving two or more inputs is related to the idea of an interaction. We say there is an *interaction* between input 1 and input 2 *in their effect on the output* if the relationship between input 1 and the output depends on the level input 2 and vice versa. Note that there can be an interaction between input 1 and 2 in their effect on the output even if there is no correlation between input 1 and 2 (that is, if input 1 and 2 vary independently of one another). We see the interaction between input 1 and input 2 clearly in the bottom panel of Figure S2.2, since the effect of changing from the low to high level of input 2 is much greater when input 1 is at its high level than when it is at its low level.

We return to the issue of the consequences of a dominant cause involving two (or more) inputs in Chapter 10 and its supplement, where we discuss finding such dominant causes, and in Chapter 14, where we consider possible solutions for problems with such a dominant cause.

S2.4 CLASSIFYING CAUSES OF VARIATION

There are several ways to classify causes of variation. For our purposes, we are interested in identifying inputs that make a large contribution to the variation of the output. We call these *dominant causes.* In the application of statistical process control (SPC), causes are classified as *common* or *special.* Taguchi (1986) uses the term *noise factor* and considers internal, external, and unit-to-unit noise. Here, we discuss the relationships among these classifications.

There is considerable confusion in SPC about the definition of special and common causes that are sometimes called assignable and chance causes. This classification depends on the control charting procedure in use and on how subgroup data are collected. A dominant cause may be either special or common. Suppose that subgroups for the control chart are based on five units produced consecutively by the process. If the dominant cause changes from part to part (that is, within the subgroups), it will be a common cause. If the dominant cause changes more slowly (that is, between subgroups), then it will be special.

We find that control charts are not useful in the identification of dominant causes. The main difficulties are that when the chart signals, this means that some input has changed but there is no clue given as to which input. The effect of the input is large with respect to the within-subgroup variation but may not be a dominant cause of the overall variation of the output. Furthermore, it is not predictable when the chart will signal and it is difficult to organize resources to look immediately for the cause. If the search is postponed, then the input may have changed again by the time it is scrutinized and the information about its effect will be lost.

Taguchi's noise factors are varying inputs that affect the output. In other words, noise factors are what we call *causes.* What Taguchi calls *control factors* are what we call *fixed inputs* that can be changed only by deliberate intervention. The use of the word *factor* comes from the language of experimental design, which plays a large role in Taguchi's variation reduction approach. He classifies causes as internal, external, and unit-to-unit. *Internal causes* are inputs that change over the life of the product. *External causes* are inputs that change in the usage environment, and *unit-to-unit causes* change from part to part at the time of production. In the refrigerator example discussed in Chapter 1, the identified external causes of frost buildup were ambient temperature and humidity, the amount of food added to the refrigerator, and the number of times the door was opened. An internal cause, not identified, was the deterioration of the door seals over time. The difference in compressor performance from one refrigerator to the next was a unit-to-unit cause. We discuss Taguchi's approach to variation reduction in chapters 16 and 19.

In general, we do not believe that using an experiment is an efficient way to *identify* a dominant cause of variation. Instead, we propose to use much simpler process investigations that do not involve changing fixed inputs or controlling varying inputs to search for such a cause. Within the Statistical Engineering algorithm, experimental design is needed to verify the contribution of a potential dominant cause, and in determining the feasibility of some of the variation reduction approaches.

S2.5 PROCESS CAPABILITY

There is an alphabet soup of capability ratios, such as P_{pk}, C_{pk}, and so on, that are used to quantify the relationship between the process center and variation, and the specification limits. See Kotz and Johnson (2002) and their discussion for a review of the merits and liabilities of various capability ratios. We give a brief discussion of P_{pk} since it is sometimes used to set the goal for a process improvement problem.

The index P_{pk} is defined as

$$P_{pk} = \frac{\text{distance from the process average to the closest specification limit}}{3 \times stdev}$$

where *stdev* is the standard deviation of the process output. Larger values of P_{pk} indicate smaller variation. The index is increased by better locating the average near the middle of the specification limits and/or by reducing the variation among the parts (that is, decreasing *stdev*). Note how P_{pk} is a function of both kinds of variation (off-target and part-to-part variation) and gets bigger as one or another kind of variation gets smaller.

For the angle error data discussed in the context of the camshaft lobe runout example in Chapter 2, the average and standard deviation are given as –21.3 and 71.5, respectively. The specification limits for angle error are ±400; thus the specification limit closest to the average is –400. Hence, the value of the capability index P_{pk} is

$$P_{pk} = \frac{-21.3 - (-400)}{3 \times 71.5} = 1.77$$

Such a large value of P_{pk} indicates that the closest specification limit is far from the average relative to the process variation, as described by three times the output standard deviation. Since the histogram for angle error data is bell shaped, this means that it is highly unlikely that any of the observed angle errors will be outside of the specification limits. This can be clearly seen in the histogram of angle error given in Figure S2.3, where we added dashed lines to show the specifications.

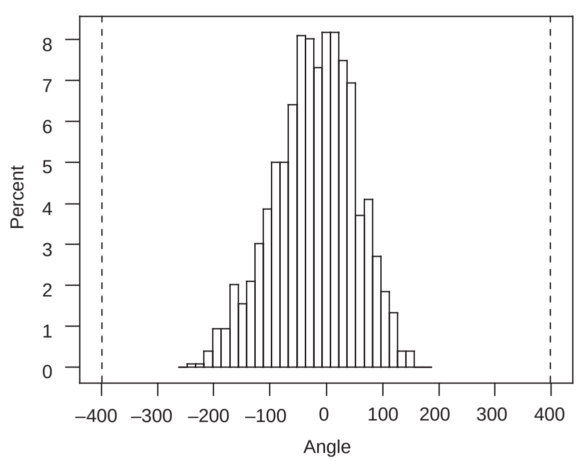


Figure S2.3 Histogram of angle error with specification limits.

Like all other summary measures of process performance, capability ratios depend on how the data are collected. If the sample is collected over a short time or when conditions are exceptional, the calculated capability ratio may be very misleading about the process performance in general.

S2.6 RELATING THE TWO KINDS OF VARIATION

We have described two kinds of variation, deviation of the average from the target and variation among parts as measured by the standard deviation. Suppose we have a set of n values for an output given by $y_1, y_2,..., y_n$. We provide a formula that relates the two kinds of variation to the variation from the process target T. The average and standard deviation are

$$avg = \frac{y_1 + ... + y_n}{n}, \quad stdev = \sqrt{\frac{(y_1 - avg)^2 + ... + (y_n - avg)^2}{n - 1}}$$

We define the root mean squared deviation (*RMSD*) of the data from the target by

$$RMSD = \sqrt{\frac{\left(y_1 - T\right)^2 + \left(y_2 - T\right)^2 + ... + \left(y_n - T\right)^2}{n}}$$

Note that *RMSD* is a measure of the average variation from the target. To relate *RMSD* to the two kinds of variation, we need the following result:

$$\left(y_1 - avg\right) + ... + \left(y_n - avg\right) = 0$$

That is, the sum of deviations of a set of numbers from their average is zero. We can show this result by noting that we can add the pieces before subtracting. That is,

$$\begin{aligned} \left(y_1 - avg\right) + ... + \left(y_n - avg\right) &= y_1 + ... + y_n - (avg + ... + avg) \\ &= n * avg - n * avg \\ &= 0 \end{aligned}$$

Now to decompose the *RMSD* into the two kinds of variation, we split each term in the numerator of the sum into three pieces.

$$\begin{aligned} (y_i - T)^2 &= [\left(y_i - avg\right) + \left(avg - T\right)]^2 \\ &= \left(y_i - avg\right)^2 + 2\left(y_i - avg\right)\left(avg - T\right) + \left(avg - T\right)^2 \end{aligned}$$

We can add the three pieces separately. The sum of the first pieces is

$$\left(y_1 - avg\right)^2 + \ldots + \left(y_n - avg\right)^2$$

Note that in the second and third pieces (*avg* – *T*) is a constant that is the same for all terms in the sum (i.e., it is the same for all *i*). The sum of the second pieces is

$$2\left(avg - T\right)\left[\left(y_1 - avg\right) + \ldots + \left(y_n - avg\right)\right] = 0$$

The third piece is the sum of *n* constants and is $n(avg - T)^2$. Combining the pieces we have

$$\begin{aligned} RMSD &= \sqrt{\frac{\left(y_1 - avg\right)^2 + \ldots + \left(y_n - avg\right)^2}{n} + \left(avg - T\right)^2} \\ &= \sqrt{\frac{n-1}{n} stdev^2 + \left(avg - T\right)^2} \\ &\approx \sqrt{stdev^2 + \left(avg - T\right)^2} \end{aligned}$$

In words, the squared average deviation from the target is approximately the square root of the sum of squares of the measures of the two kinds of variation. The relative contributions of the two kinds of variation can be assessed from this formula. For example, if the *stdev* is large and the *avg* is close to target, we can make very little gain by moving the process average closer to the target.

If the target *T* is an ideal value for the output, then there is likely to be some cost to the supplier or loss to the customer when the actual output deviates from the target. The *RMSD* has the property that it is approximately linear in the deviation from target $\left|y_i - T\right|$ when y_i is relatively far from *T* (relatively here means compared to the other output values). Close to the target, the contribution of $\left|y_i - T\right|$ to the *RMSD* is small. In other words, *RMSD* is a reasonable surrogate for the cost associated with a deviation from target.

S2.7 VARIATION WITHIN GROUPS AND GROUP TO GROUP

In the camshaft lobe runout example, we showed how aligning the average angle error for each lobe could reduce the overall standard deviation. Using the angle error data as an example, we give a general formula to connect the overall standard deviation to the variation group to group and within groups. Here a group is defined by the lobe position.

As in the previous section, we need a bit of algebra to demonstrate the decomposition. Recall that there are 108 camshafts, each with 12 lobes. Let y_{ij} represent the angle error for lobe position *j* (= 1, 2, ..., 12), and camshaft *i* (= 1, 2, ..., 108). Then the overall standard deviation is

$$stdev = \sqrt{\frac{\left(y_{1,1} - avg\right)^2 + \ldots + \left(y_{108,12} - avg\right)^2}{108 \times 12 - 1}}$$

where *avg* is the average of the 1296 values. For any term in the sum, if avg_j is the average of the angle errors at lobe position *j,* we can write the squared deviation from the average as three pieces:

$$\begin{aligned}(y_{ij} - avg)^2 &= \left[(y_{ij} - avg_j) + (avg_j - avg)\right]^2 \\ &= (y_{ij} - avg_j)^2 + 2(y_{ij} - avg_j)(avg_j - avg) + (avg_j - avg)^2\end{aligned}$$

For each lobe position, we can add the pieces separately. Note that the sum of the middle piece is zero because it is a constant times the sum of deviations from the *j*th position average. Looking at the sum for position *j,* we have

$$(y_{1,j} - avg)^2 + \ldots + (y_{108,j} - avg)^2 = \left[(y_{1,j} - avg_j)^2 + \ldots + (y_{108,j} - avg_j)^2\right] + 108(avg_j - avg)^2$$

In this equation, the first part of the sum on the right is directly related to the standard deviation of the output within position *j*, denoted $stdev_j$. That is,

$$(y_{1,j} - avg_j)^2 + \ldots + (y_{108,j} - avg_j)^2 = 107\,stdev_j^2$$

Hence, adding over all positions, we have

$$\begin{aligned}&(y_{1,1} - avg)^2 + \ldots + (y_{108,12} - avg)^2 \\ &= 107(stdev_1^2 + \ldots + stdev_{12}^2) + 108\left[(avg_1 - avg)^2 + \ldots + (avg_{12} - avg)^2\right] \\ &= 107\left(stdev_1^2 + \ldots + stdev_{12}^2\right) + 108 \times 11\, stdev(avg_1, \ldots, avg_{12})^2\end{aligned}$$

Dividing both sides by 1295 (= 108 × 12 – 1) and taking the square root we get

$$stdev = \sqrt{\frac{107\left[stdev_1^2 + \ldots + stdev_{12}^2\right]}{1295} + \frac{108 \times 11\, stdev(avg_1, \ldots, avg_{12})^2}{1295}}$$

$$\approx \sqrt{avg(stdev_1^2, \ldots, stdev_{12}^2) + stdev(avg_1, \ldots, avg_{12})^2}$$

In words, we see that the overall standard deviation is determined approximately by two pieces. The first is the average squared within-position standard deviation. The second is the squared standard deviation of the within-position averages. One way to reduce the overall standard deviation is to ensure the second term is zero. That is, we can make sure that there are no differences in the averages for the 12 lobes.

In the camshaft lobe runout example, the average and standard deviation of angle error for each position are:

Lobe	*avg*	*stdev*
1	17.76	61.16
2	45.31	67.62
3	–19.37	57.62
4	19.69	63.49
5	–30.92	59.16
6	–43.14	60.65
7	–20.58	60.03
8	–30.43	59.22
9	–28.61	61.10
10	–73.96	56.83
11	–1.46	67.42
12	–89.85	63.53

If we could control the process so that the average angle error was constant for each position, then the overall standard deviation would be reduced from 71.5 to

$$\sqrt{avg(stdev_1^2,\ldots,stdev_{12}^2)} = 61.6$$

In general, we can partition the overall standard deviation into variation within group and group to group. Aligning the group averages reduces the standard deviation by eliminating the variation due to group-to-group differences.

The division of the standard deviation into parts is the basis for the analysis of variance (ANOVA), a powerful numerical analysis method used in the book and discussed in Appendix D.

S2.8 GAUSSIAN MODEL

The Gaussian model is widely applicable. We describe some of its key properties. The Gaussian model describes the output of a process with a symmetric bell-shaped histogram. We repeat the form of the idealized curve in Figure S2.4.

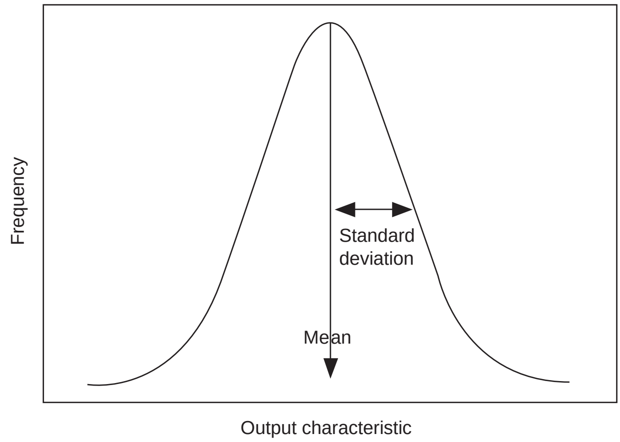


Figure S2.4 A Gaussian output model.

We prefer the name *Gaussian curve* to the more usual *Normal curve* because of the baggage that the word *normal* carries. We have been questioned many times by process engineers and production managers who look at the output from some statistical software that indicates that the process data is not "normal" and, hence, believe that something is wrong in the process. There is nothing abnormal or unusual about a process with a histogram of a different shape.

The area under the curve over any interval describes what proportion of the output values fall in that interval. We can specify the curve by two parameters, the mean (or center of the symmetric curve) and the standard deviation, a measure of the variation or spread of the values. The mean and standard deviation associated with the model are directly analogous to the *avg* and *stdev* of the histogram.

The Gaussian model has the beautiful property that it predicts that a fixed percentage of the output will fall within the interval (mean ± c standard deviation) for any constant c. For example, the percentages are approximately:

Within one standard deviation of the mean	68%
Within two standard deviations of the mean	95%
Within three standard deviations of the mean	almost 100%

There is a strong connection between the interpretation of capability measures such as P_{pk} and these percentages. If the process average is on target and we can describe the output by a Gaussian model, we can directly relate the magnitude of P_{pk} and the proportion of output that is out of specification:

P_{pk}	Parts per million out of specification
1.00	2700
1.33	64
1.67	0.6

These numbers should not be taken too seriously because a Gaussian model will never describe the process output perfectly.

In the book we rarely rely on formal statistical analysis. However, much of the formal analysis that we do present depends on the assumption that we can apply a Gaussian model. We often assume we can capture the effects of measured inputs in the mean of the model. The effects of all other varying inputs are lumped together in the standard deviation. For example, to formally analyze the position-to-position differences of the angle errors in the camshaft lobes, we assume that a Gaussian model can describe the output of the process where each position has a separate mean to explain the effect of position. We also assume that the variation within each position is the same since this variation captures the effect of all other varying inputs. Thus we assume that the standard deviation within each position is the same. We can describe the model pictorially using Figure S2.5 where we show the models for only three of the lobes to make it easier to interpret.

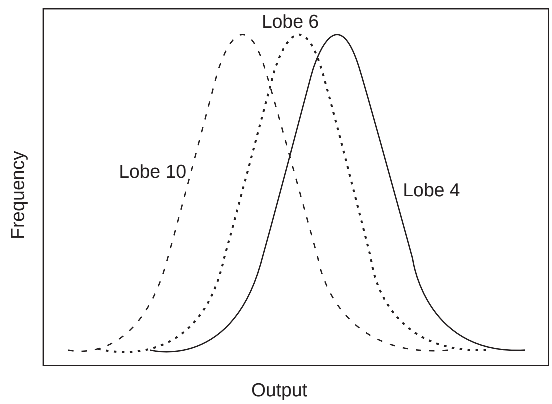


Figure S2.5 Graphical view of a model for lobe position differences.

Chapter 4 Supplement

An Algorithm for Reducing Variation

S4.1 COMPARISON OF VARIATION REDUCTION ALGORITHMS

There are many competing algorithms for solving chronic manufacturing problems and improving process performance. Some manufacturing organizations have invented or adapted an algorithm specialized to their own processes. Consulting companies have created their own versions, complete with acronyms and specialized language, driven partly by marketing considerations. We are perhaps also guilty of this sin. We have partially derived our algorithm from the one proposed by Shainin (1992, 1993).

Six Sigma (Harry and Schoeder, 2000) is a popular example. The algorithm is known by its acronym DMAIC (Define, Measure, Analyze, Improve, Control). The Define stage is sometimes omitted. The stages are divided into substages in a variety of ways that depend on which version you examine. See, for example, Breyfogle (1999).

The more detailed 12-step Breakthrough Cookbook associated with Six Sigma is described in Harry (1997, p. 21.19). A Six Sigma program is much more than the algorithm and the associated statistical methods and tools. It includes planning for implementation and ongoing management. For example, Six Sigma programs involve the training, certification, and deployment of specialized experts (Black Belts) in the use of the algorithm.

Scholtes (1988, pp. 5–19) provides a second example. The main stages of this algorithm are:

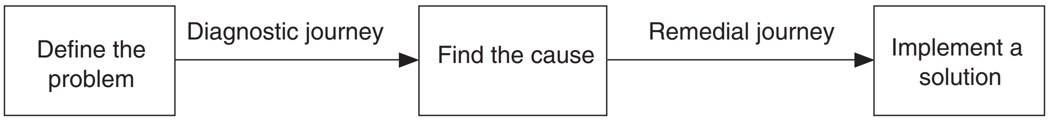
Understand the process → Eliminate errors → Remove slack → Reduce variation → Plan for continuous improvement

Each stage is further divided into a number of substages, used in completing the stage. Scholtes provides many other processes to support the algorithm, especially to help with the problem selection and the team operation.

Both the Six Sigma and Scholtes algorithms are designed to apply to processes in general and to many different kinds of problems. They are not specialized to reducing variation in high- to medium-volume manufacturing processes, as is Statistical Engineering.

In our view, these algorithms (including the one we propose) share two common elements:

- They are based on the diagnostic and remedial journeys described by Juran and Gryna (1980) and Juran (1988), as given in the figure that follows. The idea is that if we know the cause of the problem, we are more likely to find efficient and effective remedies or solutions.



- They make heavy use of empirical (statistical) methods to increase knowledge about process behavior, to identify causes, and to validate remedies.

The differences among the algorithms lie in the detailed elaboration of the three boxes in the diagnostic and remedial journey. For example, we have explicitly included a stage in which we assess the measurement system for the output to ensure that it is not the dominant cause. The DMAIC and Scholtes algorithms include this assessment within a stage and, hence, give it a lower profile. We have chosen to highlight measurement system assessment because, in our experience, there are often serious problems with the measurement systems in high- to medium-volume manufacturing processes.

The explicit focus on a dominant cause of variation, as defined in Chapter 2, is another important differences between the Statistical Engineering algorithm and the Six Sigma and Scholtes algorithms. We also emphasize the use of the method of elimination and families of variation to help find the cause in an economical and timely manner. We give more detail on families of variation and the method of elimination in Chapter 9.

One unique feature of the Statistical Engineering algorithm is the idea that the team should consider potential remedies and select a working variation reduction approach before deciding if they need to find the dominant cause of the variation. The algorithm allows for improvement without knowledge of a dominant cause. We think that this consideration will increase the efficiency of the process improvement and lead to better remedies.

See De Mast (2003 and 2004), De Mast et al. (2000), and Logothetis (1990) for a detailed methodological comparison of the Shainin, Six Sigma, and Taguchi strategies for quality improvement. Here we do not compare or contrast Statistical Engineering with the Taguchi three-phase program. See, for instance, Ross (1988) and the modification given in Taylor (1991). A direct comparison is difficult, since the Taguchi program is focused on the design of a new product or process, while the aim of Statistical Engineering is improvement in an existing process. Note, however, that two of the variation reduction approaches (desensitization and robustness—see chapters 16 and 19) use ideas from Taguchi's program.

Because the algorithms discussed here have a common basis, we believe that the choice of algorithm is far less important to success than is the organizational discipline required for routine implementation. The tools and strategies from one algorithm can, and in fact should, be incorporated into another as appropriate.

Chapter 5 Supplement

Obtaining Process Knowledge Empirically

S5.1 ATTRIBUTES

In applying QPDAC to plan investigations, we specify many numerical attributes other than averages, standard deviations, and proportions. Here we give some examples that can apply to the target and study population or the corresponding sample.

Stratification

The first complication is stratification. For example, we may have a process with distinct streams, such as two suppliers of the same component, three production teams (one for each shift), four different parallel operations within a machining operation, and so on. In the truck alignment process described in Chapter 1, there are four gages operating in parallel to measure caster and other alignment characteristics (see Figure S5.1).

We can stratify the output by gage and define attributes such as the average and standard deviation for each gage. If the averages are different, we can reduce the overall standard deviation by improving the calibration process for the gages. If the standard deviation within one gage is much larger than within the others, we can look within that gage to understand why it is behaving differently. We can estimate these attributes by applying QPDAC.

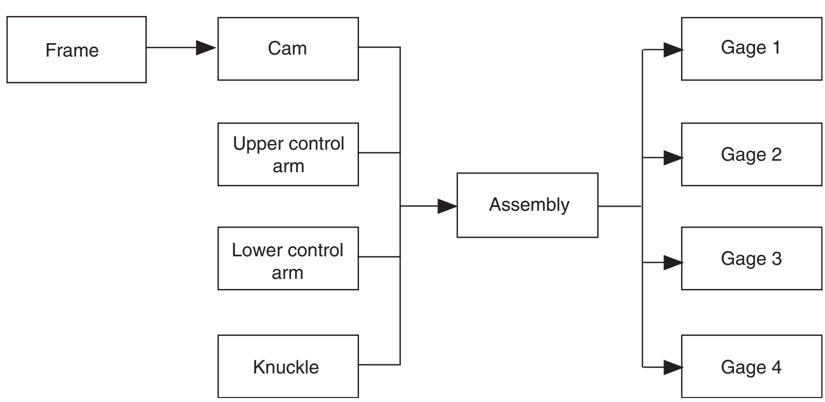


Figure S5.1 Truck alignment process.

We exploit the idea of stratification and specifying attributes by strata to generate clues about the dominant cause. See chapters 9 and 10.

Combined Attributes

Sometimes we express a question in terms of an attribute defined as a combination of two or more attributes. For example, in the supplement to Chapter 2, we defined the process performance measure P_{pk} as

$$P_{pk} = \frac{\text{distance from the process average to the closest specification limit}}{3^*\text{stdev}}$$

P_{pk} is a function of the population average and standard deviation. Another example is the capability index C_{pk}, where

$$C_{pk} = \frac{\text{distance from the process average to the closest specification limit}}{3^*\text{ within subgroup stdev}}$$

C_{pk} is a function of the average and a different attribute, the within-subgroup standard deviation.

To define within-subgroup standard deviation, we specify what we mean by a subgroup. We can choose as we please; often we define a subgroup as five consecutive parts from the process or all parts within a specified (short) period of time. In each case, we are *stratifying* the population of units into a large number of groups with no overlap. If all of the subgroups have the same size, then we define the within-subgroup standard deviation as

$$\text{within-subgroup } stdev = \sqrt{\text{average}(stdev_1^2, stdev_2^2, \ldots)}$$

where $stdev_1, stdev_2, \ldots$ are the standard deviations within subgroups 1, 2 ..., and so on. We average the squares and then take the square root to match the model behavior as discussed in Chapter 2.

We trust that you have noticed that P_{pk} and C_{pk} are different attributes because their denominators are not the same. Hence questions that we ask in terms of P_{pk} and C_{pk} are also different. We can see the connection between the denominators by noting the following result that we demonstrated in the supplement to Chapter 2.

$$stdev \approx \sqrt{(\text{within subgroup } stdev)^2 + stdev^2(\text{subgroup averages})}$$

Here we have stratified the population into subgroups. We have produced a very general formula, because the definition of a subgroup was up to us. We interpret the result by noting that the overall standard deviation is made up of a within-subgroup and a subgroup-to-subgroup component. In looking for a dominant cause of variation, we use this interpretation to eliminate from consideration causes associated with small components of the overall variation.

Note that the overall standard deviation and the within-subgroup standard deviation are almost equal when the standard deviation of the subgroup averages is small, that is, when all of the averages are about the same. In other words, the two attributes P_{pk} and C_{pk} will be almost the same when most of the variation in process output occurs within the subgroups, not subgroup to subgroup.

Scatter Plots and Fitted Lines

There is a set of useful numerical attributes associated with scatter plots. Suppose, in a population of units, we have an output characteristic y and an input characteristic x associated with each part. We do not expect y to be a simple function of x because there are many other inputs to the process. However, as shown in Figure S5.2, we often see a strong relationship in the scatter plot. In such cases, we can add a *fitted line* to describe the relationship numerically.

We fit the line using *least squares,* a procedure that minimizes the sum of squares of the vertical distances between the plotted points and the fitted line. Algebraically, if $y = a + bx$ represents any straight line, then we select the fitted line to minimize the quantity

$$\sum \left(y_i - a - bx_i\right)^2$$

where the sum is overall units in the population. We denote the fitted line in the population by the equation

$$y = \alpha + \beta x$$

The Greek letters α (alpha) and β (beta) are numerical attributes of the population. The slope β describes how much, on average, y will increase for a unit increase in x. Given a sample of parts from the study population, we estimate these attributes with the corresponding sample quantities. See Appendix E for more details on how to fit the line and create plots like Figure S5.2 for the data in the sample.

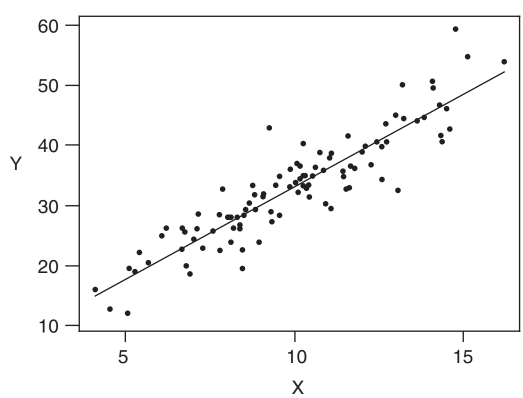


Figure S5.2 Scatter plot and fitted line.

Looking at Figure S5.2, we see that if x were held constant, the variation in y would be much smaller than when x is allowed to vary. For example, if x is held near 10, then from the scatterplot, we see that y varies from about 25 to 40. Overall, y varies from about 10 to 60. We use this idea to quantify the contribution of a single cause to the overall variation.

We can partition the standard deviation of y into two components:

$$stdev(y) = \sqrt{\beta^2 stdev(x)^2 + stdev(rest)^2}$$

The second component under the square root sign describes the variation that would remain if x were held fixed. We can picture this attribute on the scatterplot as the variation around the fitted line. The first component describes the contribution of the input x to the variation in the output y. If x is a dominant cause of variation in the output, *stdev*(*rest*) is small compared to *stdev*(y).

Given a sample from the population, we can estimate *stdev*(y) and the component contributions using MINITAB. Then we can decide if x is a major contributor to the variation in y.

S5.2 SAMPLING PROTOCOLS AND THEIR EFFECTS

In selecting the sample, the goal is to control:

- Sample error by matching the attributes of interest in the sample and the study population
- Sampling cost

There are some general considerations. First, to avoid sample error, we make sure that the sampling protocol will select units broadly across the study population. For example, if we define the study population to be one week's production, we should not entertain a sampling protocol that is limited to a single day. Second, if we choose a complex sampling scheme, we are likely to have added cost and a greater opportunity for the protocol to go wrong in the Data step.Third, we will generally incur a larger cost for larger samples, everything else being equal.

We make wide use of systematic sampling in which we deliberately select the sample over time and location. For example, we may select the next 300 units to assess process capability or we may select five consecutive pieces at the start of every two hours of production in a multivari investigation (see Chapter 11). Similarly, we may ensure that the sample has an equal number of parts from all process streams when assessing process performance. In assessing a measurement system, we select the parts to be measured repeatedly to cover the range of normal output values. We specifically sample both large and small parts. Systematic sampling has the major advantage that it is easy to organize and execute.

Random sampling is the most famous sampling protocol but has relatively few uses in process improvement. It is difficult and costly to implement. Suppose we want a sample of 100 parts selected at random from the study population, the week's production of 10,000 parts. We suppose that the units all have a unique identification code; if they do not, then we must assign such a code, which can be a daunting task. The first step is to associate each of the identification codes with a digit between 1 and 10,000. Then, using MINITAB, we generate a random sample of 100 values from the possible 10,000 numbers without replacement.

This is the easy part. Now we must find the parts corresponding to the sampled identification codes. This can be difficult if the process is complex and parts do not appear in the expected order. Finding the sampled units can take a long time and be expensive.

We once consulted on a project (by telephone only) where we naively recommended selecting a random sample of 50 water heaters from a lot of 2200 heaters. Each heater had a unique serial number. The heaters were stored in boxes in a warehouse. We provided a list of 50 serial numbers corresponding to a random sample of heaters. The sampling protocol fell apart immediately because it was not possible to locate the serial number without opening the box, and furthermore, it was very difficult to get at certain heaters due to the method of storage. In the end, a convenience sample was selected from locations spread throughout the warehouse. Note that when we discuss experimental plans in more detail, we will use random assignment extensively. This is a different use of randomization not to be confused with random sampling.

Sample size is a major issue. The most frequent question asked of a statistical consultant is, "How large a sample should I use?" The answer, like that to all statistical questions, is, "It depends!" For a given protocol, the larger the sample, the greater the chance of small sample error. However, even with a very large sample, we may be unlucky and have substantial sample error. To determine sample size, the first consideration is how the sampling fits into the rest of QPDAC. For instance, if there are likely to be large study and measurement errors, it makes little sense to try to get very small sample error. Second, there are cost and time constraints that often outweigh any consideration of sample error. We can often answer the sample size question based on what the team can afford.

There are some formal procedures that can be used to determine sample sizes if we have random sampling (or another sampling protocol where we pretend that the sampling will be random) and simple attributes in the Question step. We do not discuss these formally here. See Odeh and Fox (1975), Nelson (1985), and Neter et al. (1996). Also MINITAB has a "power and sample size" function. To understand how these methods work, consider the following example.

We were once asked how large a sample would be needed to determine if there were any defective rails in a suspect lot of 5700 rails. The defect was a vertical crack in the head of the rail that might cause derailment of a train. To detect the defect, the railway used a destructive test. The management wanted to be 99% confident that there were no defective rails in the lot. To answer the question, from a statistical perspective, the situation requiring the largest sample size would occur if there were exactly one defective rail in the lot. Suppose this was the case, that is, that there was one defective in the lot and we plan to choose a sample of n rails at random. We want to be 99% sure that we find the defective rail. Simple calculations show that we need to sample 99% of the rails. This was completely infeasible; sampling could not provide the information required in this case. The only feasible solutions were to scrap the lot of rails or to accept the risk of an accident.

In many QPDAC applications, we will use data that are already available; that is, data that have been collected for some other purpose. For example, we may examine scrap records or measurements taken as part of the process control plan. These data can be helpful but it is wise to think about how they were collected. Are we likely to be misled because of large study or sample error?

In summary, we most often recommend systematic sampling to cover the study population. Systematic sampling is relatively easy to implement. Choose a sample size as large as you can afford remembering that there are likely to be more investigations before the problem solving is complete.

S5.3 STUDY, SAMPLE, AND MEASUREMENT ERRORS

To help understand Figure 5.1 more fully, Deming (1986) distinguishes between descriptive, enumerative, and analytic investigations (see also Liberatore, 2001). In a descriptive investigation, the target population, the study population, and the sample are all the same. The only potential error is due to the measurement system. In an enumerative investigation, the target and study populations are the same. In this case, study error is not a concern. In an analytic investigation, the target population, the study population, and the sample are all different, so we must consider all types of error.

Deming used these distinctions to point out that investigations of the different types are subject to different types of uncertainty due to the possibility of the various types of errors. In our context, we are most interested in analytic investigations since the target population usually extends into the future. In analytic investigations, we must face concerns about possible study error and the trade-off with time and cost considerations.

For chronic problems involving production processes, the study process is often a time-limited version of the target process. However, in some circumstances, the study population may be generated from a different process altogether. For instance, we may draw conclusions about a proposed production process from empirical investigations of a pilot process. This difference increases the potential for study error since the pilot process may not be representative (in terms of the attribute of interest) of the production process.

We can use formal statistical methods (see Box et al., 1978) to quantify sample and measurement error. For a familiar example, think about the statement at the bottom of most public opinion polls:

A survey of this size is accurate to within 3 percentage points 19 times out of 20.

This is a statement about the likely size of sample and measurement error. For analytic investigations, Deming's point is that such a statement does not apply to study error. Our message is that we can control study error only by careful specification of the study and target populations.

S5.4 OUTLIERS

An *outlier* is an unexpectedly large or small value. The occurrence of outliers in any empirical investigation can have a large effect on the conclusions. We try to identify outliers and control their impact.

An outlier can have a large influence on the estimate of the standard deviation, one of the attributes we use most often. We use an artificial example to illustrate. Suppose we take a sample of 100 parts from a process and measure an output characteristic (denoted y). We present the data summary from MINITAB and the process histogram in Figure S5.3.

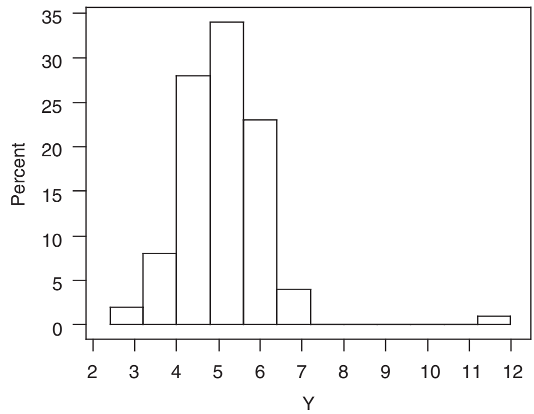


Figure S5.3 Histogram of 100 output values.

```
Descriptive Statistics: y
Variable          N      Mean    Median    TrMean     StDev   SE Mean
y               100     5.081     5.000     5.049     1.059     0.106

Variable    Minimum   Maximum        Q1        Q3
y             2.500    11.400     4.425     5.600
```

The first value in the data set is 11.4, an apparent outlier. We can study the effect of the outlier on the sample standard deviation by replacing 11.4 by a number of values between 1 and 20 (denoted x) and then recalculating the sample attribute. We plot the results in Figure S5.4.

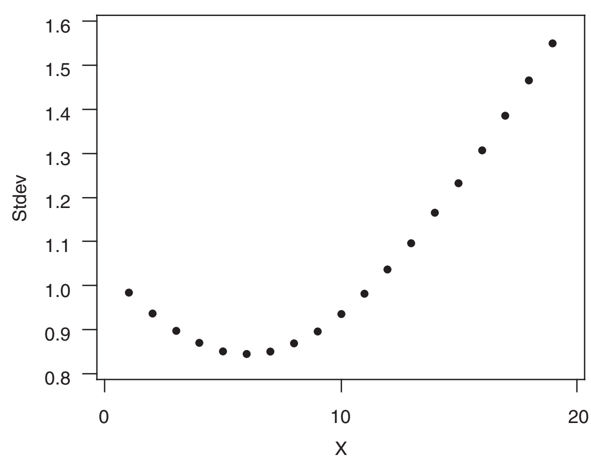


Figure S5.4 Effect of changing a single value on the sample standard deviation.

We see that the standard deviation is very sensitive to changes in a single value in a sample of size 100. If the first value in the sample had been 5 (close to the average) rather than 11.4, the standard deviation would have been reduced from 1.06 to 0.85, a 20% reduction.

What should we do if we find an outlier? We first look at the data collection procedure to ensure that no blunder occurred, such as incorrectly entering the data into the spreadsheet. If a blunder occurred and it cannot be easily corrected, we have a number of options. We can remove the observation from the analysis. This is the preferred approach for investigations such as estimating the problem baseline (see Chapter 6) where there are many observations under similar conditions. For other investigations such as designed experiments, where we make relatively few measurements under a number of combinations of inputs, we can repeat the observation or replace the outlier by an estimate of its value made from the remaining observations.

If we cannot attribute the outlier to a blunder, we have a much more difficult decision. First, if eliminating or changing the outlier has no effect on conclusions then we may ignore its presence. If the outlier does have an effect, the key question is whether the dominant cause of the outlier is the same as the cause of the problem we are trying to address. This question is unfortunately not answerable without considerable process knowledge. It is possible that the outlier comes from a different, perhaps rare, failure mechanism that we are not trying to address. On the other hand, the outlier could be an extreme example of the action of the dominant cause of the current problem. In the former case, the presence of the outlier will make finding the cause and solving the problem more difficult, since the outlier will cloud the effects we are looking for. In the latter case, the outlier may provide a great deal of useful process knowledge. Consideration of outliers is especially critical when using the idea of leverage, as defined in Chapter 9.

Chapter 6 Supplement

Defining a Focused Problem

S6.1 CONFIDENCE INTERVALS FOR THE PROCESS BASELINE ATTRIBUTES

We can use *confidence intervals* to assess uncertainty due to sampling and measurement error. A confidence interval is a range of plausible values for an attribute in the study population. The interval is based on the estimate of the attribute from the measured values in the sample. By default, we use 95% confidence intervals so that we are confident that the unknown attribute falls within the given interval.

For example, consider the V6 piston diameter data discussed in Chapter 5. The team collected diameter data for 469 pistons over a one-week period. The data are found in the file *V6 piston diameter baseline*. The attribute of interest is the standard deviation in the target population. Using MINITAB we produced the following numerical summaries:

```
Variable     N      Mean     Median     TrMean      StDev    SE Mean
diameter   469    590.85     590.90     590.83       3.32       0.15

Variable    Minimum    Maximum         Q1         Q3
diameter     581.50     602.80     588.40     593.05
```

The estimated standard deviation is 3.32 microns. We can get an approximate 95% confidence interval for the study population standard deviation of the form

$$(c_1\text{* estimated standard deviation, } c_2\text{* estimated standard deviation})$$

where c_1 and c_2 are found in Table S6.1. The given constants can be derived using a Gaussian assumption (as discussed in Chapter 2) to model the distribution of output values. The degrees of freedom are related to the sample size. Technically, in the previous calculation, the degrees of freedom are one less than the sample size. In general, for large sample sizes (greater than 30), you can replace the degrees of freedom by the sample size.

In the V6 piston diameter example, the confidence interval for the study population standard deviation is (3.08, 3.59) microns. We can be confident that the standard deviation in the study population is in this range. Even though we used a large sample size, we can only be confident that the attribute is within about ±7% of the estimate. In general, from Table S6.1 we note that the process standard deviation is poorly estimated when the sample size is small.

Table S6.1 Constants for confidence intervals for standard deviation.

Degrees of freedom	c_1	c_2
10	0.69	1.75
20	0.77	1.44
30	0.80	1.34
40	0.82	1.28
50	0.84	1.24
60	0.85	1.22
80	0.87	1.18
100	0.88	1.16
200	0.91	1.11
300	0.93	1.09
500	0.94	1.07
1000	0.96	1.05

We use the estimated standard deviation as a measure of baseline performance. In the example, if we change the process and re-estimate the standard deviation, we hope to see an estimate substantially less than 3.08 (the lower endpoint of the confidence interval) in order to be confident that the change has produced a positive benefit.

We can use the confidence interval to assess the likely size of sampling and measurement errors. The range of values for the attribute captures these uncertainties. However, the confidence interval tells us nothing about study error.

Sometimes we may want a confidence interval for an average in the study population. In the V6 piston diameter example, MINITAB has done most of the work. The estimated attribute is the sample average 590.85. To derive a confidence interval for the average, we look for the SE (denotes standard error) of the mean. In the example, the standard error of the estimate for the average (denoted "SE Mean" in the MINITAB results) is 0.15. For large sample sizes (sample sizes greater than 30), the 95% confidence interval is (approximately) sample average ± 2* (standard error of the average)

or, in this case, 590.85 ± 0.30. In other words, we can be confident that the study population average diameter falls in the range (590.55, 591.15) microns.

For many other attributes, the form of the confidence interval is the same as for the average, that is,

estimated attribute ± 2* (standard error of the estimate)

We can often find both the estimated attribute and the standard error of the estimate in the MINITAB summary.

For estimating a small proportion, such as the proportion defective, we get a rough idea of the precision of the estimate from a confidence interval of the form

$$\text{estimated proportion} \ \pm \ 2\sqrt{(\text{estimated proportion})/(\text{sample size})}$$

This formula is rough because we do not expect the underlying assumptions to hold. However, we use the formula to give some guidance about sample size. For example, if we expect the baseline defect rate to be around 0.01, then with a sample size of 1000, the confidence interval for the proportion defective is about 0.010 ± 0.006. We see there is large relative error in the estimate due to the sampling error.

We use confidence intervals when it seems necessary to quantify the uncertainty in estimated attributes. We often ignore this uncertainty, especially in the searching for the dominant cause. There are many reference books that deal with confidence intervals and other formal statistical procedures. See, for example, Box, Hunter, and Hunter (1978).

S6.2 PROCESS STABILITY AND QUANTIFYING BASELINE PERFORMANCE

There is a myth that we cannot improve process performance with Statistical Engineering methods until we have established a stable process in the sense of Statistical Process Control (SPC). We refer to this property as stability. For example, see section 6 of the AIAG QS-9000 SPC Manual (1995b).

It is true (see the discussion in Chapter 4) that we need a controlled process before it makes sense to apply Statistical Engineering. The main reason is that huge improvements can be made cheaply by ensuring that standard practices and a control plan are followed, that routine maintenance and housekeeping are carried out, that process personnel are trained, and so forth. We do not need to use the sledgehammer of Statistical Engineering to achieve the large reduction in variation that comes from fixing obvious problems.

What role does stability play in quantifying baseline performance? Consider, for example, the truck pull variation reduction problem discussed in Chapter 6. The study population was a two-month period and the sample was all trucks produced in that period. Here there is no sample error. The baseline performance measure was the pull standard deviation calculated over the two-month period. We assumed that this two-month period was long enough so that the standard deviation over this study population would describe how the process would perform into the future if no changes were made. In the language of QPDAC, we assumed that there was little study error.

For the purpose of discussion, suppose that a control chart had been in place at the four gages at the start of the project. There are many sampling protocols that could be used to generate control charts. The process may be stable using some of these protocols and unstable using others. In other words, stability is not solely a property of the process; it also depends on the sampling and charting procedure used. Suppose that in our example, the chart was based on pull measured on five consecutive trucks, once per shift. In Figure S6.1, we show the $\bar{X}$ chart based on the first five trucks produced after 2:00 A.M., 10:00 A.M., and 4:00 P.M. sampled from the data set *truck pull baseline* for all days.

The process is unstable with respect to this charting procedure. That is, there are causes (varying inputs) that change from shift to shift *within* the two-month period that produce the systematic patterns seen on the chart. The effects of these causes contribute to the overall standard deviation. Our assumption about study error is that in the future, the control chart would show similar patterns of instability. This is an assumption and the control chart here provides no help in deciding if the assumption is reasonable or not. If we were concerned with the assumption and if the data were available, we could examine a sequence of two-month periods in the past to see if such a time frame captures most of the process variation.

In summary, the selection of a study population with a time frame long enough to well represent the future is a key issue in establishing a baseline. Stability within the selected study population is not the issue. We emphasize the word *within* because the baseline standard deviation describes the whole two-month period.

Whether the process is stable or not, we may use control chart data to quantify the baseline and provide clues about the dominant cause.

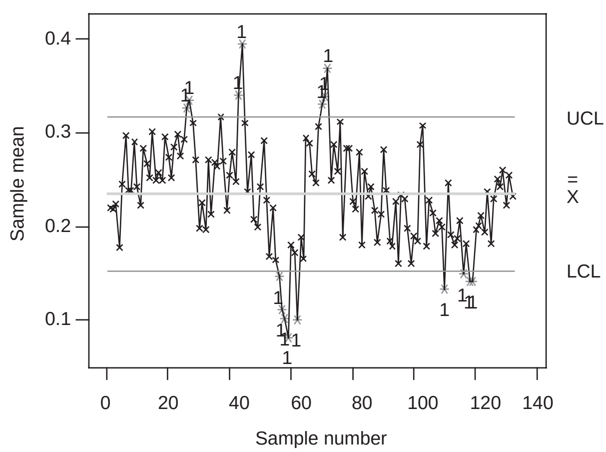


Figure S6.1 $\bar{X}$ chart for truck pull.

Chapter 7 Supplement

Checking the Measurement System

S7.1 ASSESSING A BINARY MEASUREMENT SYSTEM

With a binary measurement system, there are two outcomes we call *pass* and *fail*. We suppose here that the characteristic being measured is also binary, either good or bad. Many binary measurement systems first measure one or more continuous characteristics and then classify the part as pass or fail. If we can measure these underlying continuous characteristics, we can assess the properties of the measurement system (before discretization) using the plans proposed in Chapter 7.

Problems with a binary output may be specified in terms of excess scrap and rework or in terms of customer complaints about receiving bad parts. In the first case, the baseline performance measure is $P(pass)$, the proportion of parts passed by the measurement system. In the second case, we use $P(good|pass)$, the proportion of parts sent to the customer that are good. Note the | in the proportion indicates a conditional proportion—in this case, the proportion of good parts among all those passed by the measurement system. We can write this conditional proportion as

$$P(good \mid pass) = \frac{P(good \ and \ pass)}{P(pass)} \tag{S7.1}$$

From the baseline investigation we have a good estimate of $P(pass)$. This is analogous to obtaining an estimate of $stdev(total)$ from the baseline in a problem with a continuous output. The key formula relating the properties of the measurement system and the manufacturing process is

$$P(pass) = P(pass|good)P(good) + P(pass|bad)P(bad) \tag{S7.2}$$

The proportions $P(good)$ and $P(bad) = 1 - P(good)$ are properties of the manufacturing process. The misclassification rates $P(fail|good) = 1 - P(pass|good)$ and $P(pass|bad)$ are properties of the measurement system. For a binary output, Equation (S7.2) corresponds to the partition of the overall standard deviation into components due to the measurement system and the manufacturing process for a continuous output.

We can derive Equation (S7.2) by noting that

$$P(pass \mid good) = \frac{P(good \text{ and } pass)}{P(good)} \text{ or } P(good \text{ and } pass) = P(pass \mid good)P(good)$$

and

$$P(pass \mid bad) = \frac{P(bad \text{ and } pass)}{P(bad)} \text{ or } P(bad \text{ and } pass) = P(pass \mid bad)P(bad)$$

We finally note that the proportion of passed parts is

$$P(pass) = P(pass \text{ and } good) + P(pass \text{ and } bad)$$
$$P\left(pass\right) = P(pass \mid good)P(good) + P(pass \mid bad)P(bad)$$

If instead the baseline performance is specified by $P(good|pass)$, we can rewrite Equation (S7.1) as

$$P(good|pass) = \frac{P(pass|good)P(good)}{P(pass)}$$

to separate the effects of the measurement system from those of the rest of the process.

In summary, to assess the binary measurement system we estimate the misclassification rates $P(pass|bad)$ and $P(fail|good)$, attributes of the measurement system in the population of all future measurements. If these proportions are too high, then the measurement system is not adequate and must be improved before proceeding to the next stage of the Statistical Engineering algorithm.

We illustrate these ideas using an example in which credit cards were classified *pass* or *fail* by an automated visual inspection system. The aim was to find visual defects such as missing parts of the intended design, surface scratches, bleeding of the colors, fuzzy letters and numbers, and so on. The measurement system used a digital image of the front of each card to calculate hundreds of summary measures based on comparing the picture to a template of the ideal card. If none of the summary measures fell outside the prespecified ranges, the measurement system passed the card. In the baseline investigation, the team monitored the measurement system for more than one week and found an 89.5% pass rate. About 30,000 cards were measured each hour. The goal was to reduce the proportion of bad cards being shipped to the customer $P(pass|bad)$ without substantially decreasing the pass rate $P(pass)$.

Estimating the Misclassification Rates

To estimate the misclassification rates, we must determine whether a number of parts are good or bad. This corresponds directly to the need to know true values in the determination of bias in a continuous measurement system. We illustrate two different plans to estimate the misclassification rates using the credit card example.

Plan 1

We start with equal-sized groups of passed and failed parts. We select each group over a period of time and over a range of conditions so that they represent the long-run populations of passed and failed parts. Then we determine the true state (good or bad) of each part.

In the credit card example, the team selected 40 passed and failed cards per day for five days. Then human inspectors classified each card as good or bad. The data are summarized in the following table.

	Good	Bad	Total
Pass	195	5	200
Fail	16	184	200

From these data, we estimate $P(good|pass) = 195 / 200 = 0.975$ and $P(good|fail) = 16 / 200 = 0.080$.

With the a derivation used to obtain Equation (S7.2), we can express the proportion of good parts as

$$P(good) = P(good|pass)P(pass) + P(good|fail)P(fail)$$

Substituting the estimates for $P(good|pass)$ and $P(good|fail)$ and the overall pass rate 0.895, we estimate the proportion of good cards in the population as

$$P(good) = 0.975 \text{ ¥ } 0.895 + 0.080 \text{ ¥ } 0.105 = 0.881$$

and the first misclassification rate by

$$P(fail\,|\,good) = 1 - P(pass\,|\,good) = 1 - \frac{P(good\,|\,pass)P(pass)}{P(good)} = 1 - \frac{0.975 \text{ ¥ } 0.895}{0.881} = 0.010$$

Similarly, the estimate of the second misclassification rate is

$$P(pass\,|\,bad) = \frac{P(bad\,|\,pass)P(pass)}{P(bad)} = \frac{0.025 \times 0.895}{1 - 0.881} = 0.188$$

The team decided that the measurement system needed improvement because the estimate for $P(pass|bad)$ was large. If the team could reduce this misclassification rate, the proportion of bad cards shipped to the customer would decrease from 0.025.

Plan 2

Here we start with equal-sized groups of good and bad parts. We select each group over a period of time and over a range of conditions so that they represent the long-run populations of good and bad parts. Using the credit card example, the operators examined cards until they had 200 good cards and 200 bad cards. Next, the cards were classified by the measurement system with the following results.

	Pass	**Fail**	**Total**
Good	190	10	200
Bad	8	192	200

From these data, we can directly estimate the misclassification rates as $P(fail|good)$ = 10/200 = 0.05 and $P(pass|bad)$ = 8/200 = 0.04. Since we also know $P(pass)$ = 0.895, we can estimate $P(good)$ using Equation (S7.2):

$$P(pass) = P(pass|good)P(good) + P(pass|bad)P(bad)$$

Substituting the estimates, we have 0.895 = 0.95$P(good)$ + 0.04(1 –$P(good)$), and solving, we find

$$P(good) = \frac{0.895 - 0.04}{0.95 - 0.04} = 0.940$$

Comments

The second plan gives direct estimates of the misclassification rates and an indirect estimate of $P(bad\,|\,pass)$, whereas the first plan provides indirect estimates of the misclassification rates and direct estimates of $P(bad|pass)$. Which plan is preferred depends on costs and the way the baseline is specified. The second plan will not be feasible in many cases because it will be difficult to get representative samples of good and bad parts.

If we decide that the misclassification rates are too high, we can investigate families of causes such as operator-to-operator or time-to-time by repeating the investigation over different time periods or with different operators.

The estimates produced by these plans will be imprecise unless we have large samples. In the first plan using the credit card example, we can be confident that the estimate $P(pass\,|\,bad) = 0.188$ is within about ±0.06 of the actual misclassification rate. To increase the precision of the estimate, we need to determine the characteristic (good or bad) of more than 200 passing and failing cards.

See Boyles (2001) and AIAG (1995a) for more details on alternative ways of assessing variation due to binary measurement systems. The AIAG method is not feasible in the credit card example because it assumes a measurable continuous output has been discretized to give the binary output.

S7.2 ASSESSING A DESTRUCTIVE MEASUREMENT SYSTEM

Here we discuss the planning and analysis of an investigation to assess a destructive measurement system for a continuous characteristic. A measurement system is destructive if we change the true value of the characteristic by making the measurement. Tensile strength is a good example. We cannot make repeated measurements on the same part when the system is destructive.

To assess measurement variation, we select parts that we hope have identical or very similar characteristic values. If we have many similar units, we recommend the plan and analysis in sections 7.2 and 7.3. If there are only pairs of similar units, we can use an investigation like an Isoplot (see Section S7.4).

The following example of assessing a destructive measurement system for the tensile strength of tin-plated steel is motivated by Spiers (1989). In the baseline investigation, the team found that the overall variation in strength was 2.5 KSI (thousands of pounds per square inch) relative to the tolerance ±6.0 KSI measured from nominal.

The next step was to check the measurement system. To determine tensile strength, the operator:

1. Cut a standard sized sample from a sheet of steel
2. Calculated the cross sectional area of the sample with a micrometer
3. Pulled the sample apart in a tensometer

The tensile strength is the ratio of the maximum load to the cross-sectional area.

In the measurement investigation, 30 standard-sized pieces were cut from three different sheets of steel, parallel to the rolling direction and one-quarter distance from the edge of the sheet to eliminate variation in thickness and tensile strength due to edge or crown effects. The sheets were chosen to cover the full extent of variation in strength seen in the baseline investigation. The 30 pieces from each sheet were then randomly divided into three groups of 10. The three appraisers each measured the tensile strength of all 30 pieces (10 pieces from each sheet). We give the data in the file *tin plate strength measurement*.

In the analysis, we assume that all the pieces cut from the same sheet have the same tensile strength. This is analogous to measuring the same part a number of times in the

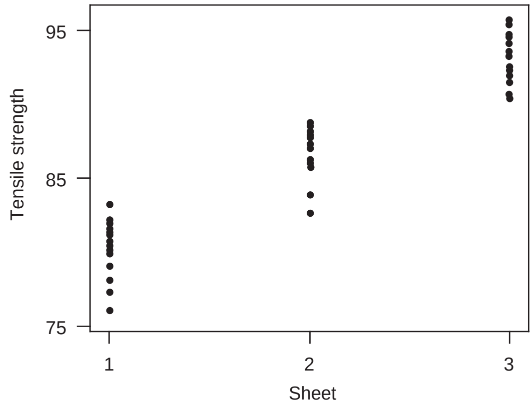


Figure S7.1 Plot of tensile strength by sheet.

nondestructive measurement case. Figure S7.1 shows the measured tensile strength versus the sheet number. A numerical summary is:

```
Descriptive Statistics: tensile strength by sheet

Variable   sheet          N      Mean     Median     TrMean      StDev
tensile    1             30    80.620     80.900     80.762      1.500
           2             30    87.177     87.450     87.354      1.413
           3             30    93.477     93.550     93.531      1.406

Variable   sheet    SE Mean    Minimum    Maximum         Q1         Q3
tensile    1          0.274     76.100     83.200     80.000     81.600
           2          0.258     82.600     88.800     86.300     88.300
           3          0.257     90.600     95.800     92.425     94.700
```

From the MINITAB results, we estimate *stdev*(due to measurement) as

$$\sqrt{\left(1.5^2 + 1.413^2 + 1.406^2\right)/3} = 1.44$$

Using the estimate *stdev* (total) of 2.50, obtained from the baseline investigation and Equation (7.2), we estimate

$$stdev(\text{process}) = \sqrt{stdev(\text{total})^2 - stdev(\text{due to measurement})^2} = 2.04$$

so the estimated discrimination ratio D is 1.42.

We need to be careful in the interpretation of D, because of the destructive nature of the measurement system. The term *stdev*(due to measurement) includes two components, one due to the measurement variation and one from the variation in true values of the samples from each sheet. If the ratio is large—that is, $D > 3$—we can conclude that the measurement system is acceptable because the combined effect of the two components is small. If the

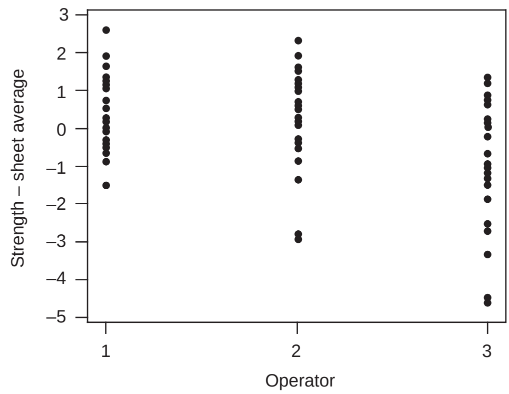


Figure S7.2 Tensile strength minus sheet average by operator.

ratio is small—that is, $D < 3$—we cannot be sure if the fault lies in the measurement system or if there is large variation within the samples taken from the same sheet.

In the example, the team concluded that they should improve the measurement system before continuing further with the Statistical Engineering algorithm. In further analysis of the measurement results, as shown in Figure S7.2, the team found that some of the measurement variation was due to the differences between the operators, with operator three exhibiting more variation than the other two operators.

To estimate measurement bias with a destructive measurement system, it is necessary to use parts with known characteristic value. These standard parts are difficult to obtain unless there is some other (better) nondestructive measurement system.

S7.3 REPEATABILITY, REPRODUCIBILITY, LINEARITY, AND STABILITY

For much of the automotive industry, as part of QS-9000 (Automotive Industry Action Group [AIAG], 1998), suppliers are required to periodically assess measurement systems used for process control and product inspection. The suppliers routinely consider *repeatability* and *reproducibility.* Less frequently, they examine *linearity* and *stability.* Here we define these attibutes of the measurement system and relate them to measurement variation and bias as described in Section 7.1. Other than stability (which we incorporate into our definition of measurement variation and bias), we feel these additional attributes are only useful if we need to improve a measurement system.

From the Automotive Industry Action Group (AIAG) *Measurement Systems Analysis* manual (1995, pp. 17–18), we quote the following definitions:

> *Repeatability:* the variation in measurements obtained with one measurement instrument when used several times by one appraiser while measuring the identical characteristic on the same part.

> *Reproducibility:* the variation in the average of measurements made by different appraisers using the same measuring instruments when measuring identical characteristics on the same part.

Stability: the total variation in the measurements obtained with a measurement system on the same master or part when measuring a single characteristic over an extended time period.

Linearity: the difference in bias values through the expected operating range of the gage.

There is some ambiguity in these definitions that is evident if we try to create the corresponding attributes. The target population is all measurements to be made in the future with the measurement system over a wide range of parts, time, operators, and environmental conditions.

The output is the measurement error for each act of measuring. We assume that the measurement system has a single gage.

Repeatability is the variation in the measurement errors when a single operator measures the same part. The definition makes no mention of which operator, which part, over what time frame, and under what conditions. One way to make the definition more precise is to consider repeatability a short-term measure of the variation in the measurement errors when the operator, part, and conditions are held fixed. That is, we think of repeatability as the average over time, operators, parts, and conditions of the standard deviation of the measurement errors in the consecutive measurement of the same part.

Reproducibility is variation from operator to operator. We can specify reproducibility more precisely as the variation of bias from operator to operator in the system. Bias for a particular operator is the average measurement error for that operator over a wide range of times, parts, and environmental conditions. We think of reproducibility as the standard deviation of the operator biases.

Stability as defined by AIAG is the variation in the measurement errors if the same part is measured repeatedly over a long period. To make this notion precise, we have to average these standard deviations over different parts, operators, and environmental conditions. Stability is to the long term what repeatability is to the short term. Stability is a badly chosen name, since it is not connected to the Statistical Process Control (SPC) use of the word as in a stable process.

Linearity is a measure of the variation in bias over part size. More precisely, for each part, we define the bias as the average measurement error over operators, time, and environmental conditions. Linearity is the standard deviation over all parts of these biases. It is not clear why the definition of linearity does not include possible changes in measurement variation over part size.

The four attributes are a strange collection. Taken together, we cannot use them to estimate the overall variation in the measurement system as defined in Chapter 7. Each attribute is one component of the overall variation, but there are many other components. If there is no time, part, or environmental component to the overall variation of the measurement errors, then we can combine repeatability and reproducibility to estimate the overall variation of the system.

We can assess these attributes (except linearity as defined earlier in terms of bias) using the plan for assessing the overall variation described in Section 7.2. Consider the camshaft example in which there were three parts, three operators, and two weeks. On one day in the week, each operator measured every part three times. We assess repeatability by the average variation within each part, operator, and week and reproducibility by the variation in the operator averages over all parts, weeks, and repeated measurements. We assess stability graphically by comparing the within-week variation over the two weeks.

To estimate repeatability we use a one-way ANOVA model. See Appendix D for further discussion of ANOVA models. The relevant MINITAB ANOVA results are:

```
One-way ANOVA: diameter versus repeat

Analysis of Variance for diameter
Source     DF        SS        MS        F        P
repeat     17  5024.511   295.559   841.83    0.000
Error      36    12.639     0.351
Total      53  5037.151
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ------+---------+---------+---------+
 1          3   -11.658     0.449   *)
 2          3     1.289     0.150                  (*
 3          3    11.940     1.020                               (*)
 4          3   -10.886     0.561    *)
 5          3     0.024     0.242                (*)
 6          3    12.581     0.792                                (*)
 7          3   -10.778     0.876    (*
 8          3     1.533     0.872                  (*)
 9          3    13.058     0.141                                (*)
10          3   -10.566     0.547    (*)
11          3     1.124     1.156                  *)
12          3    12.494     0.245                                (*
13          3   -11.219     0.251   (*)
14          3     1.358     0.410                  (*)
15          3    12.196     0.571                               (*)
16          3   -10.300     0.274    (*)
17          3     1.895     0.150                   *)
18          3    13.843     0.397                                 (*)
                                   ------+---------+---------+---------+
Pooled StDev =    0.593                -8.0       0.0       8.0      16.0
```

We estimate the repeatability as 0.593. Since the estimate of the overall measurement standard deviation was found earlier to be 0.756, the repeatability is a large component.

We give the operator averages in the following MINITAB summary:

```
Descriptive Statistics: diameter by operator

Variable   operator          N      Mean    Median    TrMean     StDev
diameter   1                18      0.76      1.20      0.81      9.84
           2                18      0.67      0.65      0.64      9.86
           3                18      1.55      1.80      1.58     10.10

Variable   operator    SE Mean   Minimum   Maximum        Q1        Q3
diameter   1              2.32    -12.20     12.90    -10.80     12.05
           2              2.32    -11.50     13.30    -10.95     11.85
           3              2.38    -11.50     14.10    -10.07     13.13
```

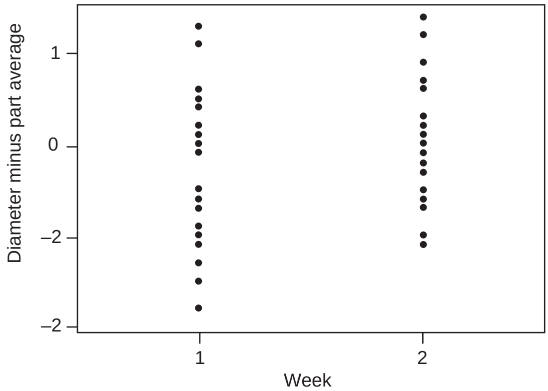


Figure S7.3 Deviation from part average versus week.

The standard deviation of the operator averages is 0.484. To estimate reproducibility, we calculate

$$\sqrt{0.484^2 - (0.593)^2/18} = 0.463$$

also a large component of the overall variation.

We can assess stability by plotting the deviations from part averages versus week as in Figure S7.3.

If we know the true values of the parts being measured, as in Section 7.3, we can assess linearity by plotting the average measurement error (estimated bias) for each part, where the average is calculated over all days, operators, and repeats, versus the part size.

In summary, the repeatability, reproducibility, stability, and linearity of a measurement system do not by themselves play an important role in the assessment of the system for use in the Statistical Engineering algorithm. What we need is an estimate of the overall measurement variation (and perhaps the bias) to determine if the measurement system is the home of the dominant cause of deviation.

S7.4 GAGE R&R AND ISOPLOT INVESTIGATIONS TO ASSESS MEASUREMENT VARIATION

In sections 7.2 and 7.3, we recommended a plan and an analysis to estimate the variation and bias of a measurement system. The plan involved repeatedly measuring three specially selected parts using different operators, and possibly different gages, over both the short term and long term. Here we discuss two alternatives for estimating the measurement variation: gage R&R and Isoplot investigations.

Gage Reproducibility and Repeatability (R&R)

To execute a gage R&R investigation, as described by AIAG (1995a) and Farnum (1994), we:

- Choose 10 parts from regular production
- Select two or three production operators
- Have each operator measure each part two or three times

We then use an analysis of ranges to estimate the repeatability, the reproducibility, and the R&R, the combined effect of the two sources of variation. The R&R is an estimate of the overall measurement variation only if the time-to-time component is negligible. In the standard analysis, we also use the same data to estimate *stdev*(due to process).

This plan and the basic analysis have several deficiencies:

- There is no time-to-time component in the plan. Most R&R studies are carried out as quickly as possible and miss the day-to-day or week-to-week component of the measurement variation. The R&R likely underestimates the overall measurement system variation.

- The use of 10 parts selected over a short time is not sufficient to produce a reliable estimate of the variation of the true part dimensions in the process. Rather, we should use the available baseline information.

- The analysis of ranges is inefficient and should be replaced by the corresponding analysis of variance. The ANOVA analysis is the default in MINITAB. See Wheeler (1992).

- The numerical summaries should be augmented by appropriate graphs—see the MINITAB results that follow. This allows useful checks for outliers and interaction effects.

To illustrate, we use the camshaft journal diameter context for demonstration purposes. Data generated to be realistic are given in the data file *camshaft journal diameter measurement gageRR*. We simulated the use of three operators who measure each of the 10 parts twice for a total of 60 measurements.

The plan we recommended in Section 7.2 requires the same number of measurements in total as the typical R&R investigation but uses fewer parts and takes longer. We can add multiple gages and time periods to the R&R plan—see the AIAG *Measurement System Analysis* manual (1995)—but then the standard MINITAB gage R&R routine cannot be used.

In any case, using the traditional gage R&R analysis *stdev*(due to process) is poorly estimated. In Table 6.3, the relative precision for estimating a standard deviation with a sample size of 50 is about ±20% For a sample size of 10, relative precision is approximately ±50%. In other words, there is a huge uncertainty.

Alternately, we recommend estimating the overall variation in the baseline investigation that uses hundreds of parts. We can then calculate the variation in the true values, *stdev*(due to process), using the square root formula

$$stdev(\text{due to process}) = \sqrt{stdev(\text{total})^2 - stdev(\text{due to measurement})^2}$$

To use this idea in the MINITAB gage R&R analysis, we need to enter the historical standard deviation. The detailed MINITAB results for the gage R&R are:

Gage R&R Study - ANOVA Method
Gage R&R for measurement

Two-Way ANOVA Table With Interaction

Source	DF	SS	MS	F	P
part	9	2569.27	285.474	668.967	0.00000
operator	2	12.98	6.490	15.209	0.00014
operator*part	18	7.68	0.427	0.887	0.59664
Repeatability	30	14.43	0.481		
Total	59	2604.36			

Two-Way ANOVA Table Without Interaction

Source	DF	SS	MS	F	P
part	9	2569.27	285.474	619.608	0.00000
operator	2	12.98	6.490	14.087	0.00002
Repeatability	48	22.12	0.461		
Total	59	2604.36			

Gage R&R

Source	VarComp
Total Gage R&R	0.762
Repeatability	0.461
Reproducibility	0.301
operator	0.301
Part-To-Part	47.502
Total Variation	48.264

Source	StDev (SD)	Study Var (5.15*SD)	%Tolerance (SV/Toler)	%Process (SV/Proc)
Total Gage R&R	0.87304	4.4962	17.98	14.42
Repeatability	0.67877	3.4957	13.98	11.21
Reproducibility	0.54906	2.8277	11.31	9.07
operator	0.54906	2.8277	11.31	9.07
Part-To-Part	6.89219	35.4948	141.98	113.83
Total Variation	6.94726	35.7784	143.11	114.74

Number of Distinct Categories = 11

From this investigation, we obtain an estimate for *stdev*(due to measurement) of 0.87. The most critical result is given in the row labeled "Total Gage R&R" and the column labeled "%Process (SV/Proc)" in the final ANOVA table. This value estimates 100 * *stdev*(due to measurement)/*stdev*(due to process), that is, 100/D as defined by Equation (7.2).

In the example we have $100/D = 14.42$, or $D = 6.9$. Thus, we conclude that the measurement system is adequate. Note that by providing an estimate for *stdev*(overall) we have not used the estimate for *stdev*(due to process) derived just from the gage R&R (6.89 in the example).

The estimate for *stdev*(due to measurement) derived from the gage R&R results differs from the 0.756 we obtained in Section 7.2 based on the same data. The gage R&R results are determined using a *random effects* ANOVA model, while the Section 7.2 results come from a *fixed effects* ANOVA model with different assumptions. See Feder (1974) for a discussion of the difference between fixed and random effects models. For our purposes, neither model is ideal. However, since we are looking for large effects, either model would suffice.

The AIAG standard for measurement systems is that the measurement variation must be less than 30% of the overall variation. This criterion corresponds to a discrimination ratio D, as defined by Equation (7.2), greater than 3.2. We recommend a more liberal cutoff point for problem solving, since by using the baseline data we have a better estimate for *stdev*(process). The more stringent AIAG standard applies to measurement systems used for process control and part inspections.

The gage R&R analysis should be supplemented by graphical displays to check assumptions underlying the numerical calculations. The useful default plots automatically provided by MINITAB are given in Figure S7.4.

We can examine the various graphs looking for evidence of outliers and interactions between operators and parts. In the example, we see no concerns.

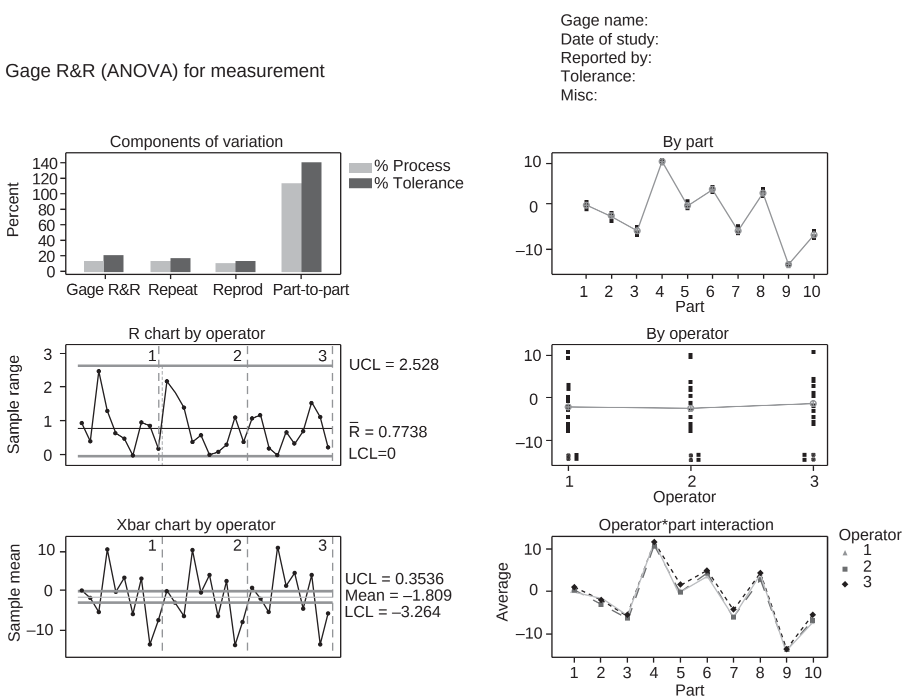


Figure S7.4 Default MINITAB graphical gage R&R output.

Isoplot

A second method for assessing a measurement system is an Isoplot investigation (Traver, 1995). An Isoplot investigation has a simple plan:

- Select 30 parts from production to reflect the baseline variation
- Measure each part twice using different operators, gages, and time periods

Note that if we use the same operator and gage to measure each part over a short time, we are assessing repeatability only. We analyze the data using ANOVA and a scatter plot. We give Isoplot data generated to be realistic for the camshaft journal diameter example in the data file *camshaft journal diameter measurement isoplot*.

The (edited) ANOVA results for the Isoplot investigation are as follows. We use only a *part* term in the ANOVA model.

```
One-way ANOVA: diameter versus part

Analysis of Variance for diameter
Source     DF        SS        MS        F        P
part       29  2551.910    87.997   128.14    0.000
Error      30    20.602     0.687
Total      59  2572.512
                                    Individual 95% CIs For Mean
                                    Based on Pooled StDev
Level       N      Mean     StDev   ------+---------+---------+---------+
 1          2   -10.998     0.030    (*-)
 2          2    -8.322     0.168       (-*)
 . . .
29          2    10.150     0.976                               (-*)
30          2    -2.352     0.508                (*-)
                                    ------+---------+---------+---------+
Pooled StDev =    0.829                -8.0       0.0       8.0      16.0
```

From the ANOVA results, we estimate *stdev*(due to measurement) as 0.829 and we can estimate *stdev*(due to process) as $6.55\left(\sqrt{6.603^2 - 0.829^2}\right)$. The plan uses 30 parts to estimate the process variation. This is an improvement over the gage R&R. With only 30 parts, the relative precision for estimating a standard deviation is approximately ±25%. We prefer to derive an estimate for *stdev*(due to process) by combining the results of the baseline and measurement investigations.

We also plot the first versus second measurement on each of the selected parts as in Figure S7.5. This scatter plot is sometimes called an Isoplot. We can see the process variation by looking at the spread of values on either axis. Note that a point on the 45° line corresponds to the two measurements on that part being identical. If all the plotted points are clustered tightly around the 45° line, then we know that the measurement system standard deviation (or the appropriate component, depending on the plan) is small relative to the process variation.

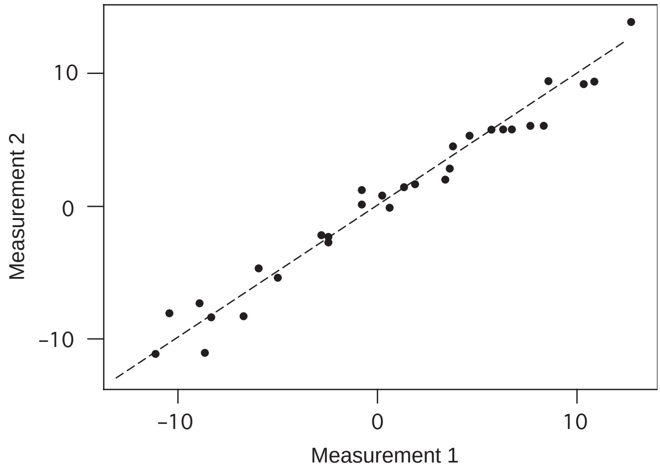


Figure S7.5 Scatter plot of the measurement results (dashed line represents "measurement 1" = "measurement 2," the 45° line).

The scatter plot is valuable because we can easily see the comparison of the measurement variation and process variation. We can also identify any outliers on the plot.

The Isoplot method has several deficiencies:

- There is no time-to-time component in the plan. Most Isoplot investigations are carried out as quickly as possible and miss the day-to-day or week-to-week component of measurement variation.
- Since all second measurements are conducted under identical conditions, only one family of causes that influence measurement variation is assessed at a time.
- The use of 30 parts and 30 repeated measurements is not enough to get good estimates of the process and measurement variation, respectively.

Bhote and Bhote (2000) recommended the use of the Isoplot and a minimum measurement discrimination ratio D of 5 before proceeding to the next stage of their process-improvement algorithm. We believe this is too conservative.

Comparison to Proposed Measurement Assessment Plan in Chapter 7

In comparison to both gage R&R and Isoplot, our proposed measurement assessment plan uses fewer parts with more measurements per part. In this way, the investigation focuses on estimating just the measurement variation (rather than the variation due to the process as well). The available baseline information gives an estimate of the total variation, and using Equation (7.2), we can solve for the variation due to the process. In this way we obtain more precise estimates of the variation both due to the measurement system and due to the process. This allows better decisions regarding the adequacy of the measurement system.

S7.5 INTERPRETING MEASUREMENT VARIATION

In the camshaft diameter example in Section 7.2, the estimated measurement variation is 0.756 microns. What does this tell us about the measurement system?

If we know the bias is close to zero, then a series of measured values on the same part vary about the true value with standard deviation close to 0.756 microns. In other words, we would be surprised to find a measured value more than two standard deviations—that is, 1.63 microns—away from the true value.

When the bias is not known, we cannot interpret the measurement variation in terms of how close the measured value will likely be to the true value. Instead we consider the difference in two measurements on the same part. Using the results from the models described in Chapter 2, the difference in two measurements on the same part will have mean zero (differencing eliminates the unknown bias and true value) and standard deviation close to $\sqrt{2}$ ¥ 0.756 = 1.07. In other words, we would be surprised to see two diameter measurements on the same part differ by more than two standard deviations—that is, 2 ¥ 1.07 = 2.14 microns.

S7.6 EFFECT OF MEASUREMENT VARIATION

We have argued that if the ratio of the process-to-measurement system standard deviations (that is, the discrimination ratio, D) falls between 2 and 3, we should improve the measurement system before we proceed, even though it is not a dominant cause of the overall output variation. Here we justify this recommendation.

Measurement variation makes finding a dominant cause more difficult. For instance, measurement errors will influence estimates of the contributions of causes to the total variation. The importance of measurement variation depends on the approach and is of greater concern for approaches whose implementation requires ongoing measurement, such as feedforward control, feedback control, and 100% inspection.

The effect of measurement variation depends on the goal of the process investigation. To keep the discussion simple, we formally compare the effect of measurement variation on our ability to detect differences in mean output level between two groups of parts. We may compare groups when trying to find or verify a dominant cause of variation and when assessing the proposed variation reduction approaches.

As measurement variation increases, it becomes more difficult to determine if any observed difference between the groups is due to chance or due to a real difference in average level. We quantify the effect of measurement variation based on the increase in sample size (of the two groups) needed to detect the same size difference in the average level with the same power or probability. Our ability to detect a particular size difference in the average level depends on the inherent variation (*stdev*(total)) in the dimensions that here, for simplicity, we assume is the same in the two groups. Part of the inherent variation is due to measurement variation. We compare the case where there is no measurement error—that is, *stdev*(total) = *stdev*(due to process)—to the case where there is measurement error—that is, *stdev*(total), is given by Equation (7.1).

To detect differences in the two groups, we compare the observed sample averages in the two groups. Under conceptual repeated sampling and measuring, the standard error of an average decreases by a factor $1/\sqrt{n}$ as the sample size n increases. As a result, the sample size needed to maintain the desired power is greater by a factor

$$\left(stdev\left(\text{due to process}\right)^2 + stdev\left(\text{due to measurement}\right)^2\right) \Big/ stdev\left(\text{due to process}\right)^2$$

if measurement variation is present.

Thus, using $D = stdev(\text{due to process})/stdev(\text{due to measurement})$, measurement variation results in a $100/D^2$ percent increase in the sample size requirements over the case with no measurement error. For example, if *stdev*(due to measurement) = 0.3 *stdev*(due to process), that is, $D = 3.3$, the sample sizes needed to detect average differences are roughly 9% larger than if there were no measurement error. This seems a fairly small effect. To be somewhat conservative, and acknowledging that there may be some error in our of D, we set the minimum value at 3.

S7.7 FINDING A DOMINANT CAUSE OF MEASUREMENT VARIATION

There are many other attributes of the measurement process that may be of interest, especially if we judge the system to be inadequate. The basic idea is to divide the measurement bias and variation into components that can be assigned to inputs such as operators, parts, time, and the environment. We demonstrate the decomposition using operators.

Suppose that the measurement system includes two or more operators. We have defined the measurement variation and bias in terms of measurement errors based on a broad target population of measurements made on many parts over a long period of time. We can stratify this population according to which operator makes the measurements and, *for any particular operator,* we define the measurement bias and variation of the measurement system as the average and standard deviation of the measurement errors if that operator made the measurements.

If we assume that each operator makes the same proportion of the measurements, we can decompose the idealized histogram of the measurement errors into histograms for each operator such as the one shown in Figure S7.6.

In general (and as seen in Figure S7.6) it is true that:

- The measurement system bias is the average of the individual operator biases.
- The measurement variation depends on the measurement variation for individual operators and the relative biases among the operators.

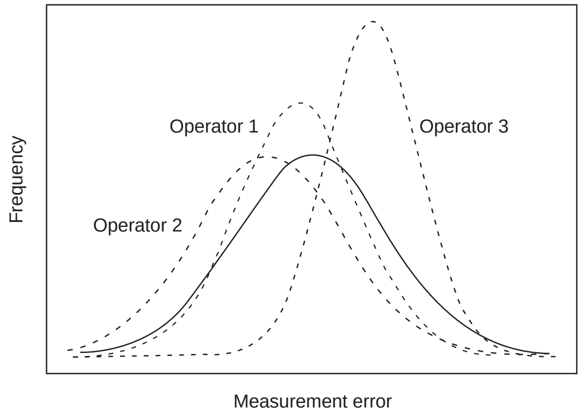


Figure S7.6 Decomposition of measurement errors by operator (solid line shows distribution of overall measurement error).

The overall system bias can be zero even if each operator bias is not. Also, the differences in operator bias contribute to measurement variation. This is another example of the combination of group-to-group and within-group variation that we discussed in Chapter 2 and its supplement.

In looking for opportunities to improve a measurement system, we can stratify by operator. Suppose that we identify large differences among the operator-specific attributes. Then, we know that these differences arise because of differences in the methods among the operators. If we can standardize the method, we will see substantial improvement in the overall measurement system attributes.

We use the camshaft journal diameter data from the plan described in Section 7.2 to look at differences among operators. In this plan, each operator measured each part six times. Since we do not know the true value of the diameters, we calculate

$$\text{deviation} = \text{measured value} - \text{part average}$$

for each of the 54 measured diameters. We show these data plotted by operator in Figure S7.7.

The three operators have roughly the same measurement variation, but some (a small component) of the total variation comes from differences in relative bias between the operators, since the average "diameter minus part average" differs between operators.

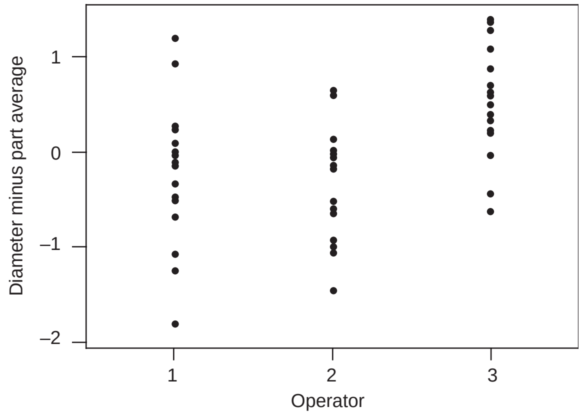


Figure S7.7 Diameter (deviation from part average) by operator.

We may see many different patterns in plots such as Figure S7.7 that will lead us to different actions such as:

- No differences among operator measurement bias and variation—look elsewhere for opportunities to improve the system.
- Different measurement variation or different relative biases for each operator—look for systematic differences among the methods used by the different operators.

We can do the same analysis with other inputs to the measurement process such as gages, environmental inputs, and time.

Chapter 9 Supplement

Finding a Dominant Cause Using the Method of Elimination

S9.1 COMPARISON OF STRATEGIES FOR FINDING A DOMINANT CAUSE

A standard strategy to find the (dominant) cause of a problem is to:

- Create a list of all possible causes, using people experienced with the process.
- Prioritize the list of causes in terms of the likelihood that each has a large effect.
- Investigate the causes in order of priority, either singly or in groups.

The first step often involves brainstorming (Evans and Lindsay, 1993), a formal process designed to elicit ideas from a group of people. We then use a cause-and-effect diagram (Ishikawa, 1982) to categorize the causes into what we call families (main branches) and subfamilies (twigs). We can prioritize the causes using multivoting (Scholtes, 1988), a nominal group technique (Brassard, 1988), or some other process, democratic or otherwise. We then have relatively few causes left to investigate. We can use available data and experimental or observational investigations to determine the effects of the top-ranked causes.

In Figure S9.1, we give a cause-and-effect diagram that was constructed by a team trying to reduce variation in the diameter runout on a transmission output shaft. The diagram is unusual in that it does not use the standard main branches (man, machine, material, method, measurement, and environment). Given this list of potential dominant causes, the team struggled with what to do next. In the end, they abandoned the idea of finding the dominant cause and instead looked to make the process more robust by changing a number of fixed inputs in an experimental investigation. This investigation was a failure.

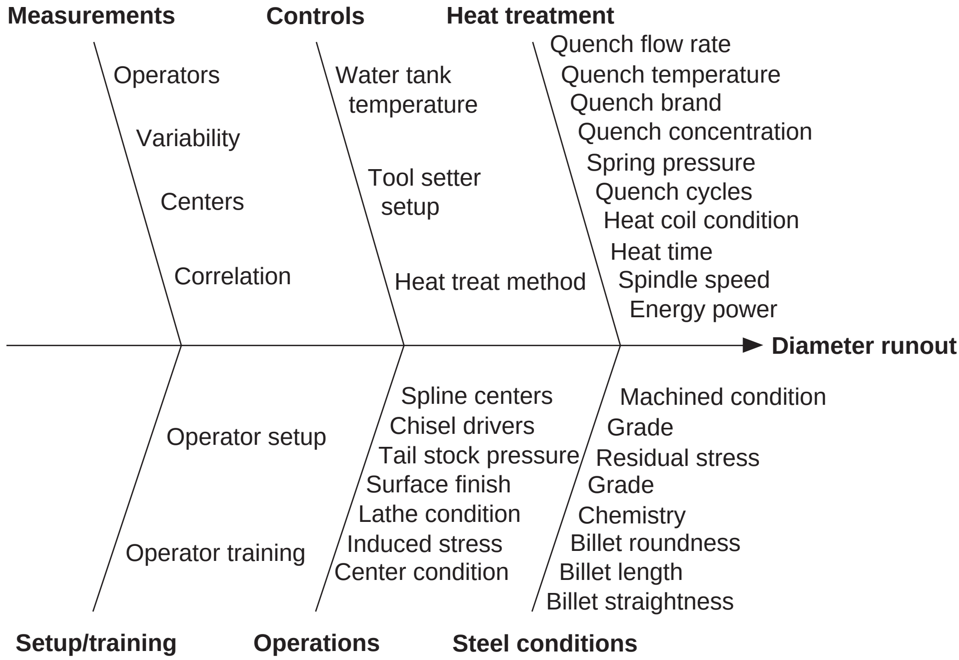


Figure S9.1 Example of conventional brainstorming to list all potential dominant causes.

We have recommended a different strategy, the method of elimination, based on forming and eliminating families of causes using available data and observational plans. The strategy involves:

- Dividing the causes into two or more families
- Using simple process investigations to rule out all but one family
- Iterating until one suspect (or a few at most) remains as a possible dominant cause
- Verifying the suspect(s) as the dominant cause

We expect that we will require several iterations to rule out most causes and home in on the dominant cause. At each iteration, we split the remaining causes into families based on process knowledge and the ease with which we can investigate the process. We never list causes in families that have been eliminated.

There are two major differences between the two strategies.

The cause-and-effect diagram is a hindrance for the method of elimination. The families are not constructed with the idea of elimination. In the output shaft example, we cannot think of an investigation that would eliminate all but one of the five families. If the main branches are considered families, there are too many families to start and time is wasted getting the detailed list of causes within each main branch. We have been involved in projects where considerable time was spent arguing about the labels for the main branches of the cause-and-effect diagram and which causes went where.

The second difference is the method of prioritizing the causes. With the method of elimination, at each step there are only a few families, and we can let the process do the voting by carrying out an appropriate investigation. On the other hand, with a cause-and-effect diagram and multivoting, for example, there is no guarantee that the dominant cause is anywhere near the top or even on the final list.

S9.2 NO SINGLE DOMINANT CAUSE

In the method of elimination, we assume there is a single dominant cause of the baseline variation. This assumption can be wrong in two ways:

- There are two (or perhaps three) causes, each with a large effect.
- There is no dominant cause, only a large number of causes, each with a relatively small effect.

If there are two (or three) dominant causes, each with a large effect, that live in different families, we cannot rule out either of those families. We then investigate the families separately to search for the dominant causes. Once we isolate the dominant causes, we may need to choose a different variation reduction approach to address each of them.

The second case in which there is no dominant cause contradicts the Pareto Principle. We believe this to be a rare occurrence. We may find a dominant family but no single cause within that family. We may rule out any of these families and then decide, in the next split, that there are two dominant families. As we partition the causes further, we will never be able to explain the full extent of variation found in the baseline investigation with a family of only a few causes.

In our experience, the search for the dominant cause is often abandoned in these cases. The consequence is that none of the caused-based variation reduction approaches is an available remedy. We must select a working approach that does not require the identification of the cause.

Chapter 10 Supplement

Investigations to Compare Two Families of Variation

S10.1 ANALYSIS OF VARIANCE

To find the dominant cause, we need to identify the family that explains a large component of the overall variation. In the examples presented in Chapter 10, we relied on graphical displays (for example, scatter plots, box plots, and multivari charts). If needed, we can use a one-way analysis of variance (ANOVA) to estimate the components of variation due to each family. We describe ANOVA in more detail in Appendix D.

Consider the rod thickness example. The data are given in the file *rod thickness baseline*. The ANOVA results correspond to Figure 10.3.

```
Analysis of Variance for thickness
Source      DF        SS        MS        F        P
position     3   46256.5   15418.8   241.45    0.000
Error      796   50831.0      63.9
Total      799   97087.5
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N      Mean     StDev  ------+---------+---------+---------+
1          200    37.820     8.040                            (-*-)
2          200    37.975     8.286                            (-*-)
3          200    21.580     7.559  (-*-)
4          200    40.925     8.062                                 (-*-)
                                    ------+---------+---------+---------+
Pooled StDev =     7.991               24.0      30.0      36.0      42.0
```

The important result in the ANOVA output is the Pooled StDev, which is an estimate of the process standard deviation if all of the position averages are equal. In this example, the Pooled Stdev equals

$$\sqrt{\frac{stdev(\text{due to position 1})^2 + \ldots + stdev(\text{due to position 4})^2}{4}}$$

Pooling the standard deviations within each position is the method that we used to estimate the measurement system variation in Chapter 7, where we pooled the standard deviations within each part.

Since the baseline standard deviation for rod thickness was 11.023 (see Chapter 6), we can make substantial improvement by aligning the position averages. From the plots presented earlier, we need to increase the thickness center for position 3. We can adopt the Move the Process Center approach to learn how to make the change. We can use the ANOVA results to assess the maximum benefit from this process change.

We consider a second example to illustrate some further complexities in interpreting ANOVA results.

Block Porosity

Consider the block porosity example in Chapter 10. The data are given in the file *engine block porosity multivari*. Recall that porosity was measured on five blocks molded consecutively, once per hour over one shift. The goal was to identify which of the two families, hour-to-hour or mold-to-mold, was the home of the dominant cause. We apply a one-way ANOVA with porosity as the output and hour as the only term in the model. The results are:

```
One-way ANOVA: porosity versus hour

Analysis of Variance for porosity
Source     DF        SS        MS        F        P
hour        7    282168     40310     7.72    0.000
Error      32    167133      5223
Total      39    449301
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ------+---------+---------+---------+
1           5     20.80     26.34   (------*------)
2           5    102.60     48.86          (-----*------)
3           5      7.40      5.22  (------*-----)
4           5      8.40      3.71  (------*-----)
5           5    224.80    127.12                    (-----*------)
6           5      8.40      6.99  (------*-----)
7           5     14.20      8.96   (-----*------)
8           5    190.00    149.57                 (------*------)
                                   ------+---------+---------+---------+
Pooled StDev =    72.27                  0       100       200       300
```

The Pooled StDev, 72.3, at the bottom of the MINITAB results gives an estimate of the variation due to causes that act in the mold-to-mold family.

Using the overall standard deviation normally available from the baseline investigation, we estimate the component of the variation due to the hour-to-hour family as

$$stdev(\text{due to hour-to-hour}) = \sqrt{stdev(\text{total})^2 - stdev(\text{due to mold-to-mold})^2}$$

Here the team did not conduct a baseline investigation for the new measure of porosity. They judged that they had seen the full extent of variation during the one-day investigation, because the proportion of scrapped blocks matched the historical level. As a result, we estimate the total variation by looking at the variation in porosity over the course of the investigation. We can use MINITAB to produce this estimate *stdev*(total) = 107.3:

```
Variable       N       Mean     Median     TrMean      StDev    SE Mean
porosity      40       72.1       16.5       58.7      107.3       17.0

Variable       Minimum    Maximum         Q1         Q3
porosity           0.0      418.0        9.0      102.0
```

An alternative is to take into account the special structure in the data. We can indirectly estimate the total variation using the available data by specifying hour as a random effect in the one-way ANOVA. Here the MINITAB results are:

```
General Linear Model: porosity versus hour

Factor     Type Levels Values
hour     random      8 1 2 3 4 5 6 7 8

Analysis of Variance for porosity, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
hour        7     282168     282168      40310    7.72  0.000
Error      32     167133     167133       5223
Total      39     449301

Variance Components, using Adjusted SS

Source    Estimated Value
hour                 7017
Error                5223
```

Based on this result, we estimate the variation associated with the hour-to-hour family as $\sqrt{7017} = 83.8$, and of the total variation as

$$stdev\left(\text{total}\right) = \sqrt{stdev(\text{due to hour-to-hour})^2 + stdev(\text{due to mold-to-mold})^2} = 110.6$$

In this example, the two methods give similar estimates for the total variation.

Here, neither family is dominant. This is not surprising since we only resort to ANOVA when the conclusion from the plots is not obvious. However, we see that if we could eliminate the mold-to-mold variation (or at least a substantial portion of it), we could substantially reduce the overall variation. The major advantage of using ANOVA when there is no obvious dominant family is that we can quantify the contributions of the various families. We can then decide which family we want to pursue. We expect there to be a dominant cause within each family.

Note also that ANOVA Pooled StDev does not capture the interaction between causes in the two families, which is apparent in Figure 10.9. This interaction was the most important discovery in the investigation. We can see the interaction in the ANOVA results since the StDev values for each hour are very different (for example, 26.3 for hour 1 and 149.6 for hour 8). The calculated Pooled StDev masks these differences.

S10.2 POSSIBLE EXPLANATIONS FOR OBSERVED CHANGES IN VARIATION

In some investigations, we observe a change in both process center and variation when the output is stratified by location or time. The engine block porosity problem provides an example. From Figure 10.9 we see that both the average level and variation of porosity change from hour to hour. We learn a lot about the nature of the dominant cause from this observed pattern.

Here we explore possible reasons for a change in both average and variation. We give two simple explanations in Figure S10.1, where the relationship between the dominant cause and the output is roughly linear. The horizontal and vertical axes give the full range of variation for the cause and the output.

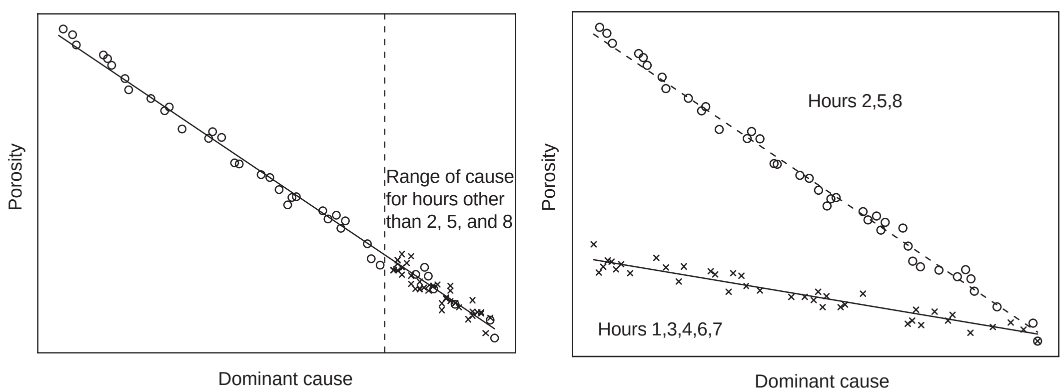


Figure S10.1 Possible explanations for engine block porosity results (*o*—observations for hours 2, 5, and 8; *x*—observations for other hours).

In the left-hand panel, the range of values for the dominant cause for hours 2, 5, and 8 is different than the range for the other hours. In the right-hand panel, the relationship between the cause and porosity depends on the time period, and in hours 2, 5, and 8 the process is much more sensitive to variation in the dominant cause. This pattern in the right panel indicates there is a strong interaction involving the plotted cause and some other input that changes from hour to hour.

In the example, the range of the dominant cause was different in hours 2, 5, and 8. At these times, the pouring temperature was lower and varied substantially, since during breaks the iron cooled.

There is a similar explanation for a change in variation between groups when there is no corresponding change in the process center, as in the hypothetical V6 piston diameter illustrated in Figure 10.17. The two possibilities are illustrated in Figure S10.2, where the horizontal and vertical axes give the full range of variation for the dominant cause and the output. We have:

- Left panel: The variation in the dominant cause is different for streams *A* and *B*
- Right panel: The effect of changes in the dominant cause is much greater for stream *B* than stream *A*

We can also have a combination of the two scenarios.

If we observe such patterns, we cannot rule out the within-strata or strata-to-strata families. However, we do get a strong clue about the nature of the dominant cause. To proceed, for example in the V6 piston investigation, we consider causes within stream *B* and look for differences in causes or effects between the two streams. We have a big advantage because we know it is possible to run process (that is, stream *A*) in a way that results in substantially less output variation.

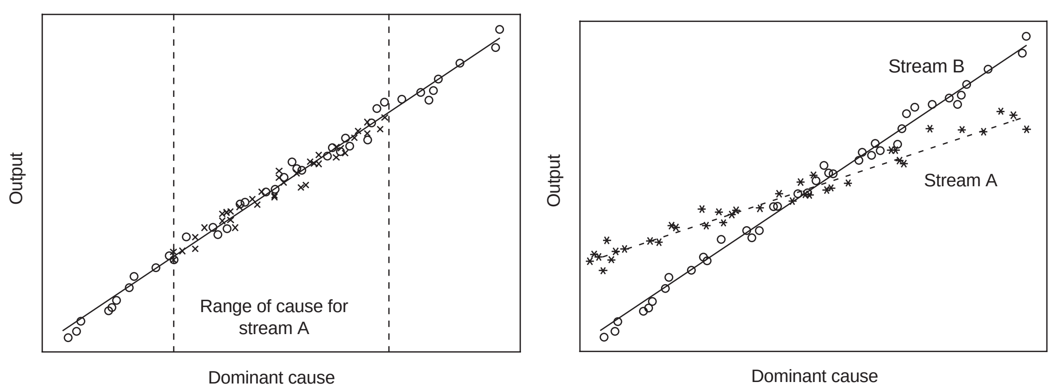


Figure S10.2 Possible explanations for pattern in Figure 10.17 (observation from streams *A* and *B* are denoted with *x* and *o,* respectively).

S10.3 WHY FORMAL HYPOTHESIS TESTS ARE NOT RECOMMENDED

Formal hypothesis tests, such as t-tests and the F-test given in ANOVA output, are used to determine if there is a *statistically significant* difference between the output averages when stratified by the value of an input. In this context, by a *statistically significant* difference we mean that we cannot explain the difference in strata averages by the variation within the strata. Consider the rod thickness example discussed in sections 10.1 and S10.1. We give again the ANOVA results that correspond to Figure 10.1.

```
One-way ANOVA: thickness versus position

Analysis of Variance for thickness
Source     DF        SS       MS        F        P
position    3   46256.5  15418.8   241.45    0.000
Error     796   50831.0     63.9
Total     799   97087.5
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean    StDev   ------+---------+---------+---------+
1         200    37.820    8.040                             (-*-)
2         200    37.975    8.286                             (-*-)
3         200    21.580    7.559   (-*-)
4         200    40.925    8.062                                  (-*-)
                                   ------+---------+---------+---------+
Pooled StDev =    7.991                24.0      30.0      36.0      42.0
```

We see that:

- There are statistically significant differences among the position average thicknesses, since the p-value 0.000 in the F-test is so small
- Some of the 95% confidence intervals for the position means do not overlap in the plot

We conclude that there are position-to-position differences among the averages. We cannot conclude, however, that position is a dominant cause.

We see the same difficulties, as illustrated here for ANOVA, with nonparametric tests based on end-counts (due to Tukey, 1959) as recommended by Bhote and Bhote (2000) and others. The tests can identify groups with statistically significant differences but do not tell us if the dominant cause acts in the group-to-group family.

In general, with large sample sizes, a formal hypothesis test will show that there are statistically significant differences among group averages. The small difference may have no *practical significance,* however. We are searching for a dominant cause that explains a large part of the variation in the output, not for causes with small but statistically significant effects. The dominant cause will produce significantly different averages, but the converse is not necessarily true. For this reason, we do not rely on hypothesis tests to help in the search for a dominant cause.

S10.4 OTHER PLANS FOR COMPARING UPSTREAM AND DOWNSTREAM FAMILIES

In Chapter 10, we looked at a plan to compare upstream and downstream families by measuring the output at an intermediate operation and at the end of the process for the *same* parts. Additional references concerning variation transmission investigations include Agrawal (1997), Lawless et al. (1999), and Agrawal et al. (1999).

Here we consider two other plans for comparing upstream and downstream families that do not require measuring the output mid-process. These plans, operations swap and randomized sequencing, have limited applicability. There are more complex versions if we want to compare several processing steps simultaneously.

Operations Swap

Operations swap (Ingram, 1993) is feasible if:

- Differences between two parallel processes are a dominant cause of variation.
- We can swap parts between the two parallel processes at some intermediate processing step.

The idea is simple. We swap the production path of parts moving through the process at a given processing step (for example, halfway through the process). Figure S10.3 illustrates the plan with two parallel lines and two operation steps. The solid lines show the usual process path through lines 1 and 2. The dashed lines show how we propose to swap the path of parts between operations *A* and *B* temporarily for the purposes of the investigation.

We keep track of the parts that are swapped. Then, we compare the output of the two streams of swapped parts with the performance of the regular processes. We know there was a substantial difference in performance between the two production lines before the swapping. If the performance of the swapped parts that went through operation *A* in line 1 and operation *B* in line 2 matches the performance of line 1 before the swapping

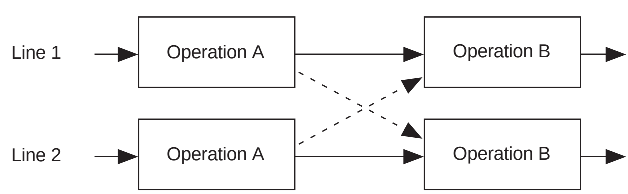


Figure S10.3 Operations swap illustration.

investigation, and similarly for line 2, then we can conclude that the dominant cause of the difference between the two lines is in operation *A* or earlier.

An operations swap investigation can be logistically difficult. We need to continue swapping, tracking swapped parts, and measuring finished parts (though it is not necessary to swap all parts in the given time period) until we see the full extent of variation in the output.

Swapping parts from line 1 to line 2 *and* vice versa is not necessary. We could draw conclusions just by swapping in one direction. However, the double swap is useful to ensure nothing else has changed in the process and, for logistical reasons, to balance the load on the two production lines.

Randomized Sequencing

Randomized sequencing (Asher, 1987) is feasible if:

- The process output exhibits a known pattern of variation—for example, a drift—or sustains shifts.
- We can change the processing order of parts at an intermediate operation and later identify the order used.

Suppose the output drifts with a predictable pattern over each day. To conduct the randomized sequencing plan, we:

- Select a sample of parts spread out over one day after operation *A*.
- Record the time for each sampled part and set it aside.
- Randomize the order of these parts.
- Process and track the parts through operation *B*.
- Measure the output.

This plan is illustrated in Figure S10.4.

If the dominant cause lives in operation *A* or upstream, we will see the predicted drift when we plot the output against the time of processing through operation *A*. If we do not see the predicted drift, then we know the dominant cause lies in operation *B*.

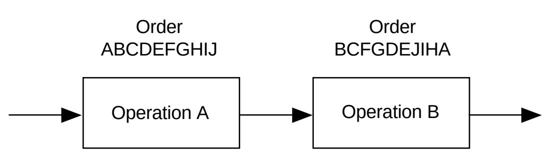


Figure S10.4 Randomized sequencing illustration.

S10.5 APPLYING LEVERAGE IN COMPARING ASSEMBLY AND COMPONENT FAMILIES

To compare the assembly and component families for an assembled product, we recommend the following plan:

- Select two parts with output at the extremes of the baseline distribution.
- Disassemble and reassemble each part at least three times and measure the output each time.
- Plot the results and look for changes in the output.

Large changes in the performance of the product after disassembly and reassembly indicate that the assembly family is the home of the dominant cause. It is surprising that we can make such a claim with such a small sample size. This is a good example of leverage, where we ensured that we would see the full extent of variation in the sample.

We can describe the output for any assembly as a function of a contribution from the components and a contribution from the assembly. The standard deviation of the output can be expressed as

$$stdev(\text{output}) = \sqrt{stdev(\text{due to components})^2 + stdev(\text{due to assembly})^2}$$

Suppose, for the moment, that the assembly variation is dominant—that is, *stdev*(due to assembly) is much larger than *stdev*(due to components). When we select two parts with extreme output values, we can be confident that the reason the values are extreme is that the contributing values from the assembly are also extreme. This must be true since, by our hypothesis, the contributions from the components are not highly variable and hence cannot produce the extreme values.

When we disassemble and reassemble the parts, the contribution of the components does not change. However, the assembly contributions will change, and since we started with extreme values, we expect that there will be large changes to the output.

On the other hand, if the dominant cause lives with the components—that is, *stdev*(due to components) is much larger than *stdev*(due to assembly)—then the reason for the extreme values is a large component contribution. When we disassemble and reassemble such parts, we will continue to see extreme values for the output since the component contribution is not changed.

Chapter 11 Supplement

Investigations to Compare Three or More Families of Variation

S11.1 ANALYSIS OF VARIANCE FOR MULTIVARI INVESTIGATIONS

We use multivari investigations to examine the effects of a number of families of causes simultaneously. If there are three or more families, our strategy is to examine several multivari charts to eliminate families as the home of a dominant cause. If there are many families, or, more important, no single dominant family, we can use analysis of variance to quantify the contribution of each family to the overall variation.

We illustrate the use of ANOVA with three examples where the additional analysis is warranted.

Casting Thickness

At a foundry, a team was assigned the task of reducing variation in thickness, specified at four locations on each casting. A baseline investigation found that the full extent of variation in thickness (as measured from nominal) was –20 to 35 thousandths of an inch. The team found that the measurement system was acceptable and decided to look for a dominant cause of thickness variation.

They conducted a multivari investigation in which they measured thickness at the four positions for three consecutive castings from each of the six cavities in the mold. They sampled each hour for six hours on two different days to get a total of 864 thickness measurements on 216 castings. The data are available in the file *casting thickness multivari*. This investigation will allow us to look at the following families and their interactions:

- Casting-to-casting
- Position-to-position
- Cavity-to-cavity
- Hour-to-hour
- Day-to-day

Plotting a histogram of the multivari data, as shown in Figure S11.1, we see close to the full extent of variation from the baseline.

We give a number of multivari charts in Figure S11.2 looking at the individual families. We have defined a new input *time* that sequentially numbers the 12 time periods representing all the possible combinations of hour and day and a new input *group* to uniquely identify the sets of three consecutive thickness measurements within a position at the different time periods. We used group = (time – 1)*24 + (cavity – 1)*4 + position.

The multivari charts are not easy to interpret, since there is no single dominant cause. We can attribute a large amount of variation to the position-to-position, cavity-to-cavity, and casting-to-casting families (as shown by the multivari chart by group). In this example, we can quantify the components of variation that can be attributed to the various families. Using the model from Chapter 2, if the causes in separate families act independently, we have

$$sd\left(\text{total}\right)=\sqrt{sd\left(\text{due to family1}\right)^2+sd\left(\text{due to family2}\right)^2+\ldots+sd\left(\text{due to rest of causes}\right)^2}$$

In the ANOVA calculation, the total sum of squares is similarly partitioned into sums of squares associated with the various families. See Neter et al. (1996) and Box et al. (1978) for a general background on ANOVA. See also de Mast et al. (2001) for a more formal discussion of the connection between multivari charts and ANOVA.

We propose to fit an ANOVA model with all possible terms involving the inputs whose effects are expected to be systematic. For example, we might expect a systematic difference (in the average thickness) among the four positions. We would not expect systematic variation among the measurements taken at one position from consecutive castings. That is, we would be surprised if the average of the first measurement over all cavities and times was substantially different from the average of the second measurement.

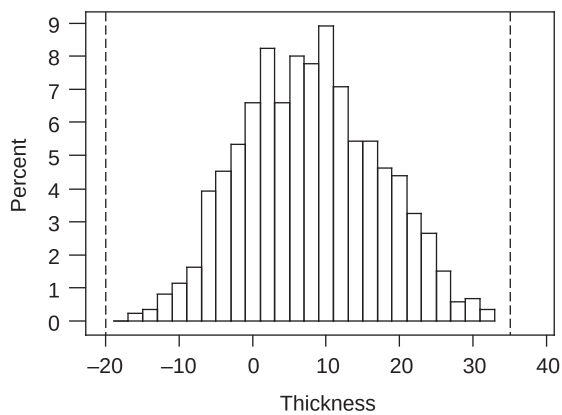


Figure S11.1 Histogram of casting thickness multivari data.

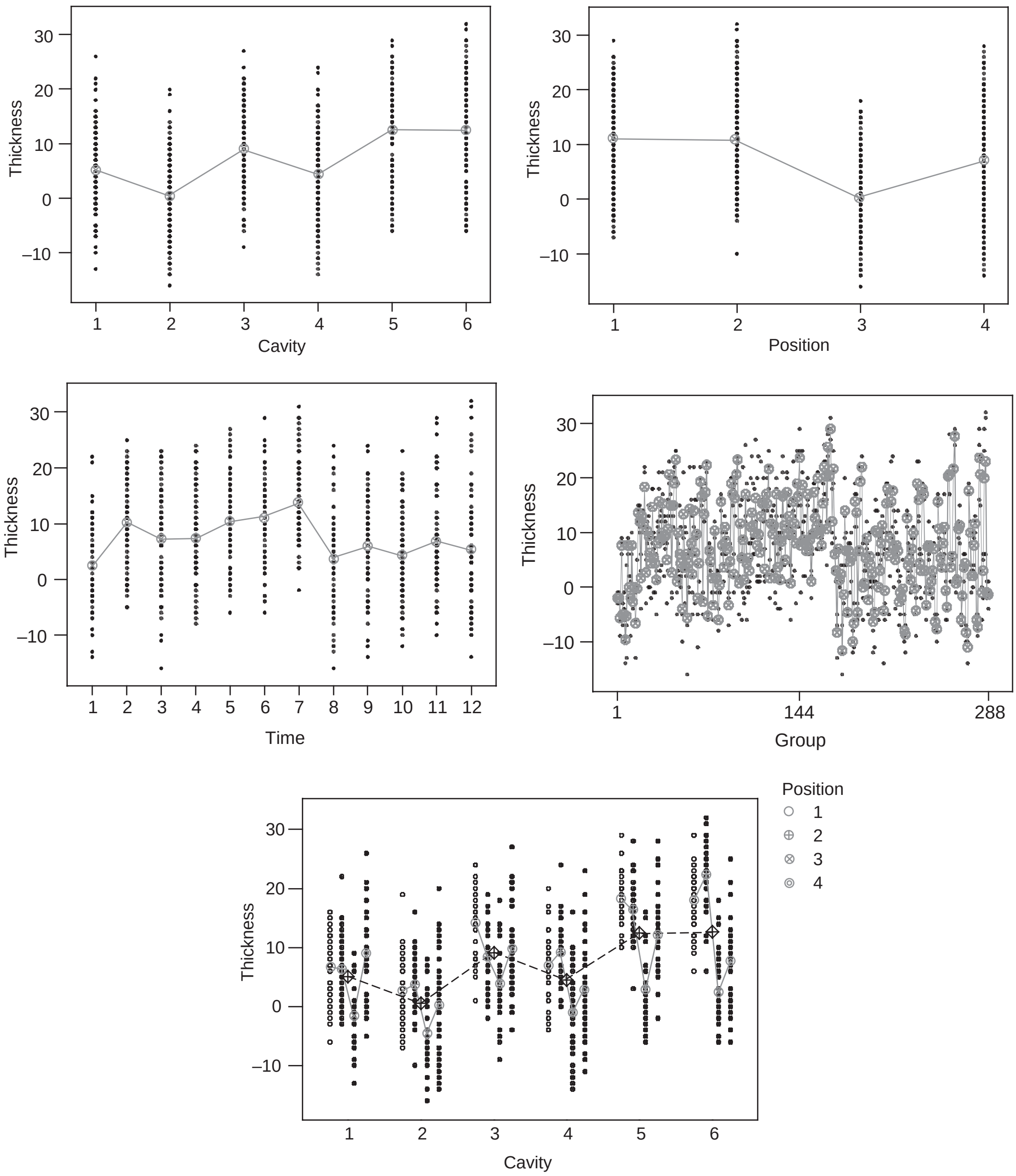


Figure S11.2 Multivari charts for casting thickness.

See Appendix D for details on how to fit the ANOVA model in MINITAB. We get the following results when we fit an ANOVA model with time, cavity, and position (the three inputs expected to have systematic effect) and all interactions among these three inputs.

```
ANOVA: thickness versus time, cavity, position

Factor     Type Levels Values
time      fixed     12      1      2      3      4      5      6      7
                            8      9     10     11     12
cavity    fixed      6      1      2      3      4      5      6
position  fixed      4      1      2      3      4


Analysis of Variance for thickness

Source                   DF          SS          MS        F       P
time                     11     9008.82      818.98    31.81   0.000
cavity                    5    16994.99     3399.00   132.00   0.000
position                  3    16697.24     5565.75   216.15   0.000
time*cavity              55     1544.02       28.07     1.09   0.311
time*position            33     7580.19      229.70     8.92   0.000
cavity*position          15     5363.43      357.56    13.89   0.000
time*cavity*position    165     1634.06        9.90     0.38   1.000
Error                   576    14832.00       25.75
Total                   863    73654.75
```

To compare the relative sizes of the families, we look at the sum of squares (SS) column. We see large, roughly equal-sized effects due to position, cavity, and error. This matches our conclusion from the multivari charts. The calculated sum of squares tell us approximately how much we could expect to reduce the total sum of squares if we could eliminate all variation due to the given family.

It is not easy to translate these sums of squares into estimates for the standard deviation attributable to each family, but there is a rough correspondence. Recall that we are looking for dominant causes. For example, if we eliminate cavity-to-cavity variation, the total sum of squares is reduced by 16994.99 to 56659.76 and the estimated overall standard deviation is reduced from 9.24 ($\sqrt{73654.75/863}$) to 8.10 ($\sqrt{56659.76/863}$).

The error sum of squares is relatively large at 14832. This sum of squares includes the effect of the casting-to-casting family (recall that the casting-to-casting family is expected to have haphazard effect) and all interactions between the casting-to-casting family and the other families. We look more at this family in the next section.

As there was no single dominant family, the team proceeded by trying to address each of the three large families separately.

Block Bore Diameter

We can also use an ANOVA when the volume of data in a multivari investigation is large. With large amounts of data, there are so many points on the multivari charts that they are difficult to interpret, especially when we are looking at interactions.

In an investigation to try to find the dominant cause of engine block bore diameter variation, the team had access to a large volume of data that was automatically collected on every block. The team sampled from the complete database to isolate the effects of several families. They selected diameter measurements from:

- Three consecutive blocks selected from the process each hour for 80 hours
- All eight bores for each block
- Three heights and two orientations at each height for each bore

We give the 11,520 observations in the file *block bore diameter multivari*. To reduce the number of families, we define a new input, called *position*, with six values, which labels the two different orientations at three heights. If the position-to-position variation is found to be large, we can look more closely at this family.

We find multivari charts, such as Figure S11.3, difficult to interpret because of the large number of time periods used in the investigation. The chart does show that there are several odd values about a third of the way through the data set that we should investigate further.

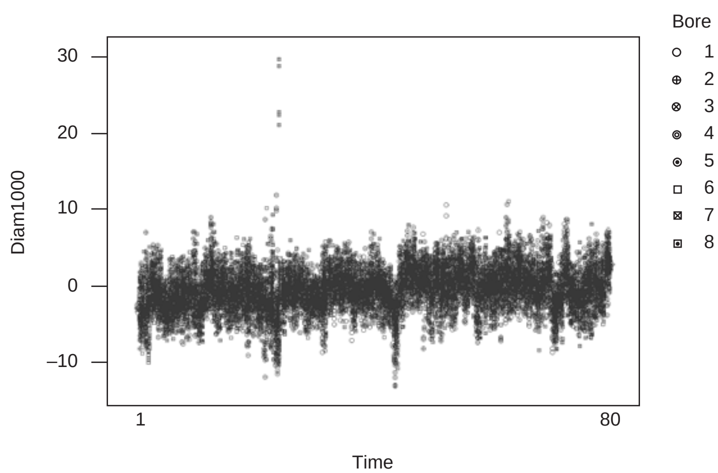


Figure S11.3 Multivari chart for bore by time.

To try to understand the results, we fit a full ANOVA model using all families (except block-to-block) and their interactions:

```
ANOVA: diam1000 versus time, bore, position

Factor     Type Levels Values
time      fixed     80      1      2      3      4      5      6      7
                            8      9     10     11     12     13     14
                           15     16     17     18     19     20     21
                           22     23     24     25     26     27     28
                           29     30     31     32     33     34     35
                           36     37     38     39     40     41     42
                           43     44     45     46     47     48     49
                           50     51     52     53     54     55     56
                           57     58     59     60     61     62     63
                           64     65     66     67     68     69     70
                           71     72     73     74     75     76     77
                           78     79     80
bore      fixed      8      1      2      3      4      5      6      7
                            8
position  fixed      6      1      2      3      4      5      6


Analysis of Variance for diam1000

Source                DF          SS          MS       F      P
time                  79    24602.93      311.43   56.34  0.000
bore                   7      951.65      135.95   24.60  0.000
position               5    10614.80     2122.96  384.09  0.000
time*bore            553    15568.13       28.15    5.09  0.000
time*position        395     5344.71       13.53    2.45  0.000
bore*position         35     4417.75      126.22   22.84  0.000
time*bore*position  2765     2935.01        1.06    0.19  1.000
Error               7680    42449.46        5.53
Total              11519   106884.44
```

The two largest sums of squares correspond to Error and time. We conclude that dominant causes of variation act in the block-to-block family (or in an interaction involving the block-to-block family) and the time-to-time family. The team proceeded to look for dominant causes within each of these families.

Camshaft Journal Diameter

In Chapter 11, we found a large grinder-to-grinder effect in the camshaft journal diameters. Looking at the part-to-part family with the multivari charts given in Figure 11.13 is difficult due to the large number of possible values for the input group that indexes the 192 sampling points. Here we fit an ANOVA model to the diameter data with inputs grinder, batch, hour, and position, and all possible interactions among these inputs. This model allows us to quantitatively assess the relative sizes of the effects due to each of the families and interactions among the families.

```
ANOVA: diameter versus grinder, batch, position, hour

Factor     Type Levels Values
grinder   fixed      2      A      B
batch     fixed      3      1      2      3
position  fixed      8     1f     1r     2f     2r     3f     3r     4f
                           4r
hour      fixed      4      1      3      5      7


Analysis of Variance for diameter

Source                          DF         SS         MS       F      P
grinder                          1   11855.39   11855.39 1230.72  0.000
batch                            2       2.15       1.07    0.11  0.894
position                         7     298.55      42.65    4.43  0.000
hour                             3      12.40       4.13    0.43  0.732
grinder*batch                    2      22.54      11.27    1.17  0.311
grinder*position                 7      47.14       6.73    0.70  0.673
grinder*hour                     3      53.57      17.86    1.85  0.136
batch*position                  14     223.45      15.96    1.66  0.060
batch*hour                       6      85.25      14.21    1.48  0.184
position*hour                   21     238.67      11.37    1.18  0.261
grinder*batch*position          14      92.57       6.61    0.69  0.789
grinder*batch*hour               6     112.93      18.82    1.95  0.070
grinder*position*hour           21     153.74       7.32    0.76  0.770
batch*position*hour             42     487.06      11.60    1.20  0.180
grinder*batch*position*hour     42     300.26       7.15    0.74  0.886
Error                          768    7398.07       9.63
Total                          959   21383.75
```

From the ANOVA results, we see (as expected) that the largest sum of squares is associated with the family of causes that differ between the two grinders. The sums of squares for any other systematic family (that is, batch, hour, position, and interactions) are much smaller. The error sum of squares, representing the part-to-part family, is also fairly large. We explore the part-to-part family further in the next section.

Comments

We can use a random-effects ANOVA model that ignores the systematic nature of the effects as an alternative analysis. See Feder (1974) for a background on random-effects models. However, it is incorrect to assume a random effect for a family like position in the casting thickness example since there are only four positions in total. The random-effects

ANOVA model provides estimates of the variance components attributable to the various families. In other words, the random-effects model provides estimates of *stdev*(due to a particular family).

In the casting thickness example, the results from the ANOVA model and the multivari charts match closely. The ANOVA has the advantage of quantifying the contributions of the various families and, in one step, allowing us to look at interactions among the families. However, note that ANOVA models are looking for variation caused by differences in average output for different level of the inputs. They are not good at detecting other patterns such as changes in output variation across different levels of the inputs or outliers. We recommend first looking at graphical summaries of the data such as multivari charts before fitting an ANOVA model.

If there are no families included in the multivari investigation that can be expected to have a haphazard effect, the proposed analysis strategy where we fit an ANOVA model with all inputs and interactions will not work in MINITAB. Although the appropriate sums of squares (which we use to draw conclusions) could still be calculated, MINITAB will not proceed because there are no degrees of freedom left to estimate the error sum of squares. If there are no haphazard families, we recommend leaving out one of the families in the ANOVA model. The error sum of squares is then attributable to the left-out family plus all interactions between the left-out family and other families in the model. If the error sum of squares is large, we redo the analysis leaving out a different family, continuing until we find the home of a dominant cause.

S11.2 HANDLING FAMILIES EXPECTED TO HAVE A HAPHAZARD EFFECT

For large data sets or multivari investigations with many families, we have difficulty displaying the effect of families, such as part-to-part, that are expected to have a haphazard rather than systematic effect. To explore the variation due to a haphazard family, we define a new input that uniquely numbers the groups of consecutive parts. For instance, in the cylinder head example from Chapter 11 we defined group as (time – 1)*4 + pattern. Similarly, in the fascia cratering example we used group = (time – 1)*2 + (machine – 1).

We illustrate the analysis to explore haphazard families with two examples.

Casting Thickness

The casting thickness example was introduced in the previous section. We expect the effect of the casting-to-casting family to be haphazard. We define a new input group that numbers the 288 different sets of consecutive thickness measurements from each different combination of time, cavity, and position—that is, we use group = (time – 1)*24 + (cavity – 1)*4 + position. Fitting a one-way ANOVA model with group gives:

```
One-way ANOVA: thickness versus group

Analysis of Variance for thickness
Source     DF        SS        MS        F        P
group     287   58822.7     205.0     7.96    0.000
Error     576   14832.0      25.7
Total     863   73654.7

                                       Individual 95% CIs For Mean
                                       Based on Pooled StDev
Level       N      Mean     StDev   --+---------+---------+---------+----
  1         3    -2.000     1.000         (---*---)
  2         3    -2.000     1.000         (---*---)
  3         3    -5.667     3.512      (---*---)
        .
        .
        .
286         3    23.000    14.731                        (---*---)
287         3    -1.000     1.732         (---*---)
288         3    -1.333     2.517         (---*---)
                                    --+---------+---------+---------+----
Pooled StDev =    5.074
```

The ANOVA divides the overall variation (as measured by the total standard deviation $\sqrt{73654.7/863} = 9.24$) into two components.

Pooled StDev is an estimate of the variation attributable to the casting-to-casting family and all interactions between the casting-to-casting family and other families. In other words, if we could eliminate all group-to-group differences due to changes in time, cavity, and position, we estimate that the process standard deviation would be reduced to 5.074. That is, there is a dominant cause acting in the group-to-group family. We examined the components of this family (time, cavity, and position) in the first section of this supplement.

Here, the pooled standard deviation is relatively small. In other cases, there may be a dominant cause in the part-to-part family. To demonstrate the methodology, suppose there is a dominant cause in the casting-to-casting family and we explore the casting-to-casting family further. We define a new output characteristic, the within-group standard deviation for each group. We then repeat the multivari analysis using the new output *group stdev*. To do this we need to define new input characteristics that correspond to the levels of the other families associated with each group. In the example, we define *stime*, *scavity*, and *sposition*, which correspond to the time, cavity, and position associated with each group, respectively. See the discussion on multivari charts in Appendix C for details on how to do this easily with MINITAB.

In the multivari chart, we are looking for any systematic change in the casting-to-casting variation over the other families. Note that these changes reflect an interaction with the casting-to-casting family. Looking at the multivari charts in figures S11.4 and S11.5 we conclude that the largest effect is due to a time-by-cavity interaction. At some particular

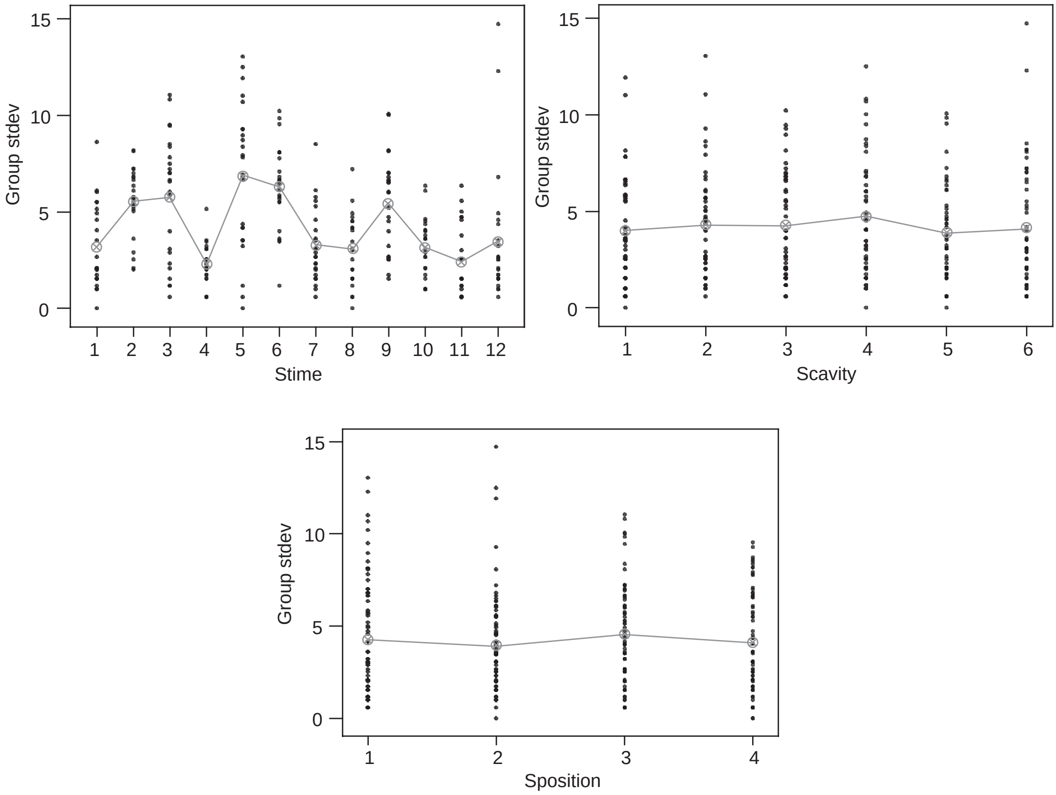


Figure S11.4 Multivari charts for casting thickness group standard deviation.

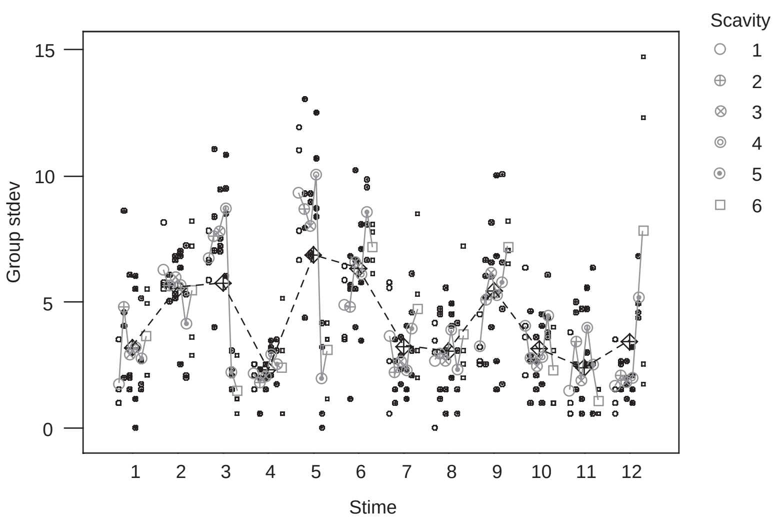


Figure S11.5 Multivari chart for *group stdev*, showing cavity by time interaction.

times and for some cavities, the variation in thickness for three consecutive castings changes substantially. We are looking for a cause that can explain such process behavior.

The interpretation of the results is difficult because the output is already a measure of variation. We see that for some times (for example, time 4) the variation in each group is small relative to other times. Also, for some times (for example, times 3, 5, and 12) the variation within a group varies for different cavities, while at all other times the variation within a group is similar for all cavities. From the interaction plot in Figure S11.5, we see that at some times, cavities 5 and 6 result in either more or less variation within a group than the other cavities. This matches the conclusion from fitting an ANOVA model with all but the three-input interaction term, as given:

```
ANOVA: group stdev versus stime, scavity, sposition

Factor     Type Levels Values
stime     fixed     12     1     2     3     4     5     6     7
                           8     9    10    11    12
scavity   fixed      6     1     2     3     4     5     6
sposition fixed      4     1     2     3     4


Analysis of Variance for group stdev

Source              DF         SS         MS       F      P
stime               11    705.022     64.093   22.19  0.000
scavity              5     22.070      4.414    1.53  0.184
sposition            3     14.738      4.913    1.70  0.169
stime*scavity       55    715.179     13.003    4.50  0.000
stime*sposition     33    288.226      8.734    3.02  0.000
scavity*sposition   15     70.297      4.686    1.62  0.073
Error              165    476.570      2.888
Total              287   2292.102
```

The conclusion is that the dominant cause acts casting to casting, but its effect varies with cavity and time. The team now must ask, "What causes vary in a way that matches this pattern?"

Camshaft Journal Diameter

The camshaft journal diameter example was introduced in Chapter 11 and also discussed in the previous section of this supplement. In the multivari analysis, we found the part-to-part variation was large, though not as large as the grinder-to-grinder variation. Here, we explore the part-to-part variation further by calculating the standard deviation within each group of parts by position, grinder, and batch. We define group = (grinder number – 1)*96 + (batch – 1)*32 + (hour–1)*8/2, since grinders *A* and *B* are coded 0 and 1, there are three batches, and the possible values for hour are 1, 3, 5, and 7. We then plot multivari charts using the within-group standard deviation as the output.

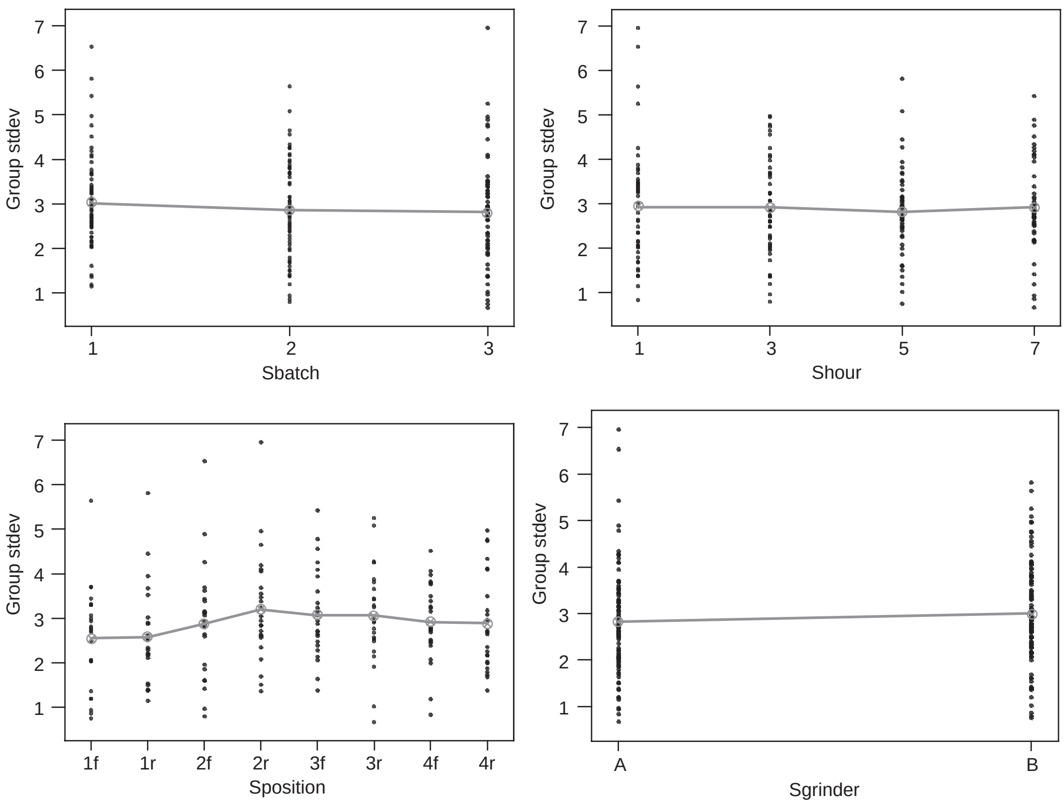


Figure S11.6 Multivari chart for *group stdev* for camshaft journal diameter multivari.

We conclude from Figure S11.6 that there is no evidence of any effects that change the within-group standard deviation. The dominant cause of the part-to-part variation acts in a similar way over all grinders, batches, and times.

S11.3 REGRESSION ANALYSIS FOR VARIATION TRANSMISSION INVESTIGATIONS

In Chapter 11, we analyzed the results of a variation transmission investigation using scatter plots. In most cases we find this graphical analysis sufficient. We can, however, supplement the graphical analysis with a numerical analysis based on regression models. Depending on the observed pattern in the scatter plots, we can use a variety of forms for the model. If the relationship appears roughly linear, as in Figure S11.7 which shows the V6 piston diameters after Operations 270 and 310, we can use a model of the form:

$$\text{diameter after OP310} = b_0 + b_1 * \text{diameter after OP270} + \text{residual},$$

where b_0 and b_1 are unknown model parameters that represent the y-intercept and slope, respectively, of a straight line that summarizes the relationship between the two diameter measurements. The parameters b_0 and b_1 are estimated based on the available data.

Fitting a regression model (see Appendix E) to the data gives:

```
Regression Analysis: diameter after OP310 versus diameter after OP270

The regression equation is
diameter after OP310 = 64.3 + 0.884 diameter after OP270

Predictor        Coef     SE Coef          T        P
Constant        64.27       22.29       2.88    0.005
diameter      0.88358     0.03739      23.63    0.000

S = 1.224       R-Sq = 85.6%     R-Sq(adj) = 85.4%

Analysis of Variance

Source           DF          SS          MS        F        P
Regression        1      836.67      836.67   558.58    0.000
Residual Error   94      140.80        1.50
Total            95      977.47
```

The estimated model parameters for b_0 and b_1 are given as 64.3 and 0.884, respectively. The estimated slope parameter (0.884) can be interpreted as the change in the average diameter after Operation 310 for every unit change in the diameter after Operation 270. Often, as in this example, the y-intercept parameter estimate will have no useful interpretation, since it tells us the average diameter after Operation 310 if the diameter after Operation 270 is zero. The regression equation is a straight line that best summarizes the relationship between the piston diameters after Operation 270 and Operation 310. As we see in Figure S11.7, in this case the relationship is strong. The value of $s = 1.224$ from the MINITAB output is an estimate of the standard deviation in the final diameter if we hold the diameter after

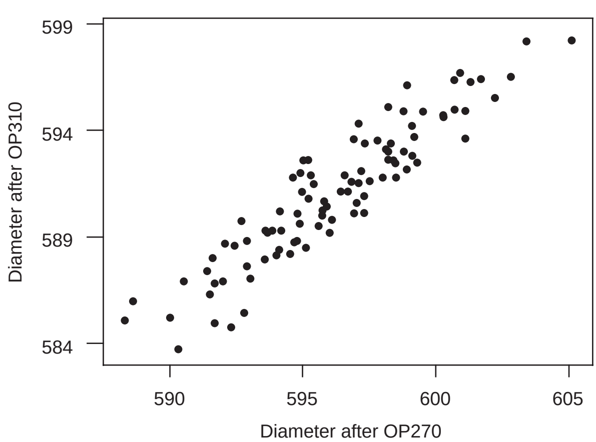


Figure S11.7 Plot of final diameter versus diameter after Operation 270.

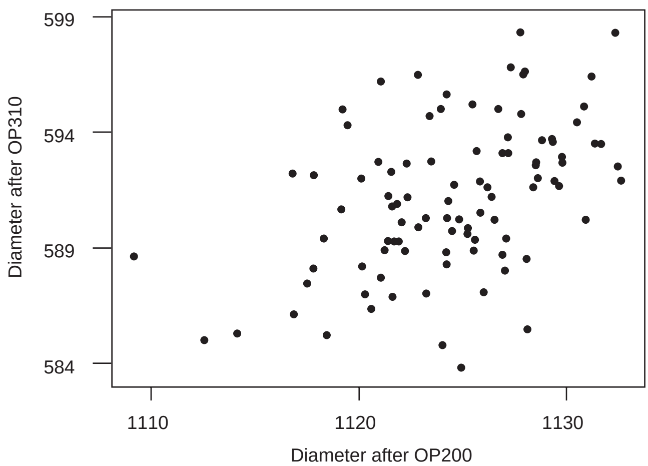


Figure S11.8 Plot of final diameter versus diameter after Operation 200.

Operation 270 fixed. We can compare this standard deviation to the baseline value 3.32 to determine that the input (in this case the diameter after Operation 270) is a dominant cause.

In the same example, we can also fit a regression model using the diameter after Operation 200 as the input. From MINITAB, we get the plot in Figure S11.8 and the following numerical results:

```
Regression Analysis: diameter after OP310 versus diameter after OP200

The regression equation is
diameter after OP310 = 225 + 0.325 diameter after OP200

Predictor        Coef     SE Coef          T        P
Constant       225.12       72.72       3.10    0.003
diameter      0.32542     0.06467       5.03    0.000

S = 2.862       R-Sq = 21.2%     R-Sq(adj) = 20.4%
```

From these results, it is clear that the relationship between the diameter after Operation 200 and the final diameter (after Operation 310) is much weaker than between the diameter after Operation 270 and the final diameter. In the second regression model, the estimate of the residual variation, $s = 2.862$, is close to the baseline standard deviation 3.32. If we were to hold the diameter at Operation 200 fixed, we would see little change in the variability after Operation 310. We can rule out all causes that act up to and including Operation 200 as possible dominant causes of final diameter variation.

Comments

In the context of variation transmission investigations, we do not recommend fitting a regression model to try to explain the final output as a function of two or more intermediate output measurements. Strong correlation between the inputs in the regression model (that is, the intermediate diameter measurements) makes interpretation of the results difficult. The residual variation (given by *s* in the MINITAB results) would estimate the residual standard deviation if we could hold all the inputs fixed.

We use regression models in Chapter 12 and discuss them further in the supplement to Chapter 12.

S11.4 OTHER COMPONENT-SWAPPING INVESTIGATIONS

In Chapter 11, we propose what we believe is the preferred plan, in most circumstances, for comparing component families. Dorian Shainin first proposed an alternative component-swapping plan in 1956. More recently the method has been promoted and discussed by Shainin and Shainin (1988a), Amster and Tsui (1993), Ingram (1993), and Bhote and Bhote (2000). Cotter (1979) provides some motivation for the Shainin type of component-swap experiments.

We consider the four-stage Bhote and Bhote (2000) version here. The four stages are confirmation, elimination, capping, and analysis. There are complicated rules for dealing with all possible eventualities that we do not repeat here.

To start, we choose two units with extreme and opposite performance. In the confirmation stage we check that repeated disassembly and reassembly does not affect performance. We use this stage to confirm that a dominant cause lies in the component rather than the assembly family of causes. This matches the procedure described in Section 10.4.

We start the elimination stage by ranking the components in descending order of likely importance. Then, we swap components between the two extreme parts, one at a time, starting with the most likely. After each swap, we determine the performance of the two assemblies. If the low and high output assemblies remain unchanged, we swap back the particular component and move on to swap the next most likely component. If the performance of the two units follows the swapped component, we have found the component family that is the home of a dominant cause and we can move to the final stage. If the performance partially moves with swapped component, we continue swapping with that particular component identified as part of a dominant cause involving two or more inputs.

When a dominant cause involving two components has been tentatively identified, we move to the capping stage. The capping run involves simultaneously swapping the two components identified as important. If the performance follows the double swap, the dominant cause is an interaction between two inputs with one in each component.

The final stage is called the analysis stage. In the analysis stage, no more swaps are conducted. The existing results are summarized graphically in 2 ¥ 2 tables to determine the magnitude and direction of the effects.

Comments

With the Bhote and Bhote (2000) procedure, the length of the search depends on engineering judgment regarding which components are likely to be most important. Also, Amster and Tsui (1993) provide some examples of when the Bhote and Bhote method would yield incorrect conclusions.

As suggested by Tippett (1934), Taguchi (1987), and Parmet and Steinberg (2001), we could also conduct a component-swapping experiment without using the idea of leveraging; that is, by choosing components from regular production. This strategy requires considerably more pairs of parts and a greater number of disassemblies and reassemblies.

The Bhote and Bhote (2000) component swap procedure matches the factorial component swap experiment as described in the next section, where we do not conduct all swaps once we believe the results are clear.

We believe iteratively forming two groups of components and using the method of elimination is a more efficient strategy (see Section 11.3).

S11.5 COMPONENT-SWAPPING WITH THREE GROUPS OR SUBASSEMBLIES

Here we look at component swap investigations with three groups. This may be useful if the assembly naturally divides into three subassemblies. We suppose for the moment that the three groups make up the entire assembly and that we have not yet eliminated any components as the home of the dominant cause.

We consider a component swap investigation as a factorial experiment, regardless of the number of groups. See Chapter 13 for a full discussion of factorial exeriments.We label the groups of components based on whether they come from the assembly that originally gave a low or high output. For instance, a component swap investigation with two groups, as presented in Chapter 11, corresponds to a (full) factorial experiment with two inputs, two levels for each input, and four treatments:

Treatment	G1	G2
1	Low	Low
2	Low	High
3	High	Low
4	High	High

Treatments 1 and 4 correspond to the two original assemblies and treatments 2 and 3 correspond to swapping the components in G1 (or G2) between the two assemblies.

With three groups of components, we disassemble and reassemble to get all eight possible combinations:

Treatment	G1	G2	G3
1	Low	Low	Low
2	Low	Low	High
3	Low	High	Low
4	Low	High	High
5	High	Low	Low
6	High	Low	High
7	High	High	Low
8	High	High	High

Treatments 1 and 8 correspond to the original assemblies. Since a swap of components corresponds to two runs of the experiment, the proposed plan is equivalent to conducting three swaps.

We can summarize the results of component swap experiments using two 2 ¥ 2 tables, similar to those shown in Chapter 11. The eight cells correspond to the eight treatments.

Had we used three groups in the power window buzz example described in Chapter 11, we would have to start:

	G2	
G1	**High**	**Low**
High	7	
Low		
G3 High		

	G2	
G1	**High**	**Low**
High		
Low		1
G3 Low		

We fill the other six cells of the table with the results from three swaps. We might get:

	G2	
G1	**High**	**Low**
High	7	1
Low	7	2
G3 High		

	G2	
G1	**High**	**Low**
High	7	1
Low	6	1
G3 Low		

Here, since the results in the two tables are similar, we can eliminate all the components in G3 and interpret the result in either table as in Chapter 11. In this example, the dominant cause acts in G2.

To help us make the correct interpretation, we can generate three pairs of tables similar to the previous set using a different group to stratify the pair of tables. If none of the three pairs of tables gives similar results across the two 2 ¥ 2 tables, the dominant cause involves a component (or more than one) in each of the three groups. This is unlikely.

Chapter 12 Supplement

Investigations Based on Single Causes

S12.1 MATRIX SCATTER PLOTS AND DRAFTSMAN PLOTS

To analyze the results of a group comparison investigation, we create plots of each of the inputs stratified by the two possible values of the output. This can be done using a series of box plots, where we select the option to plot the individual data points. To look simultaneously at two inputs, we create scatter plots of one input versus the second using a different plotting symbol for each of the two groups as defined by the output value. The number of such plots can be large.

In examples such as the window leak problem described in Chapter 12, where the number of inputs is less than 10, we can use a matrix scatter plot (see Appendix C) of all pairs of the (continuous) inputs. We give such a plot in Figure S12.1, where leakers are denoted by circles and nonleakers by plus signs. The plot in the right panel of Figure 12.2 is in the lower left corner of the matrix scatter plot in Figure S12.1. If the number of plots in the matrix scatter plot is too large, we find it difficult to see any patterns.

In input/output investigations, we want to examine scatter plots of the (continuous) output versus each of the inputs. In MINITAB, we can use a draftsman plot (see Appendix C) to automatically create all the desired scatterplots. Consider the crossbar dimension example described in Chapter 12. Figure S12.2 gives the draftsman plot with crossbar dimension and the five inputs. The left panel of Figure 12.3 is given in the middle of the draftsman plot.

With a large number of inputs, the aspect ratio of the draftsman plot makes interpretation difficult.

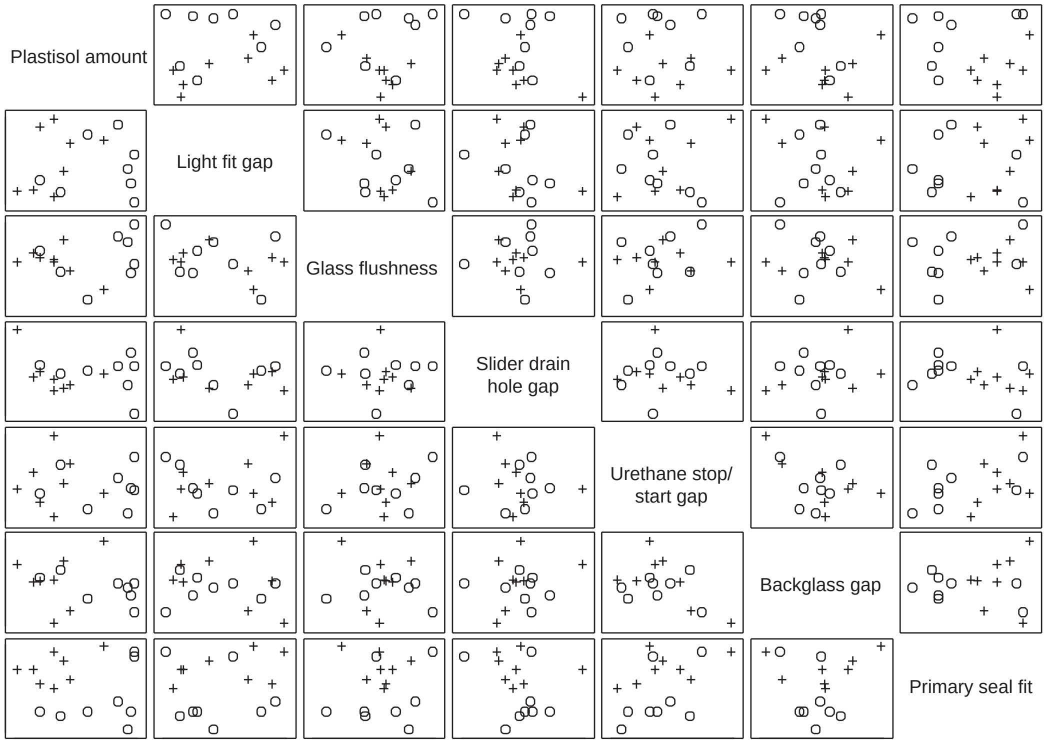


Figure S12.1 Matrix scatter plot for window leaks group comparison (leakers are denoted by o and nonleakers by +).

In output/input investigations with large numbers of inputs, examining all the scatter plots can be tedious, even when using a draftsman plot. In this context, another useful tool in MINITAB is best subsets regression (see Appendix E). With best subsets regression, we can ask MINITAB to fit all possible regression models involving a single input (or larger numbers of inputs, but that is not needed in this context). The results rank the possible models by the estimated residual standard deviation *s*. For illustration, results from using the best subsets regression routine for the crossbar dimension example are:

```
Best Subsets Regression: dimension versus die temp, nozzle temp, ...

Response is dimension

                                                d n b h c
                                                i o a y a
                                                e z r d v
                                                  z r r i
                                                t l e a t
                                                e e l u y
                                                m     l
Vars   R-Sq    R-Sq(adj)       C-p          S   p t t i p
```

```
1   78.9         78.3       2.9   0.25439      X
1   14.8         12.6     121.0   0.51123          X
1    8.2          5.8     133.1   0.53061        X
1    0.8          0.0     146.9   0.55174    X
1    0.1          0.0     148.1   0.55353  X
```

We see that barrel temperature is a dominant cause, because the residual standard deviation for the model with barrel temperature is so small relative to the baseline value of 0.46.

In the crossbar dimension example, there are only five inputs. Using best subsets regression is unnecessary, since we get the same information from Figure S12.2. One danger with relying on the best subsets regression is that the models are ranked according to how well a linear model fits the data. As a result, nonlinear relationships may be missed. We strongly recommend that you look at all of the scatter plots of the output versus the selected inputs.

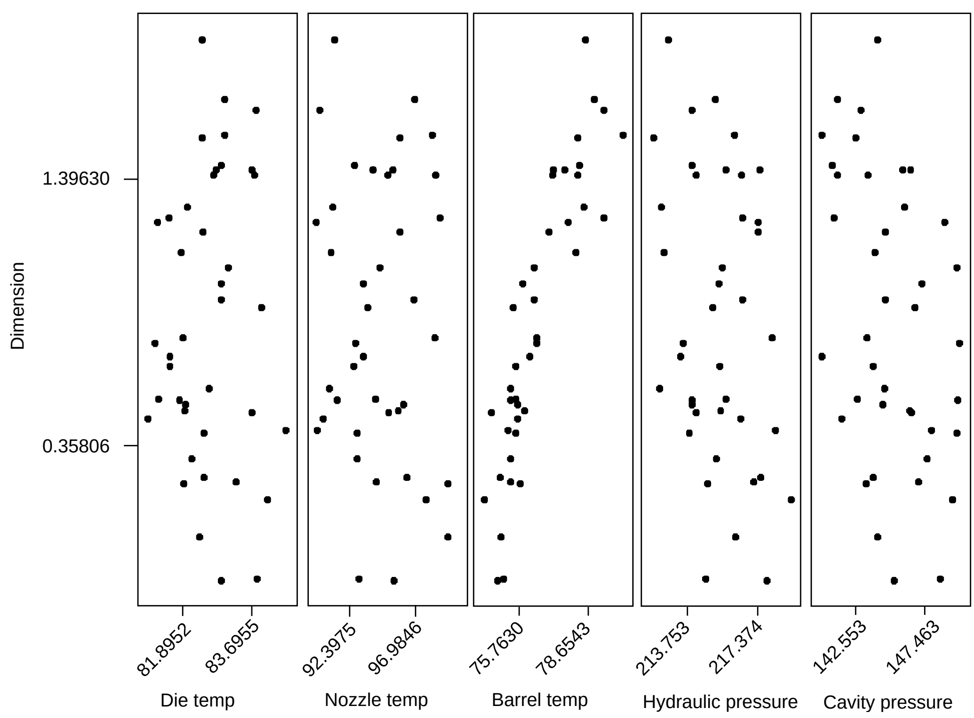


Figure S12.2 Draftsman plot for crossbar dimension regression investigation.

S12.2 GROUP COMPARISON VERSUS PAIRED COMPARISON

Bhote and Bhote (2000) promote a plan and analysis tool they call *paired comparison,* sometimes referred to as *group comparison.* The tool is similar to the proposed group comparison method discussed in Section 12.1. Like group comparison, the goal of paired comparison is to compare the effects of a number of inputs for parts with binary output.

There are some important differences in the two methods. First, in the paired comparison, a hypothesis test based on end-count is recommended to compare the two groups of parts. As discussed in the supplement to Chapter 10, we do not believe that formal hypothesis tests are necessary or appropriate when searching for a dominant cause of variation. Significant differences do not (necessarily) correspond to dominant causes.

A second difference is that paired comparison involves an arbitrary pairing of the parts, one from each group. In the search for the dominant cause, pairing makes little sense. Since the causes of variation are unknown, determining how to pair is problematic. In addition, the goal of the comparing extreme parts is to help identify a dominant cause of variation in the output. If we pair in such a way that a dominant cause is held constant within the pairs, the analysis will fail to identify the dominant cause. With artificial pairing, the analysis results depend on the way pairs are produced.

We do not recommend paired comparison as an alternative to group comparison.

S12.3 REGRESSION EXTENSIONS

Regression is a flexible analysis tool that can be extended and generalized in many ways. We consider a few useful extensions here.

Regression models can be used to analyze the designed experiments we encounter in Chapter 13. They are especially useful if, for example, we lose a run and the resulting experimental data are not balanced. The general linear model ANOVA option in MINITAB is based on fitting a regression model. Regression models can also be useful when assessing the feasibility or implementing a feedforward control.

Multiple Regression Model

We can fit regression models with more than one input, although in a search for a dominant cause, we do not, as a general rule, look at inputs simultaneously. If we think there is a dominant cause that involves two or more inputs, we may fit a model with all inputs or all pairs of inputs using best subsets regression. We may also add terms to a model if we suspect there is an nonlinear input/output relationship, as in the manifold sand scrap example in Chapter 12.

We are interested in understanding the remaining output variation if the input or inputs in the regression model are held fixed. This is given by the residual standard deviation. Whenever we consider including more than one input in a regression model, we need to consider the form of the model. We call this *model building*. An experienced analyst should conduct regression model building. Modeling assumptions can be checked using residual plots based on the estimated residuals from a regression model (Montgomery et al., 2001; Box et al., 1978). The estimated residuals are defined as the difference between the observed output value and the value predicted by the regression model.

Categorical Inputs and Indicator Variables

Suppose we have a categorical input x, such as a machine with three values: A, B, C or 0, 1, 2. We cannot use this input directly in a regression model. Instead, we replace x with two indicator variables. Let $x_B = 1$ if the part is from machine B; otherwise, let $x_B = 0$. Also, let $x_C = 1$ if the part is from machine C; othewise, let $x_C = 0$. We know that if $x_B = 0$ and $x_C = 0$, then the part is from machine A. Now we include x_B and x_C in the regression model.

In general, we use $k - 1$ indicator variables if there are k categories. (See Montgomery et al., 2001, for more detailed discussion.)

Logistic Regression for Binary Outputs

Regression analysis can be relatively easily extended to handle binary outputs. Logistic regression is one example (Hosmer and Lemeshow, 2000). See the MINITAB regression menu. Logistic regression can be a useful alternative or complement to the graphical group comparison analysis described in Chapter 12. Logistic regression analysis may be helpful when the group sizes are large.

Other Regression Extensions

Another useful extension is regression with count data (Cameron and Trivedi, 1998). Both logistic regression and regression with count data are special cases of generalized linear models (McCullagh and Nelder, 1989; Hamada and Nelder, 1997).

Chapter 13 Supplement

Verifying a Dominant Cause

S13.1 UNDERSTANDING REPEATS AND REPLICATES

There is great confusion about the difference between replicates and repeats. Replicates are different runs with the same treatment. Repeats are different parts within a run. The measured output values on repeats vary due to causes that act within a run. Between replicates, the measured output values vary because of causes that act both within runs and from run to run.

In traditional applications of experimental design, the experimenters do not know the within-treatment variation and hence need an estimate from within the experiment itself. If there are no replicates and they calculate an estimate based on the variation of the repeats within each run, they are almost certain to underestimate the within-treatment variation. As a consequence, they are likely to conclude a suspect cause is dominant when it is not.

The same problem occurs when we use an experiment to assess the effects of changing one or more fixed inputs in Chapter 15, where we search for an adjuster to move the process center.

We start the analysis in Chapter 13 (and in later uses of factorial experiments) by plotting the output values by treatment. To make this plot, we find it convenient to store the data with one row for each repeat. The plot includes output values from both replicates and repeats. We use the plots to look for large effects and promising treatments. We include horizontal lines showing the full extent of variation on the plot to get an assessment of the performance of the process with each treatment.

In the analysis for full factorial designs (and later fractional factorial designs—see Chapter 15), we use ANOVA to calculate, plot, and rank the effects of the inputs and their interactions. These calculations and plots are correct as long as the experimental design is (close to) balanced in terms of replicates and repeats (that is, there are the same number of runs per treatment and the same number of repeats per run) as we recommend. However, because of the way we have stored the data, the ANOVA program cannot separate replicates from repeats and hence the internal estimate of the within-treatment variation, derived from the residual variation, is too small. As a consequence, the formal hypothesis tests given in the output may to be misleading.

S13.2 RANDOMIZATION, REPLICATION, AND BLOCKING

Suppose we plan to verify a single suspect as a dominant cause. In making the decision to verify, we are implicitly or explicitly concerned about some other cause, identified or not, that acts in the same family as the suspect. We want to conduct the experiment to confirm the suspect and rule out any other possibility.

We say that other causes in the same family as the suspect are *confounded* with the suspect. That is, the pattern of behavior that we have seen in the output (in the observational investigations used in the search for a dominant cause) can be explained by the suspect or by some other member in the same family. In the verification experiment, we want to eliminate confounding.

Blocking is one tool that we can use to prevent confounding. If we hold the other members of the family fixed and change only the suspect, and if the output varies over its full extent, we know that the suspect is the dominant cause. For example, suppose the suspect is in the hour-to-hour family; that is, it changes relatively slowly. We can rule out other members of the family by changing the suspect quickly, say from one part to the next. From part to part we know that the other members of the hour-to-hour family are not changing much. This is blocking. By holding other possible causes fixed and varying the suspect, we can verify that the suspect is a dominant cause. Note that in other contexts blocking is used to increase the precision of the conclusions.

Sometimes we need more than blocking because we do not know how to hold all of the other causes in the suspect family fixed. For example, in the truck water leak example, a suspect was identified as an interaction between a gap and the plastisol application. These suspect inputs vary from truck to truck in a haphazard pattern. Many other inputs associated with any of the body components, the window, and the assembly vary in the same way. These causes had not been ruled out in the group comparison discussed in Chapter 12. We can use blocking to hold the body components fixed. That is, we use one truck and change the plastisol application and the gap with a different window. Using a single truck keeps all of the body components fixed. However, the causes in the assembly and window family change as we change the gap and plastisol levels.

We use randomization and replication to reduce the risk of confounding. For example, we select a number of trucks, say 10, and for each truck we change the plastisol and gap as described previously. Now we have 10 blocks. For each truck, we apply the four treatments in a random order. This random ordering over the 10 trucks will, with high probability, break the link between changes in the treatments and changes in other causes in the assembly family.

We use 10 different pairs of windows (one with a large gap and one with a small gap) to change the gap. There are other causes in the window family that match changes in the gap. We may be fooled in identifying the gap as a dominant cause when, in fact, there is some other characteristic of the window that is the cause. We accept this risk here because it is difficult to change the gap except by selecting the appropriate windows.

Randomization is likely to be effective only if there is sufficient replication—we recommend at least three runs at each level of the suspect and preferably five or more runs. Repeats are not helpful in reducing the risk of confounding, since their order cannot be randomized across runs.

S13.3 VARIABLES SEARCH

Bhote and Bhote (2000) highly recommend variables search for finding a dominant cause if there are five or more suspects. With fewer suspects, they recommend a full factorial experiment. Variables search competes with fractional factorial designs—see Chapter 15 and its supplement (also Shainin and Shainin, 1988).

With variables search, we conduct an experimental investigation using a procedure similar to component swap, as described in Chapter 11. We start with:

- A list of suspects ranked in order of expected importance
- Two levels for each suspect chosen at the extremes of their usual range of values
- Two treatment combinations that produce output that spans most of the full extent of variation

In the next stages, we swap levels of the suspects in a series of experimental runs starting with the highest-ranked suspect. The variables search method focuses on finding high-order interactions.

We cannot recommend variable search because:

- The length of the search depends on how well we order the suspects.
- All suspects not yet ruled out must be held fixed at one of the two levels for each run of the experiment. With a large number of suspects, all of which normally vary, this can be a daunting task.
- Substantial effort is required to find the initial two treatments that produce output that spans the full extent of variation. No direction is given on how to find these treatments.

See Ledolter and Swersey (1997a) for a critical view of variables search. We believe there is no good experimental plan when the list of suspects is long. Rather, we recommend continuing to use the method of elimination until the number of suspects is small.

Chapter 15 Supplement

Moving the Process Center

S15.1 FRACTIONAL FACTORIAL EXPERIMENTS

We give a brief description of the planning and analysis of fractional factorial experiments. We discuss only two-level fractional factorial designs and rely heavily on MINITAB. For a more complete description see Montgomery (2001); Box, Hunter, and Hunter (1978); or Wu and Hamada (2000). See Appendix F for details on how to set up the designs and conduct the analysis in MINITAB.

We use fractional factorial designs to estimate the effects of a number of inputs (usually four or more) simultaneously when the number of experimental runs is limited. In the language of factorial designs, inputs are called *factors* and the values used in the experiment are called the *levels* of the inputs. A *treatment* or a *treatment combination* is a set of particular levels, one for each input, that can be used to run the process.

We find it helpful to present the experimental plan using a code of –1 and +1 for the two levels of each input. To illustrate, consider the brake rotor balance verification experiment discussed in Chapter 13. This experiment was a full factorial design but we use it here to show how the coding works. We denote the inputs by the letters *A, B,* and *C,* as shown in Table S15.1.

We can represent any treatment combination using the code for each input. We list the eight possible treatments in Table S15.2. We call this plan a 2^3 design, because there are three inputs each at two levels and there are $2^3 = 8$ treatments used in the experiment.

Table S15.1 Input levels and coding for brake rotor verification experiment.

Input (factor)	Label	Low level (–1)	High level (+1)
Tooling	A	Old tooling	New tooling
Core position	B	Offset	Nominal
Thickness variation	C	30 thousandths	Nominal

Table S15.2 A list of all possible treatments in a 2^3 design.

Treatment	A	B	C
1	–1	–1	–1
2	–1	–1	+1
3	–1	+1	–1
4	–1	+1	+1
5	+1	–1	–1
6	+1	–1	+1
7	+1	+1	–1
8	+1	+1	+1

We generate the additional columns in Table S15.3 by multiplication. First we look at how this works and then we provide an interpretation. To get the AB column, for example, we multiply the corresponding plus or minus ones for each row in the table using the *A* and *B* columns. We repeat this calculation to get four additional columns corresponding to the products AB, AC, BC, and ABC. Note that we have put the columns in a standard order. We call the seven columns in Table S15.3 the *contrast matrix* for the 2^3 design. Taguchi (1987) calls this matrix an orthogonal array and labels it L_8.

We can calculate all the effects using the contrast matrix and the output from the experiment. In Table S15.4, we add one extra column that gives the output value for each treatment in the experiment. Here we assume there is a single run of the experiment corresponding to each treatment. With replication (that is, more runs), we replicate the rows in the contrast matrix. Note that the treatments in the contrast matrix are written in a specific order and the runs in the experiment likely occur in a different random order.

Table S15.3 Contrast matrix for the 2^3 design.

Treatment	A	B	AB	C	AC	BC	ABC
1	–1	–1	+1	–1	+1	+1	–1
2	–1	–1	+1	+1	–1	–1	+1
3	–1	+1	–1	–1	+1	–1	+1
4	–1	+1	–1	+1	–1	+1	–1
5	+1	–1	–1	–1	–1	+1	+1
6	+1	–1	–1	+1	+1	–1	–1
7	+1	+1	+1	–1	–1	–1	–1
8	+1	+1	+1	+1	+1	+1	+1

Table S15.4 Contrast matrix for the 2^3 design with output data.

Treatment	A	B	AB	C	AC	BC	ABC	Weight
1	–1	–1	+1	–1	+1	+1	–1	0.56
2	–1	–1	+1	+1	–1	–1	+1	0.17
3	–1	+1	–1	–1	+1	–1	+1	0.44
4	–1	+1	–1	+1	–1	+1	–1	0.08
5	+1	–1	–1	–1	–1	+1	+1	1.52
6	+1	–1	–1	+1	+1	–1	–1	0.37
7	+1	+1	+1	–1	–1	–1	–1	1.34
8	+1	+1	+1	+1	+1	+1	+1	0.03

Recall from Chapter 13 that the *main effect* of a particular input is the difference between the average output over runs at the high level and the average output over runs at the low level of the input. In the example, the main effect of *A* is

$$\frac{1.52+0.37+1.34+0.03}{4}-\frac{0.56+0.17+0.44+0.08}{4}=$$
$$\frac{-0.56-0.17-0.44-0.08+1.52+0.37+1.34+0.03}{4}=0.5025 \qquad \text{(S15.1)}$$

We have changed the order of the terms in the sum of the second expression in this calculation to match the treatment order in Table S15.4. In terms of the contrast matrix, to get the numerator of Equation (S15.1), we apply the signs from column *A* to the data (that is, multiply the two columns element by element) and add.

To find an *interaction effect,* we compare the main effects for one input at the two different levels of a second input. For example, to look at the AC interaction, we have

$$\text{when C} = +1\text{, the effect of A is } \frac{0.37+0.03}{2}-\frac{0.17+0.18}{2}\text{, and}$$

$$\text{when C} = -1\text{, the effect of A is } \frac{1.52+1.34}{2}-\frac{0.56+0.44}{2}$$

The interaction effect is half the difference. That is,

$$\frac{1}{2}\left\{\left[\frac{0.37+0.03}{2}-\frac{0.17+0.18}{2}\right]-\left[\frac{1.52+1.34}{2}-\frac{0.56+0.44}{2}\right]\right\}=$$
$$\frac{0.56-0.17+0.44-0.08-1.52+0.37-1.34+0.03}{4}=-0.4275 \qquad \text{(S15.2)}$$

Table S15.5 Contrast matrix for the 2^3 design with estimated effects.

Treatment	A	B	AB	C	AC	BC	ABC	Weight
1	–1	–1	+1	–1	+1	+1	–1	0.56
2	–1	–1	+1	+1	–1	–1	+1	0.17
3	–1	+1	–1	–1	+1	–1	+1	0.44
4	–1	+1	–1	+1	–1	+1	–1	0.08
5	+1	–1	–1	–1	–1	+1	+1	1.52
6	+1	–1	–1	+1	+1	–1	–1	0.37
7	+1	+1	+1	–1	–1	–1	–1	1.34
8	+1	+1	+1	+1	+1	+1	+1	0.03
Effect	0.503	–0.182	–0.078	–0.803	–0.428	–0.033	–0.048	

Again we have reordered the terms in the numerator of the second expression in Equation (S15.2) and you can see that to get this sum, we apply the signs from the AC column of the contrast matrix to the data and add.

All main effects and interactions can be found by applying the signs from the appropriate column of the contrast matrix to the data, adding, and then dividing by half the number of runs. Every effect is the difference of two averages, where each average includes exactly half the data. We show all of the effects in Table S15.5, where we add an extra row to Table S15.4.

In the analysis, we look for large (positive or negative) effects. For two-level factorial and fractional factorial designs, MINITAB will calculate and rank the absolute value of the effects in a Pareto diagram. We can use this diagram to isolate large effects. The effects in the example are plotted in Figure S15.1. We see that the main effects of *C* and *A* are relatively large, as is the interaction effect AC. Since there is evidence of interaction, we look at the effects of *A* and *C* simultaneously.

We now use the contrast matrix to explain fractional factorial designs. As an example, we start with the contrast matrix for a 2^4 design, that is, a design with four inputs each at two levels. We give the contrast matrix in Table S15.6. Note the order of the columns.

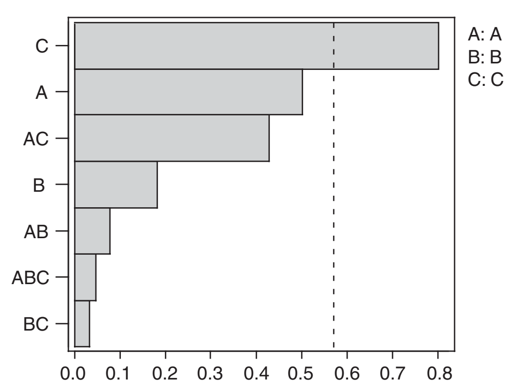


Figure S15.1 Pareto chart of effects in rotor balance experiment.

Table S15.6 Contrast matrix for a 2^4 design.

Treatment	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
1	–1	–1	+1	–1	+1	+1	–1	–1	+1	+1	–1	+1	–1	–1	+1
2	–1	–1	+1	–1	+1	+1	–1	+1	–1	–1	+1	–1	+1	+1	–1
3	–1	–1	+1	+1	–1	–1	+1	–1	+1	+1	–1	–1	+1	+1	–1
4	–1	–1	+1	+1	–1	–1	+1	+1	–1	–1	+1	+1	–1	–1	+1
5	–1	+1	–1	–1	+1	–1	+1	–1	+1	–1	+1	+1	–1	+1	–1
6	–1	+1	–1	–1	+1	–1	+1	+1	–1	+1	–1	–1	+1	–1	+1
7	–1	+1	–1	+1	–1	+1	–1	–1	+1	–1	+1	–1	+1	–1	+1
8	–1	+1	–1	+1	–1	+1	–1	+1	–1	+1	–1	+1	–1	+1	–1
9	+1	–1	–1	–1	–1	+1	+1	–1	–1	+1	+1	+1	+1	–1	–1
10	+1	–1	–1	–1	–1	+1	+1	+1	+1	–1	–1	–1	–1	+1	+1
11	+1	–1	–1	+1	+1	–1	–1	–1	–1	+1	+1	–1	–1	+1	+1
12	+1	–1	–1	+1	+1	–1	–1	+1	+1	–1	–1	+1	+1	–1	–1
13	+1	+1	+1	–1	–1	–1	–1	–1	–1	–1	–1	+1	+1	+1	+1
14	+1	+1	+1	–1	–1	–1	–1	+1	+1	+1	+1	–1	–1	–1	–1
15	+1	+1	+1	+1	+1	+1	+1	–1	–1	–1	–1	–1	–1	–1	–1
16	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1

Suppose we want to conduct an experiment and we can afford only 16 runs. With four inputs, *A, B, C,* and *D,* we can run all treatments. The four columns labeled *A, B, C,* and *D* give the levels of the four inputs for each treatment. If, however, we have a fifth input, *E,* there are 32 possible treatments and we have the resources to conduct an experiment using only half of the possible treatments. The question is, which half? The resulting design is often referred to as a 2^{5-1} design since we are using the half fraction (2^{-1}) design with five inputs at two levels each.

One possibility is to assign *E* to the column denoted ABCD in the contrast matrix in Table S15.6. Then, we can read the 16 treatments we will use in the experiment by looking at the five columns labeled *A, B, C, D,* and *E.* For example, with treatment 7, we have the levels *A:* –1, *B:* 1, *C:* 1, *D:* –1, *E:* 1. With this choice for *E,* exactly half the 16 runs have *E* at the high level, that is, $E = +1$, and the other half have *E* at the low level, that is, $E = -1$. The same is true for the other four inputs.

To see the consequences of this choice, suppose we carry out the experiment and measure the output for each of the selected 16 treatments. We can calculate the main effects of each of the five inputs by applying the appropriate column of plus and minus ones to the data, adding and dividing by eight. We carry out the same calculation for every column.

To calculate the effect of the four-input interaction ABCD and the main effect for *E,* we use the same column of the contrast matrix and we say these two effects are *confounded.* If the calculated effect is large, this may be due to the change in input *E* or the four-input interaction involving inputs *A, B, C,* and *D.* We cannot separate these two effects using the data. However, we usually assume that three- and four-input interaction effects are small and so, in this case, we would attribute the large effect to the main effect of *E.*

You should be wondering by now where to find the interactions involving *E.* To find the two-input interaction AE, we multiply the *A* and *E* columns as before. We found the column of signs for *E* by multiplying the corresponding columns for *A, B, C,* and *D.* We use the convenient notation $E = ABCD$. Hence we have

$$AE = A \text{ ¥ } ABCD = BCD$$

since, if we multiply column *A* by itself, we get a column of plus ones, which has no effect on the overall product. In other words, the interaction effect AE is confounded with the three-input interaction BCD. If this column produces a large effect, we cannot tell if this is due to AE or BCD. Again we will attribute the effect to the lower-order interaction—in this case, the two-input interaction AE.

You can quickly find all other two-input interactions involving *E.* Every column of the contrast matrix corresponds to two effects (since we used a half fraction design). MINITAB will produce a list of the confounded effects for the design. For the example, the list of confounded effects (MINITAB calls these effects aliases) is:

```
Design Generators:  E = ABCD

Alias Structure
```

```
I + ABCDE

A + BCDE
B + ACDE
C + ABDE
D + ABCE
E + ABCD
AB + CDE
AC + BDE
AD + BCE
AE + BCD
BC + ADE
BD + ACE
BE + ACD
CD + ABE
CE + ABD
DE + ABC
```

Any effects in the same row, linked by a + or –, are confounded. See Appendix F for instructions on how to create this list. You might also wonder what happens if we start by assigning *E* to a different column, say *E* = *ABC*. Since the ABC column gives the level of *E* for each run, we now have a different set of 16 treatments. We use MINITAB to produce the list of confounded effects. This time, we ask that four- and five-input interactions be suppressed. The confounded effects with this design are:

```
Design Generators:  E = ABC

Alias Structure (up to order 3)

I
A + BCE
B + ACE
C + ABE
D
E + ABC
AB + CE
AC + BE
AD
AE + BC
BD
CD
DE
ABD + CDE
ACD + BDE
ADE + BCD
```

Table S15.7 Design resolution interpretation.

Resolution	Meaning
III	Main effects are confounded with two-input interaction effects
IV	Main effects are confounded with three-input or higher-order interactions, and two-input interaction effects are confounded with other two-input interaction effects
V	Main and two-input interaction effects are confounded only with three-input or higher-order interactions

With this choice, we see that main effects are confounded with three-input interactions and, more important, pairs of two-input interactions such AB and CE are confounded. This plan is less desirable than the design based on assigning *E* to ABCD.

We say that the design with *E* = ABC has resolution IV, because at least one pair of two-input interactions is confounded. If we assign *E* to column AB, then a main effect is confounded with a two-input interaction and the design has resolution III. The first design discussed earlier with *E* = ABCD has resolution V. The higher the resolution, the less likely it is that important effects will be confounded. We summarize the meaning of *resolution* in Table S15.7.

For a fixed number of runs, the greater the number of inputs, the lower the highest possible resolution (see Figure S15.2).

In using MINITAB, we normally assign the letters A, B, C, and so on to the candidates and +1 and –1 to the two levels for each. It does not matter which letter is assigned to each candidate. Then we ask MINITAB to generate the design, a worksheet of the treatments selected for the experiment. Sometimes we want to ensure that one particular treatment is included in the experiment—for example, the current process levels of the candidates. In

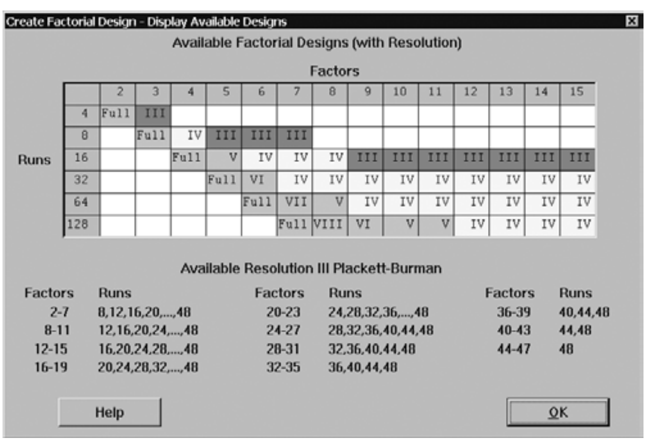


Figure S15.2 Two-level factorial experiments available in MINITAB.

this case, we generate the design first and then assign the letters and levels so that one of the treatments in the design has the required levels.

We recommend an analysis based on a plot of the output by treatment number, the Pareto plot of the effects, and main effect and interaction plots. MINITAB will produce the table of confounded effects to help interpret the important effects.

Despite the confounding, factorial experiments are useful because of the *scarcity of effects principle,* which states that when there are many inputs, there are likely only a few large effects. These effects are commonly the main effects and low-order interactions.

Resolving Confounded Effects

In fractional factorial experiments, all effects are confounded with at least one other effect. In some circumstances, to strengthen the conclusions from the experiment, we conduct additional runs with new treatments to break the confounding for important effects. This is especially necessary if the confounded effects are interactions of the same order. Simple follow-up experiments that use *fold-over* are covered in Montgomery (2001). A more sophisticated method is given in Meyer et al. (1996). Typical follow-up experiments involve 8 or 16 runs, depending on the number of inputs and degree of fractionation used in the initial design.

Use of Center Points

If the inputs are quantitative (that is, speed, voltage, amount, and so on), we have the option of adding center points to the design. The level of each input at the center point is the average of the high and low values. We illustrate the center point for a design with two inputs at two levels in Figure S15.3. MINITAB allows the addition of center points to factorial or fractional factorial designs with quantitative inputs.

When we add a center point to the design, we can check to see if there is curvature in the main effect of one or more of the inputs. The curvature may be important if we plan to use the input as an adjuster. The addition of the center point does not change the confounding of the effects in a fractional design. For more detail, see Montgomery (2001).

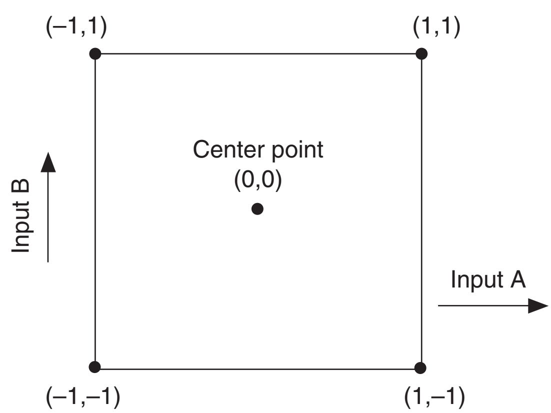


Figure S15.3 A 2^2 design with center point.

Summary

Most issues in the planning of fractional factorial experiments are the same as for full factorial experiments. We need to

- Choose the *k* inputs and their levels
- Define a run
- Use MINITAB to select the design (that is, choose the treatments, a fraction of all of the possibilities), mindful of the tradeoff between resolution and number of runs
- Randomize the run order

The choice of fraction is based on cost considerations and concerns about confounding. In fractional factorial experiments, every effect will be confounded with one or more other effects. We choose the design with the highest possible resolution to minimize problems associated with confounded effects. That is, we choose a design so that main effects and two-input interactions are confounded with high-order interactions as much as possible. If we find large effects, we typically attribute the effect to the main effect or two-input interaction.

MINITAB can help plan and analyze fractional factorial experiments. The complete confounding structure for any fractional factorial design is given by MINITAB. Your task is to assign the inputs and levels to the generic letters *A, B, C,* and so on, and the codes –1 and +1 to produce the design.

Chapter 16 Supplement

Desensitizing a Process to Variation in a Dominant Cause

S16.1 MATHEMATICAL REPRESENTATION OF PROCESS DESENSITIZATION

We can demonstrate process desensitization using a regression model. In the model given by Equation (S16.1), Y is the value of the output, X is the level of the dominant cause, c is the level of the desensitization candidate (here assumed to be quantitative), and the term R represents the (small, since X is a dominant cause) residual effect of all other varying inputs.

$$Y = b_0 + b_1X + b_2c + b_3Xc + R$$

which can be rewritten as

$$Y = b_0 + (b_1 + b_3c)X + b_2c + R \qquad \text{(S16.1)}$$

The levels of other fixed inputs determine the coefficient b_0. For a given level c of the candidate, the coefficient $b_1 + b_3c$ represents the effect of the cause X on the output. The standard deviation of the output is

$$sd(Y) = \sqrt{(b_1 + b_3c)^2 sd(X)^2 + sd(R)^2} \qquad \text{(S16.2)}$$

assuming that the effect of the other causes, R, varies independently of the dominant cause. In Equation (S16.1), we have modeled the interaction between the cause X and the candidate by the product term b_3Xc. We see from Equation (S16.2) that the standard deviation of the output *sd(Y)* is sensitive to the level of the candidate if b_3 is not equal to zero. If we can set $c = -b_1 / b_3$, then the effect of the dominant cause *sd(X)* is completely eliminated and *sd(Y)* is reduced to *sd(R)*.

In practice, such a simple model rarely describes the relationship amongst the inputs and output exactly. As well, we must estimate the coefficients b_1 and b_3 without error before we can achieve the optimal reduction of variation.

If the candidate is binary (for example, supplier 1 or 2), then *c* in Equation (S16.1) can take only one of two possible values—for example, –1 or +1—and we choose the level of the candidate with the smaller value of $(b_1 + b_3c)^2$.

S16.2 FRACTIONAL FACTORIAL EXPERIMENTS FOR DESENSITIZATION

In this section, we give a brief description of fractional factorial experiments specifically designed for desensitization of the output to the effects of a known dominant cause. See the supplement to Chapter 15 for more general information about fractional factorial designs.

We have two types of inputs: a dominant cause and the candidates. Taguchi calls these the noise and control factors, respectively. The main goal of the experiment is to look for interactions between the cause and the candidates.

We recommend that the two levels for the dominant cause be set at the extremes of the normal range of variation. If there is more than one dominant cause, then we suggest creating a pseudo-input with two levels to generate the full extent of variation in the output when the candidates are set at their original values. There is little need to use a full factorial design for the causes since we already know their effects. See the refrigerator frost buildup case in Chapter 16 for an example.

If there are four or more candidate inputs, we can use a two-level fractional factorial design to determine the treatments for the candidates. To combine the two types of inputs (cause and candidates) into a single design, we construct a *crossed design* with runs at both levels of the cause for each combination of the candidates.

As an example, suppose we have five candidates labeled *A* to *E* and a single dominant cause *X*, each at two levels. The total number of runs is the product of the number of combinations of the candidates times the number of levels of the dominant cause. Suppose we can carry out 16 runs in total. Since the dominant cause has two levels, we can have eight treatments. We use MINITAB to select the quarter fraction design of resolution III for the five candidates. The combinations are shown in Table S16.1 in a randomized order.

For the crossed design, we have two runs, one at the low level and one at the high level of *X* for each treatment, as shown in Table S16.2, where the asterisks correspond to the 16 runs of the experiment.

Table S16.1 Selected treatment combinations for a 2^{5-2} design.

Treatment	A	B	C	D	E
1	–1	+1	–1	–1	+1
2	–1	–1	+1	+1	–1
3	–1	+1	+1	–1	–1
4	+1	+1	+1	+1	+1
5	+1	+1	–1	+1	–1
6	+1	–1	–1	–1	–1
7	+1	–1	+1	–1	+1
8	–1	–1	–1	+1	+1

Table S16.2 Crossed design with 16 runs.

						Level of X	
Treatment	**A**	**B**	**C**	**D**	**E**	**Low**	**High**
1	–1	+1	–1	–1	+1	*	*
2	–1	–1	+1	+1	–1	*	*
3	–1	+1	+1	–1	–1	*	*
4	+1	+1	+1	+1	+1	*	*
5	+1	+1	–1	+1	–1	*	*
6	+1	–1	–1	–1	–1	*	*
7	+1	–1	+1	–1	+1	*	*
8	–1	–1	–1	+1	+1	*	*

In MINITAB, we need each row to correspond to a single run. We create the design by:

- Pasting a copy of the eight treatments into the next eight rows of the spreadsheet
- Adding a column for the dominant cause, denoted *X,* with eight –1s followed by eight +1s

We get the columns as shown in Table S16.3.

Table S16.3 Crossed design with 16 runs displayed in rows.

A	B	C	D	E	X
–1	+1	–1	–1	+1	–1
–1	–1	+1	+1	–1	–1
–1	+1	+1	–1	–1	–1
+1	+1	+1	+1	+1	–1
+1	+1	–1	+1	–1	–1
+1	–1	–1	–1	–1	–1
+1	–1	+1	–1	+1	–1
–1	–1	–1	+1	+1	–1
–1	+1	–1	–1	+1	+1
–1	–1	+1	+1	–1	+1
–1	+1	+1	–1	–1	+1
+1	+1	+1	+1	+1	+1
+1	+1	–1	+1	–1	+1
+1	–1	–1	–1	–1	+1
+1	–1	+1	–1	+1	+1
–1	–1	–1	+1	+1	+1

Next we use MINITAB to create a custom factorial design (see Appendix F) with the six inputs. We can get the confounding structure for this design before we collect the data by using a column of dummy output. Any set of 16 numbers will do. The given design has the following confounding structure, where we show only main effects and two- and three-input interactions:

```
Alias Structure (up to order 3)

I + ABD + ACE
A + BD + CE
B + AD + CDE
C + AE + BDE
D + AB + BCE
E + AC + BCD
X
AX + BDX + CEX
BC + DE + ABE + ACD
BE + CD + ABC + ADE
BX + ADX
```

```
CX + AEX
DX + ABX
EX + ACX
BCX + DEX
BEX + CDX
```

The crossed design has the advantage that all of the two-input interactions involving the dominant cause *X* and the candidates are not confounded with any other two-input interactions. This is a powerful motive for using the crossed design given that the goal of the experiment is to examine these interactions.

We recommend using a crossed design for desensitization experiments. There are other possible designs, perhaps using fewer runs. See Hamada and Wu (2000).

S16.3 FURTHER ANALYSIS FOR THE EDDY CURRENT MEASUREMENT EXAMPLE

In the eddy current example discussed in Chapter 16, the team concluded the measurement system was not reliable. They came to this conclusion by looking at plots of the eddy current hardness measurement versus the Brinell hardness for all the treatments. We gave an example plot in Figure 16.11. Had the results looked more promising, the team could have conducted further analysis. For instance, they could have fit a regression model of the form:

$$\textit{eddy current hardness} = a + b\,(\textit{Brinell hardness}) + \textit{residual}$$

for each of the eight treatments defined by the candidates. In the model, the residual represents the variation in the eddy current measurements not explained by Brinell hardness. A good treatment would have a large estimated slope b and a small residual standard deviation (given as s in the MINITAB regression results—see Appendix E). We use $|b/s|$ as the performance measure. To illustrate the analysis, for treatment 7 we get:

```
The regression equation is
Eddy Current Hardness = 16.3 - 2.60 Brinell

Predictor         Coef     SE Coef          T        P
Constant        16.347       2.041       8.01    0.000
Brinell        -2.5951      0.4544      -5.71    0.000

S = 0.2172      R-Sq = 34.5%     R-Sq(adj) = 33.4%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         1      1.5394      1.5394     32.62     0.000
Residual Error    62      2.9261      0.0472
Total             63      4.4656
```

Table S16.4 Eddy current desensitization experiment results.

Treatment	Frequency	Temp	Gain	Slope *b*	Residual variation *s*	Performance |*b*/*s*|
1	200	35	30	–0.545	0.104	5.2404
2	200	35	40	–0.948	0.188	5.0426
3	200	65	30	0.435	0.110	3.9545
4	200	65	40	–1.010	0.143	7.0629
5	350	35	30	–0.871	0.195	4.4667
6	350	35	40	–0.038	0.071	0.5352
7	350	65	30	–2.600	0.217	11.9816
8	350	65	40	–0.929	0.110	8.4455

The performance measure for treatment 7 is $|-2.5951/0.2172| = 12.0$. The data with the calculated performance measures are given in Table S16.4.

Treatment 7 gives the best performance. The Pareto plot of the effects given in Figure S16.1 summarizes the results across all treatments. The temperature main effect and the temperature by frequency interaction are the largest effects, but nothing stands out.

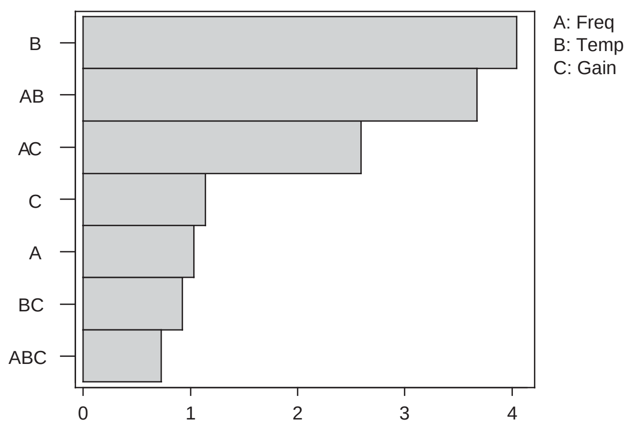


Figure S16.1 Pareto plot of the effects based on the performance measure.

We can also look at promising treatments in more detail by plotting the eddy current by Brinell hardness stratified by the four levels of the cause as defined by:

Cause combination	Day	Cleaning time
1	1	5
2	2	5
3	1	19
4	2	19

From Figure S16.2 for treatment 7, we see that the eddy current measurement system gives inconsistent output across the different levels of the cause. In addition, it does not work well for any cause combination, even well cleaned parts. We also see that the average Brinell hardness of the parts differed by day. This suggests the chemistry was indeed different across the two days.

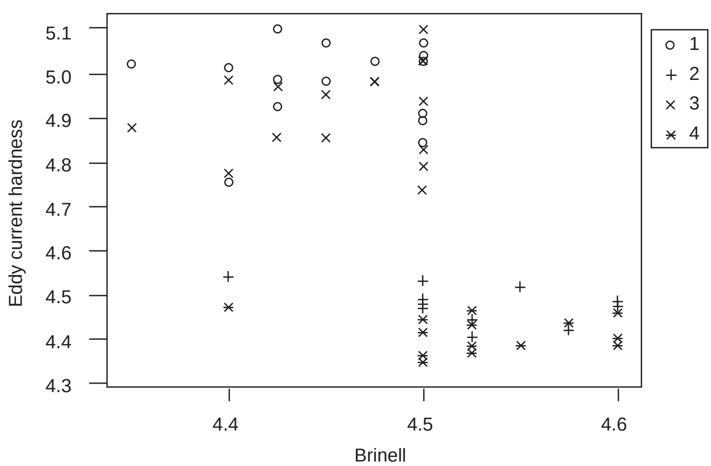


Figure S16.2 Scatter plot of eddy current hardness versus Brinell hardness for treatment 7 (plotting symbols represent the four values of the cause).

Chapter 17 Supplement

Feedforward Control Based on a Dominant Cause

S17.1 SIMULATING THE BENEFIT OF SELECTIVE FITTING

When considering implementation of selective fitting in the steering vibration example, the team needed to choose the number of bins and the location of the bin boundaries. They simulated the effect of two and three bins to determine their choice. The team chose the bin boundaries so that roughly equal numbers of components would fall into each bin. For example, with two bins they used the median center of gravity for part 2 (that is, 0.75) to define the boundary between the bins. We recommend choosing bin boundaries that result in roughly equal frequencies of components. For more advanced considerations, see Mease (2004).

We simulate the proposed selective fitting scheme using weights for the components sampled in some earlier investigation from the existing process. We start by specifying the number of bins, the bin boundaries, and the protocol for the assembly operation. In the example, we used a MINITAB macro that randomly selected a number of component pairs and calculated the distance from the center of gravity to the axis of rotation. See Appendix A for more information on writing MINITAB macros.

In the following, we give a MINITAB macro that assesses the effect of using two bins for part 2, called large and small, with bin boundary 0.75. The assembly protocol is to measure part 1. If the measured value exceeds the median for part 1, then select a part 2 from the large bin. In the simulation, the part 1 distances are sampled randomly from the baseline distribution.

```
MACRO
selectivefit2 part1 part2

mconstant I I2 cpart2 part1now part2now pnow test distc
mcolumn part2 part1 p1list allvals temp testcol

let cpart2=1
let  I2=1
let temp=1
let allvals=0

#repeat 50 times to get a reasonable estimate of the standard deviation
while I2<=50
 Sample 100 part1 p1list #get a list of 100 distances for part1 at random
  let I=1
  while I<=100
    let part1now=p1list(I)   #look at the next part1
    if part1now>0.75   #try to find a part2 that is also bigger than 0.75
       let test=0
     while test<=0
        let part2now=part2(cpart2)   #look at the next part2
        let cpart2=cpart2+1
        if cpart2>100  #end of list, start again
           let cpart2=1
        endif
        if part2now>0.75
          let test=1  #found appropriate match
        endif
        endwhile
     let distc=abso(part1now-part2now)  #vector
     endif
if part1now<=0.75   #try to find a part2 that is also smaller than 0.75
  let test=0
  while test<=0
    let part2now=part2(cpart2)   #look at the next part2
    let cpart2=cpart2+1
    if cpart2>100   #end of list, start again
      let cpart2=1
      endif
    if part2now<=0.75
      let test=1   #found appropriate match
      endif
    endwhile
  let distc=abso(part1now-part2now)   #vector components
```

```
  endif
  if (I2=1) AND (I=1)
    let allvals=distc
  else
    Stack allvals distc allvals.   #store all the distance values
  endif
  let I=I+1
  endwhile
let I2=I2+1
endwhile

let test = MEAN(allvals)    #some summaries of the combine distances
print test
Let test = STDEV(allvals)
print test

Code (-100:2) 0 (2:100) 1 allvals testcol
let test=sum(testcol)/(50*100)  #determine proportion bigger than 2
print test
ENDMACRO
```

To call the macro, open the MINITAB file *steering wheel vibration feedforward* and copy the following command into the command line, replacing “filelocation” with the location of the macro on your system:

```
% filelocation\selectivefit2.txt' 'part1' 'part2'
```

Another way to simulate the effect of selective fitting is to first build models that describe the centers of gravity for the two components. Then, in the simulation, rather than drawing samples from the existing data, we draw samples from the models. We need to be careful that the models are appropriate. For an example of the modeling option, see the Chapter 17 exercises and solutions.

S17.2 MORE ON MAKING PREDICTIONS

There are many ways to create models to predict the output characteristic from the values of the dominant cause(s). Model building, which includes choosing the appropriate inputs and their form in the model, can be complicated. Residual plots can be helpful. Complex prediction models with many inputs should be validated to check for overfitting and other modeling errors. Validation involves building the prediction equation using one set of data and checking how well it works using a separate set (Neter et al., 1996).

Multiple Regression Models

In some circumstances, the prediction of output values can be improved by including other causes in addition to a dominant cause in the regression model. In other words, we may try to fit a model of the form:

$$output = b_0 + b_0 cause_1 + ... + b_k cause_k + residual,$$

where the residual is the variation due to all other causes not explicitly included in the model. This is called a *multiple regression model.* For further discussion of regression models, see the Chapter 12 supplement and Appendix E.

We want to avoid adding nondominant causes to the model because:

- Measuring additional inputs can be expensive.
- The inclusion of these inputs in the model can lead to overadjustment and, thus, to increased output variation.

Remember there are measurement errors, prediction errors, and adjustment errors. It does not make sense to improve the prediction marginally.

Smoothers

A good smoother is the LOcally WEighted Scatterplot Smoother (LOWESS) available in MINITAB (Cleveland, 1979). A LOWESS smoother models the cause/output relationship with a smooth curve. Predictions for the output can then be determined for any input value within the usual range. This prediction does not assume a linear relationship between the cause and the output.

Time Series Models

Time series prediction models (Box et al., 1994; Abraham and Ledholter, 1983) use the current and previous values of the dominant cause to help improve the prediction of the next output value. A time series model of this sort is called a *transfer function.* Transfer function models can be useful, for instance, if there is some time-to-time variation in the dominant cause or if the cause has a delayed reaction on the output characteristic. Box et al. (1994) call a feedforward controller dynamic if the prediction of the output is based on a transfer function.

Chapter 18 Supplement

Feedback Control

S18.1 ALTERNATIVE SIMPLE FEEDBACK CONTROLLERS

There are many informal ways to implement a feedback control.

Grubbs's Rule for Feedback in a Setup-Dominated Process

Grubbs proposed a method for the adjustment of a machine to center the process output at startup. See Grubbs (1954) and Del Castillo (2002) for details. The adjustment scheme is:

- Measure the first part after setup and make an adjustment equal to the difference between observed measurement and the process target.
- Measure the second part and make half the indicated adjustment.
- Measure the third part and make a third of the indicated adjustment.
- Stop after a set number of parts.

This scheme is designed for processes where the dominant cause of variation is in the setup. That is, once the process has been centered after startup, it does not drift or suffer sudden shifts.

Precontrol and Control Charts

Precontrol (or *stoplight control*) is a feedback control scheme where the possible output values are divided into regions based on the specification limits as shown in Figure S18.1. The green zone is the middle 50% of the specification range.

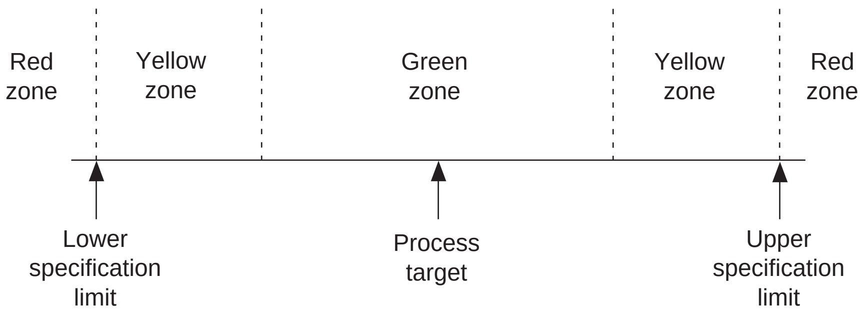


Figure S18.1 Precontrol division of the specification range.

To implement one of many versions of the rules, measure the output on one part at fixed time periods, and:

- If the measured output falls in the green zone, make no adjustment.
- If the measured output falls in the red zone, make a full adjustment and use a check (five consecutive parts in the green zone, for example) to ensure the process is properly centered.
- If the measured output falls in the yellow zone, measure another part. If the second measured value falls in the yellow (on the same side of the target) or red zones, make an adjustment. Otherwise, continue with no adjustment.

Satterthwaite (1954) first introduced Precontrol, which has received considerable attention in the research literature. For example, see Traver (1985), Shainin and Shainin (1989), Mackertich (1990), and Ledolter and Swersey (1997b). For Precontrol to be successful, the drift in the process must be relatively slow compared to the frequency of sampling and small relative to the specification range. Note that we need to add a rule to determine the size of the adjustment.

As another alternative, we can use a control chart to monitor the process and signal the need for adjustment. In the simplest version, we periodically measure one or more output values and plot the measured values on $\bar{X}$ and R charts. We make an adjustment if there is an out-of-control point on either chart. Again, we need a rule to decide on the size of the adjustment. To be successful, the sampling must be frequent enough to quickly detect changes in the process center.

Unlike Precontrol, we can also use control charts for process monitoring and for detecting the action of a cause of variation. A good comparison of process monitoring and feedback control is given in Box and Kramer (1992).

S18.2 SIMULATING THE BENEFIT OF FEEDBACK CONTROL

To quantify the potential benefit of a proposed feedback control scheme before implementation, we can simulate the effect by applying the scheme retrospectively to some historical (baseline or other) data from the existing process. The sampling plans in the proposed scheme and the historical data must match.

In the simulation, we keep track of the cumulative adjustment made to the series. Once the feedback control scheme is in operation, we observe only the adjusted series. To fix the notation we define

- y_t as the unadjusted output at time t
- y_t^* as the adjusted (observed) output at time t when using the feedback controller
- a_t as the adjustment made at time t (note the adjustment is not felt in the process until time $t + 1$)
- c_t as the cumulative adjustment up to and including time t

Then, we have:

- $c_t = \sum_{i=1}^{t} a_i = c_{t-1} + a_t$, that is, the cumulative adjustment is a sum of all previous adjustments
- $y_t^* = y_t + c_{t-1}$, that is, the observed output is the original (unadjusted) output plus the cumulative adjustment

As long as we keep track of the cumulative adjustments in the simulation, we can use these equations to go back and forth between the original and adjusted series.

To simulate the effect of applying the feedback controller, we start at time 1 and calculate adjustment a_1 and the cumulative adjustment $c_1 = a_1$. Then, at time 2, we observe the output $y_2^* = y_2 + c_1$. Next, we apply the feedback control rules to the observed series y_1^*, y_2^* to get the adjustment at time 2, a_2, and the cumulative adjustment $c_2 = a_1 + a_2 = c_1 + a_2$. And so on.

The Matlab (see http://www.mathworks.com) code that follows illustrates the retrospective application of a feedback control scheme to the flow rate example from Chapter 18. The simulation is also straightforward to implement in Microsoft Excel. We show both the simulation of a controller based on the EWMA forecast with parameter alpha (as suggested in Chapter 18) and a simpler feedback controller that adjusts back to target if the observed flow rate is outside the range [99, 107].

```
function []=feedbacksim(flowrate,alpha,dev)
%simulate the effect of using feedback on the fascia film build example
%original flow rate data given in vector "flowrate"
%example function call:  feedbacksim(flowrate,0.2,4)

%try out the feedback controller based on the EWMA forecast of unadjusted %series,
  and based on partial adjustment of deviation from target 103

cumadj1=0;  %cumulative adjustment for EWMA
cumadj2=0;  %cumulative adjustment for simple controller
z=flowrate-103*ones(1,180); %deviation from target for UNADJUSTED series

%do not make an changes for time 1
aseries1=z(1);   aseries2=z(1);   %initialize
```

```
for t=1:179,  %look at all flow rate values
     %try EWMA controller
     cumadj1=cumadj1-alpha*aseries1(t);
   aseries1(t+1)=z(t+1)+cumadj1;
   %try simple controller
     if (aseries2(t)<-dev) | (aseries2(t)>dev), %on standardized scale
     cumadj2=cumadj2-aseries2(t);
   end;
   aseries2(t+1)=z(t+1)+cumadj2;
end;

std(aseries1)   %stdev of series using EWMA based controller
std(aseries2)   %stdev of series using simple feedback controller

%plot of adjusted and unadjusted series
plot([1:180],z,'o-');  hold on
plot([1:180],aseries1,'x-');
plot([1:180],aseries2,'x--');
hold off
```

Note that when using an exponential smoother to predict the next output value (as with the flow rate data shown in Figure 18.6) and making the full adjustment, MINITAB automatically gives the standard deviation of the adjusted series as the square root of the MSD, so there is no need to simulate.

Simulating the implementation of the proposed feedback controller using historical data will most likely overestimate the potential benefit since:

- We use the same historical data to model the data and develop a prediction equation.
- The simulation assumes there is no adjustment error.

An alternative way to assess the potential benefit of feedback is to fit a time series model to the historical data and simulate new output data. With a model for the output we need to check that the simulated series without any adjustment seems reasonable when compared to any historical data we have from the existing process. An advantage of having a model is that we can repeatedly simulate the effect of the proposed feedback controller.

S18.3 PROPERTIES OF THE EXPONENTIAL WEIGHTED MOVING AVERAGE

There are many options to obtain a forecast of future output values. Suppose we have a sampling protocol that looks at parts at equally spaced times labeled 1, 2, ..., t, and at each time point, we measure the output. The goal is to predict the output at time $t + 1$ using the observed values up to time t. Mathematically, we may view this as predicting the next output y_{t+1}, using all previous outputs y_t, y_{t-1}, y_{t-2}, ..., where y_t denotes the value of the output measured at time t. For the moment, we assume there have been no adjustments. Denoting the one-step-ahead forecast as $\hat{y}_{t+1}$, two possible predictors are:

- Moving average: the average of the last k outputs; that is,

$$\hat{y}_{t+1} = \left(y_t + y_{t-1} + \ldots + y_{t-k+1}\right)/k$$

- EWMA: a weighted average of all previous outputs with exponentially decaying weights; that is,

$$\hat{y}_{t+1} = \alpha\left[y_t + (1-\alpha)y_{t-1} + (1-\alpha)^2 y_{t-2} + (1-\alpha)^3 y_{t-3} + \ldots\right] \quad \text{(S18.1)}$$

where α (the Greek letter *alpha*) is a constant and $0 < \alpha \leq 1$.

A special case of both options is using the last output value; that is, setting $\hat{y}_{t+1} = y_t$. Figure S18.2 illustrates the difference between the two forecasting options in terms of the relative weights given to past observations.

In the EWMA, a large value of α is best when the original series has small short-term variation, and thus recent observations provide a good prediction of the output value of the next unit. In situations where the process mean changes gradually, a good value for α often lies between 0.2 and 0.4 (Box and Luceno, 1997).

In any application, the best choice of forecast depends on the how the process changes over time when no adjustments occur. If there are sudden shifts, we want to use the moving average with k small. If the process shifts gradually, the EWMA is preferred. We need to determine k if we select the moving average and α if we pick the EWMA. The choice is best made empirically by simulating the performance of the possible schemes using some historical data. See Section S18.2.

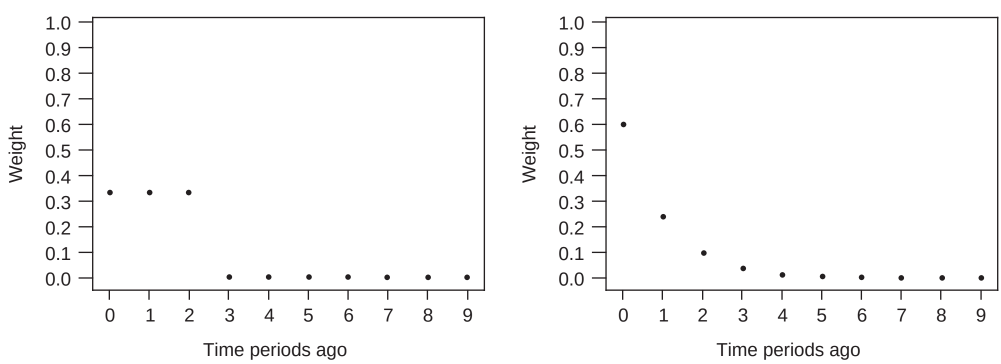


Figure S18.2 Possible weights (left panel moving average with $k = 3$, right panel EWMA with $\alpha = 0.6$).

Adjusting a Controlled Process

Once a controller is in operation, we no longer observe the original unadjusted series, and it is not obvious how to predict the next observation or what the appropriate adjustment is at each time point. In this section, we show that if we:

- Use the EWMA forecast and
- Assume perfect and complete adjustment after each observation,

then the correct adjustment is $\alpha(y_t^* - T)$ toward the target, where y_t^* is the observed output (that is adjusted series) at time *t,* and *T* is the target value for the process center. We use the same notation as in the previous section to define the adjustment at time *t*, a_t, and the cumulative adjustment at time *t*, c_t.

Without loss of generality, we assume y_t and y_t^* have been scaled so that the target value is zero. Now suppose at each time period we make an adjustment to partially compensate for the difference between the observed output y_t^* and the process target; that is, we set

$$a_t = -\alpha\left(y_t^* - 0\right)$$

Then, the cumulative adjustment at time t is given by

$$\begin{aligned}
c_t &= c_{t-1} + a_t \\
&= c_{t-1} - \alpha y_t^* \\
&= c_{t-1} - \alpha\left(y_t + c_{t-1}\right) \\
&= -\alpha y_t + (1-\alpha)c_{t-1} \\
&= -\alpha y_t + (1-\alpha)\left(c_{t-2} + a_{t-1}\right) \\
&= -\alpha y_t + (1-\alpha)\left(c_{t-2} - \alpha\left(y_{t-1} + c_{t-2}\right)\right) \\
&= -\alpha\left(y_t + (1-\alpha)y_{t-1}\right) + (1-\alpha)^2 c_{t-2} \\
&= -\alpha\left(y_t + (1-\alpha)y_{t-1}\right) + (1-\alpha)^2\left(c_{t-3} - \alpha\left(y_{t-2} + c_{t-3}\right)\right) \\
&= -\alpha\left(y_t + (1-\alpha)y_{t-1} + (1-\alpha)^2 y_{t-2}\right) + (1-\alpha)^3 c_{t-3} \\
&\ \ldots
\end{aligned}$$

so

$$c_t = -\hat{y}_{t+1}$$

Hence the cumulative adjustment is given by Equation (18.1); that is, we fully compensate for the deviation between the EWMA forecast for y_{t+1} made at time *t* and the target. Since, for the adjusted series, the predicted value at time $t+1$ equals $\hat{y}_{t+1}^* = \hat{y}_{t+1} + c_t$, the predicted output at time *t* + 1 will be on target.

Partial Adjustment

In discussing a feedback controller based on an EWMA forecast, we assume a full adjustment is made after each observation. Due to cost or other considerations, we may decide to make an adjustment only if the predicted deviation from target is large.

We can easily assess the effect of occasional adjustments by keeping track of the cumulative adjustment made up to time t (that is, c_{t-1}) so that we can recreate the series we would have observed had no adjustments been made ($y_t = y_t^* - c_{t-1}$). Using this unadjusted (and unobserved) series, we can determine an EWMA or any other forecast for the next time period. Then, translating back to the observed series, we compare the predicted observed output (that is, $\hat{y}_{t+1} + c_{t-1}$) to the process target to decide if further adjustment is needed.

S18.4 CONNECTION TO PID CONTROLLERS

There is extensive literature on feedback control in the engineering field. See for instance Aström (1970), Aström and Wittenmark (1989), and Prett and Gracia (1988). This literature does not, for the most part, address the important question of how to estimate model parameters from data. The EWMA-based controller is a discrete analog of a proportional, integral, derivative (PID) controller. A comprehensive treatment of feedback control from a statistical perspective is given in Box and Luceno (1997).

Chapter 19 Supplement

Making a Process Robust

S19.1 TAGUCHI'S METHODOLOGY

There is considerable controversy about the methods, experimental designs, and analysis suggested by Genichi Taguchi (1987); for instance, see Nair (1992) and Phadke (1989). Generally, at least in the statistical literature, it is felt that the fundamental ideas are valuable but that the specific designs and analyses can be improved upon. We briefly look at a number of ideas promoted by Taguchi.

In contrast to Statistical Engineering, with Taguchi's methodology:

- There is no recommended algorithm to reduce variation in existing processes
- There is little emphasis on understanding the patterns and causes of variation in the process output
- There is no explicit recognition of or search for a dominant cause
- There is heavy use of the robustness approach

To summarize the lack of emphasis put on identifying the cause, we quote Shin Taguchi from Nair (1992):

> *Notice that the objective of parameter design is very different from a pure scientific study. The goal in parameter design is not to characterize the system but to achieve robust function. Pure science strives to discover the causal relationships and to understand the mechanics of how things happen. Engineering, however, strives to achieve the results needed to satisfy the customer. Moreover, cost and time are very important issues for engineers. Science is to explain nature while engineering is to utilize nature.* (p. 130)

Taguchi considers using experimental design in both the product development process and to improve an existing process. This is a much broader view of improvement than is taken in Statistical Engineering, where we focus on improving an existing process.

To improve an existing process, Taguchi recommends:

- The process desensitization approach if there is a suspect cause
- The process robustness approach otherwise

In our experience, we have seen the desensitization approach fail because the cause included in the experiment was not dominant. There is little value in making the process insensitive to a cause that makes a small contribution to the overall variation.

Taguchi calls performance measures *signal-to-noise (S/N) ratios.* There is a large variety of signal-to-noise ratios that can be applied in different situations. Most often, Taguchi recommends S/N ratios of the form

$$10\log\left(\bar{y}^2/s^2\right) \text{ or } -10\log\left(s^2\right)$$

where $\bar{y}$ is the average output in a run and s is the standard deviation of the output values measured within a run. Note that $-10\ \log(s^2) = -20\ \log(s)$, so that Taguchi S/N ratio $-10\ \log(s^2)$ is a rescaled version of the performance measure we recommend to analyze robustness experiments where the goal is to reduce the within-run variation. We get identical conclusions from the analysis with either performance measure.

In some problems, the goal is to lower or increase the process center. Taguchi calls these *smaller/larger-is-better* problems and recommends the S/N ratios:

$$\text{Smaller is better: } -10\ \log[(y_1^2 + ... + y_r^2)\ /\ r]$$

$$\text{Larger is better: } -10\log[\frac{1}{r}(\frac{1}{y_1^2} + ... + \frac{1}{y_r^2})]$$

where $y_1, ..., y_r$ are the r measured values of the output characteristic on the repeats within each run. The idea is to calculate the performance measure (S/N ratio) for each run and then analyze these measures as the response in the experiment. In all cases, S/N ratios are defined so that larger values are better in terms of the goal of the problem. See Box (1988) for further discussion of signal-to-noise ratios. We recommend using two performance measures, the within-run average and log standard deviation, rather than a single S/N ratio.

Many of Taguchi's designs and methods of analysis are available in MINITAB. We do not recommend them.

APPENDICES

Using MINITAB

In the appendices we show how to create the graphs and statistical analyses described in this book using the statistical software package MINITAB. Throughout we assume the reader is familiar with a Windows-based environment.

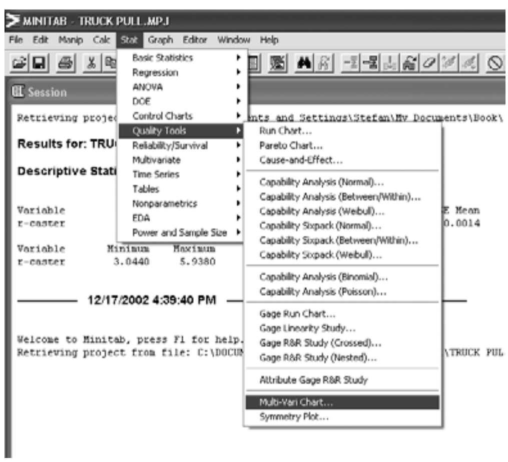
MINITAB is a general-purpose statistical software package. See http://www.minitab.com. MINITAB is a leader in the area of quality improvement and is widely used in industry. There are other suitable packages such as SAS, Splus, Statgraphics, JMP, and Systat. We do not recommend Microsoft Excel because we find it cumbersome and inefficient for producing the required analyses. In this book we illustrate MINITAB release 13.30 for Windows.

MINITAB is easy to use since its interface is based on pull-down menus and dialog boxes. The capabilities of MINITAB are well explained in the manuals (MINITAB *User's Guide 1* and *2,* 2000a and 2000b) and by the online help. In particular, in each dialog box there is an option for context-sensitive help that usually includes an example. A good reference book for MINITAB is Ryan et al. (2000).

In these appendices, we show how to select a MINITAB command through its pull-down menu structure using words and arrows. For example, a multivari chart is obtained in MINITAB using the menu selection:

Stat Æ *Quality Tools* Æ *Multi-Vari Chart*

This corresponds to menu navigation.



Finding a MINITAB command.

Appendix A

Data Storage and Manipulation

The first step is to get the data into the MINITAB worksheet. To illustrate, we use an investigation where the goal was to compare the performance of three different measurement devices—a feeler gage, a height gage, and a scale gage—for measuring the distance between a bottle lip and a label. The data are given in the file *label height measurement*.

To open a MINITAB worksheet (a file with the extension *.mtw)* we use:

File Æ *Open Worksheet*

To open an existing MINITAB project, which includes a worksheet with the data and all numerical and graphical summaries of the data we previously produced, use

File Æ *Open Project*

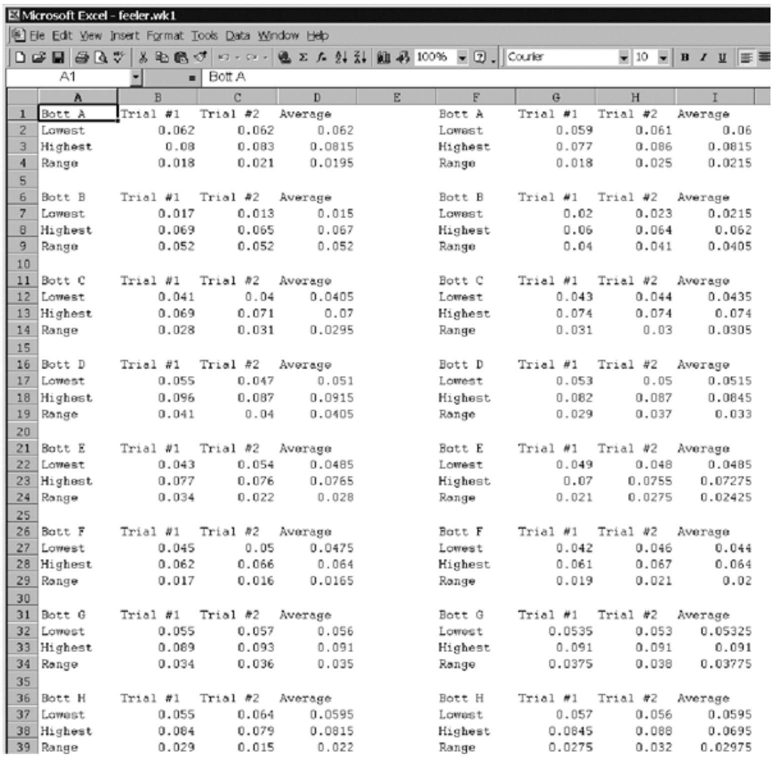
A.1 ROW/COLUMN DATA STORAGE FORMAT

In MINITAB, data are stored in a worksheet similar in appearance to a Microsoft Excel worksheet. See Figure A.1 for an example. There are, however, some important differences between Excel and MINITAB worksheets. With MINITAB, the data must be entered in the row/column format, where each column in the worksheet represents a different characteristic

Worksheet 1 ***

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9
	part	operator	feeler_low	feeler_high	height_low	height_high	scale_low	scale_high	
1	1	1	0.0620	0.0800	6.579	6.587	0.063000	0.094000	
2	1	1	0.0620	0.0830	6.580	6.588	0.094000	0.125000	
3	1	2	0.0590	0.0770	6.566	6.581	0.078130	0.093750	
4	1	2	0.0610	0.0860	6.574	6.598	0.062500	0.093750	
5	2	1	0.0170	0.0690	6.587	6.635	0.031000	0.094000	
6	2	1	0.0130	0.0650	6.586	6.626	0.031000	0.094000	
7	2	2	0.0200	0.0600	6.593	6.636	0.031250	0.062500	
8	2	2	0.0230	0.0640	6.588	6.631	0.031250	0.062500	
9	3	1	0.0410	0.0690	6.581	6.600	0.063000	0.094000	
10	3	1	0.0400	0.0710	6.577	6.605	0.063000	0.094000	
11	3	2	0.0430	0.0740	6.581	6.593	0.046875	0.078125	
12	3	2	0.0440	0.0740	6.582	6.612	0.031250	0.078125	
13	4	1	0.0550	0.0960	6.550	6.585	0.063000	0.094000	
14	4	1	0.0470	0.0870	6.575	6.618	0.063000	0.094000	
15	4	2	0.0530	0.0820	6.581	6.620	0.046875	0.078125	
16	4	2	0.0500	0.0870	6.576	6.624	0.046875	0.093750	

Figure A.1 Data in the MINITAB worksheet.



Microsoft Excel - feeler.wk1

	A	B	C	D	E	F	G	H	I
1	Bott A	Trial #1	Trial #2	Average		Bott A	Trial #1	Trial #2	Average
2	Lowest	0.062	0.062	0.062		Lowest	0.059	0.061	0.06
3	Highest	0.08	0.083	0.0815		Highest	0.077	0.086	0.0815
4	Range	0.018	0.021	0.0195		Range	0.018	0.025	0.0215
5									
6	Bott B	Trial #1	Trial #2	Average		Bott B	Trial #1	Trial #2	Average
7	Lowest	0.017	0.013	0.015		Lowest	0.02	0.023	0.0215
8	Highest	0.069	0.065	0.067		Highest	0.06	0.064	0.062
9	Range	0.052	0.052	0.052		Range	0.04	0.041	0.0405
10									
11	Bott C	Trial #1	Trial #2	Average		Bott C	Trial #1	Trial #2	Average
12	Lowest	0.041	0.04	0.0405		Lowest	0.043	0.044	0.0435
13	Highest	0.069	0.071	0.07		Highest	0.074	0.074	0.074
14	Range	0.028	0.031	0.0295		Range	0.031	0.03	0.0305
15									
16	Bott D	Trial #1	Trial #2	Average		Bott D	Trial #1	Trial #2	Average
17	Lowest	0.055	0.047	0.051		Lowest	0.053	0.05	0.0515
18	Highest	0.096	0.087	0.0915		Highest	0.082	0.087	0.0845
19	Range	0.041	0.04	0.0405		Range	0.029	0.037	0.033
20									
21	Bott E	Trial #1	Trial #2	Average		Bott E	Trial #1	Trial #2	Average
22	Lowest	0.043	0.054	0.0485		Lowest	0.049	0.048	0.0485
23	Highest	0.077	0.076	0.0765		Highest	0.07	0.0755	0.07275
24	Range	0.034	0.022	0.028		Range	0.021	0.0275	0.02425
25									
26	Bott F	Trial #1	Trial #2	Average		Bott F	Trial #1	Trial #2	Average
27	Lowest	0.045	0.05	0.0475		Lowest	0.042	0.046	0.044
28	Highest	0.062	0.066	0.064		Highest	0.061	0.067	0.064
29	Range	0.017	0.016	0.0165		Range	0.019	0.021	0.02
30									
31	Bott G	Trial #1	Trial #2	Average		Bott G	Trial #1	Trial #2	Average
32	Lowest	0.055	0.057	0.056		Lowest	0.0535	0.053	0.05325
33	Highest	0.089	0.093	0.091		Highest	0.091	0.091	0.091
34	Range	0.034	0.036	0.035		Range	0.0375	0.038	0.03775
35									
36	Bott H	Trial #1	Trial #2	Average		Bott H	Trial #1	Trial #2	Average
37	Lowest	0.055	0.064	0.0595		Lowest	0.057	0.056	0.0595
38	Highest	0.084	0.079	0.0815		Highest	0.0845	0.088	0.0695
39	Range	0.029	0.015	0.022		Range	0.0275	0.032	0.02975

Figure A.2 Poorly stored data in an Excel spreadsheet.

(output or input), and each row represents a different observation. This format is not necessarily intuitive for people who commonly use Excel. Also, in MINITAB every column (i.e. characteristic) can be given a descriptive name.

To illustrate why the row/column format is recommended, see Figure A.2, which shows how the data for the measurement investigation were initially recorded. The Excel worksheet in Figure A.2 gives the results only for the feeler gage. The results for the height and scale gages were stored in separate worksheets. In Figure A.2, the meaning of each data value depends not only on its location in the worksheet, but also on the location of various labels. For example, the value in cell B2 (0.062) gives the lowest value of trial 1 of the first operator for bottle *A* using the feeler gage. This way of storing data is not convenient when doing analysis, since determining the meaning of the various data values is difficult.

Figure A.1 shows how the same data for all three gages were stored in MINITAB. The label above each column provides the name of the stored characteristic, and each row represents a different observation, or in this case, a measurement. The 0.062 value discussed previously is now stored in the first row of column C3 (labeled *feeler_low*). We can identify the corresponding part (bottle 1) and operator *(A)* by examining the values of the other characteristics in the same row.

The more efficient row/column format approach to storing data identifies the relationships among the collected values. We can also store data in Excel in the row/column format. Translating from the original data format, shown in Figure A.2, to the row/column format, shown in Figure A.1, is a tedious process.

Another important advantage of a MINITAB worksheet is its ability to easily handle missing values. Any blank spaces in the data file become stars (symbol for *missing*) in MINITAB. All analyses in MINITAB properly deal with missing values. It is a poor idea to use special numerical codes like –99 to signify missing values. Using numerical codes for missing observations will adversely affect many data summaries.

A.2 MAKING PATTERNED DATA

It is often useful to create patterned data. The patterned data commands are shown in Figure A.3 and accessed via the MINITAB menu selection:

Calc Æ *Make Patterned Data*

In the label height measurement example, to create the column corresponding to part numbers, we fill in the dialog window as in Figure A.4. Each of 12 parts is measured four times (twice each by two different operators). Applying the command creates four 1s, four 2s, and so on in column C1. If preferred, text values (for example, *A*, *B*, *C*, and so on) can be used to represent the different bottles.

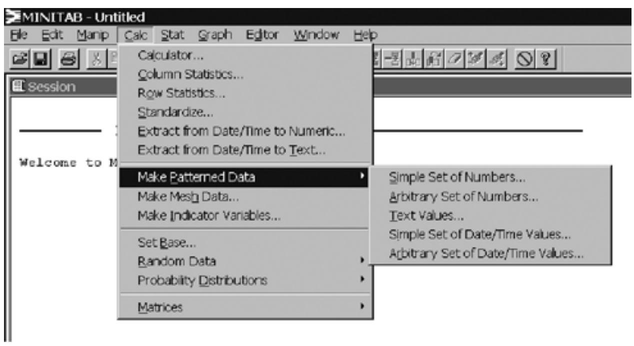


Figure A.3 Patterned Data menu.

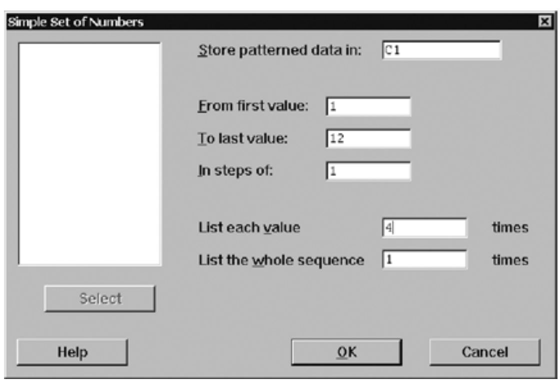


Figure A.4 Dialog box for simple set of patterned numbers.

A.3 CALCULATING DERIVED CHARACTERISTICS

To calculate new characteristics from the existing data, we use the calculator:

Calc Æ Calculator

In the label height measurement example, the minimum and maximum heights (as the bottle was rotated) were recorded. In Figure A.5, we use the calculator to record the difference (max–min) for each feeler gage measurement in column C9.

To use the calculator, we enter a formula involving the existing characteristics in the expression window. The expression can be typed in directly or identified using the *select* button. The expression may include any of the functions given in the list. Many standard statistical functions such as average and standard deviation are available. Note that (most of) these functions act simultaneously on all rows across different columns. Column summaries (for example, the mean or standard deviation of a characteristic) are discussed in Appendix B.

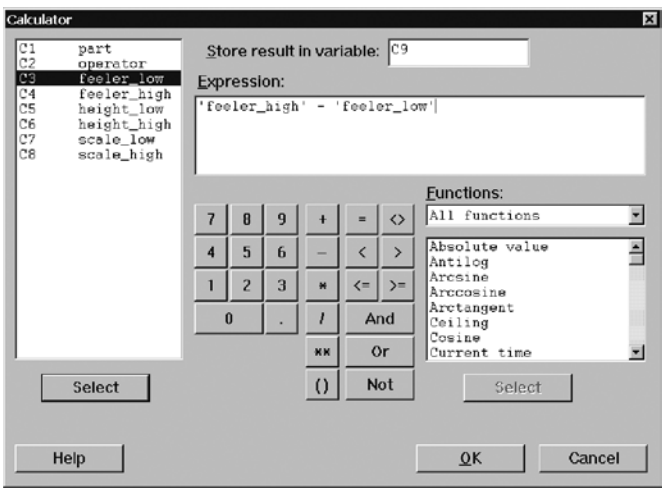


Figure A.5 Calculator dialog box.

A.4 SELECTING A DATA SUBSET

In many circumstances, it is useful to stratify the data in some way. Many numerical and graphical summaries (see appendices B and C) allow stratification without subsetting the data. We sometimes stratify data into separate worksheets using (see Figure A.6):

Manip Æ Subset Worksheet [Data Æ Subset Worksheet, in release 14.11]

or

Manip Æ Split Worksheet [Data Æ Split Worksheet, in release 14.11]

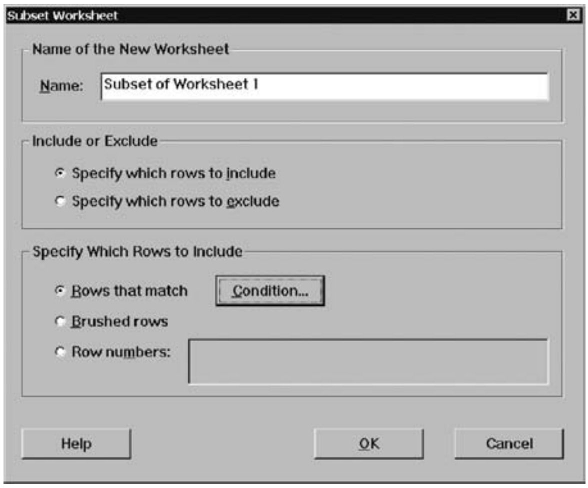


Figure A.6 Subset Worksheet dialog box.

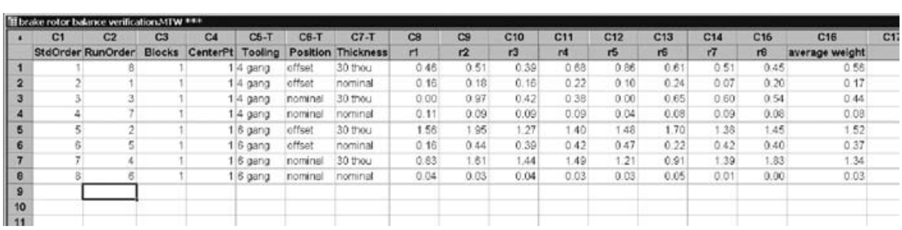
When selecting the subset, we select the data observations to either include or exclude. There are a variety of ways of specifying the subset of observations. We can give the row numbers, select the rows beforehand using brushing, or give a condition that must be satisfied.

A.5 STACKING COLUMNS

In some circumstances, it is convenient to have the data stored in more than one way in order to conduct the desired analysis. Consider the brake rotor balance example discussed in Chapter 13. In the verification experiment there were eight runs at each of the eight treatments and, within each run, eight rotors were produced. The data are stored in the file *brake rotor balance verification* and the worksheet is shown in Figure A.7. The balance weights required for the eight rotors within each treatment are stored in separate columns labeled *r1* through *r8*.

To start the analysis, we plot the weight by treatment as shown in Figure 13.4. To produce this plot, however, we need all 64 balance weights stored in a single column. We stack the columns using:

Manip Æ Stack Æ Stack Columns



brake rotor balance verification.MTW ***

↓	C1	C2	C3	C4	C5-T	C6-T	C7-T	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
	StdOrder	RunOrder	Blocks	CenterPt	Tooling	Position	Thickness	r1	r2	r3	r4	r5	r6	r7	r8	average weight	
1	1	8	1	1	4 gang	offset	30 thou	0.46	0.51	0.39	0.68	0.86	0.61	0.51	0.45	0.56	
2	2	1	1	1	4 gang	offset	nominal	0.16	0.18	0.16	0.22	0.10	0.24	0.07	0.20	0.17	
3	3	3	1	1	4 gang	nominal	30 thou	0.00	0.97	0.42	0.38	0.00	0.65	0.60	0.54	0.44	
4	4	7	1	1	4 gang	nominal	nominal	0.11	0.09	0.09	0.09	0.04	0.08	0.09	0.08	0.08	
5	5	2	1	1	6 gang	offset	30 thou	1.56	1.95	1.27	1.40	1.48	1.70	1.38	1.45	1.52	
6	6	5	1	1	6 gang	offset	nominal	0.16	0.44	0.39	0.42	0.47	0.22	0.42	0.40	0.37	
7	7	4	1	1	6 gang	nominal	30 thou	0.83	1.61	1.44	1.49	1.21	0.91	1.39	1.83	1.34	
8	8	6	1	1	6 gang	nominal	nominal	0.04	0.03	0.04	0.03	0.03	0.05	0.01	0.00	0.03	
9																	
10																	
11																	

Figure A.7 Brake rotor balance verification experiment plan and data.

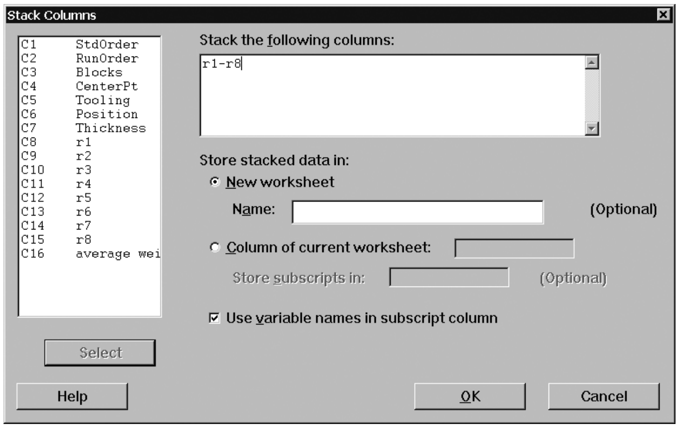
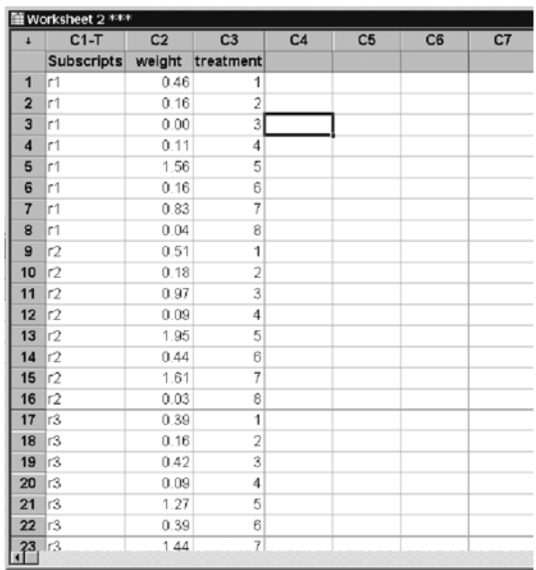


Figure A.8 MINITAB Stack Columns dialog box.



Worksheet 2 ***

↓	C1-T	C2	C3	C4	C5	C6	C7
	Subscripts	weight	treatment				
1	r1	0.46	1				
2	r1	0.16	2				
3	r1	0.00	3				
4	r1	0.11	4				
5	r1	1.56	5				
6	r1	0.16	6				
7	r1	0.83	7				
8	r1	0.04	8				
9	r2	0.51	1				
10	r2	0.18	2				
11	r2	0.97	3				
12	r2	0.09	4				
13	r2	1.95	5				
14	r2	0.44	6				
15	r2	1.61	7				
16	r2	0.03	8				
17	r3	0.39	1				
18	r3	0.16	2				
19	r3	0.42	3				
20	r3	0.09	4				
21	r3	1.27	5				
22	r3	0.39	6				
23	r3	1.44	7				

Figure A.9 Worksheet resulting from stacking columns.

The corresponding MINITAB dialog box with the required information filled in is given in Figure A.8.

Stacking the columns produces the first two columns shown in Figure A.9, where the second column has been given a descriptive label. To produce Figure 13.4, we need an additional column that gives the treatment corresponding to each weight. Figure A.9 shows the results of using the command to make patterned data as described in Section A.2.

The data were originally stored as in Figure A.7 because, for subsequent analysis, we calculate main and interaction effects based on the average balance weight.

A.6 MINITAB MACROS

Macros are useful for automating a series of commands repeated many times. Macros are also needed if you want do calculations row by row (that is, observation by observation) where the action depends on the previous value(s). For example, a macro can be used to simulate the effect of a proposed feedback control scheme on existing process data.

MINITAB macros are similar to the macros available in other software such as Excel. See the MINITAB help for more information on writing macros.

The sample macro illustrates testing a proposed feedback control scheme for the piston diameter example discussed in Chapter 18. The data are given in the file *V6 piston diameter 270*. The feedback scheme adjusts the process center back to the target of 6.7 if the observed diameter is less than 2.7 or greater than 10.7. Note that in the data file there are 200 diameter readings.

```
MACRO
macrotest diameter adjusted

mconstant cumad I diameter2 nadj
mcolumn diameter adjusted

let I=1
let cumad=0  #cumulative adjustment needed based on rule
let nadj=0

while I<=200
let diameter2=diameter(I)+cumad  #current value
let adjusted(I)=diameter2
#determine if further adjustment needed
if diameter2<2.7
let cumad=cumad+(6.7-diameter2)
let nadj=nadj+1
endif
if diameter2>10.7
let cumad=cumad-(diameter2-6.7)
let nadj=nadj+1
endif

let I=I+1
endwhile

ENDMACRO
```

We save the macro in a separate text file and call the macro using the command line editor. The command line editor is available using:

Edit Æ *Command Line Editor*

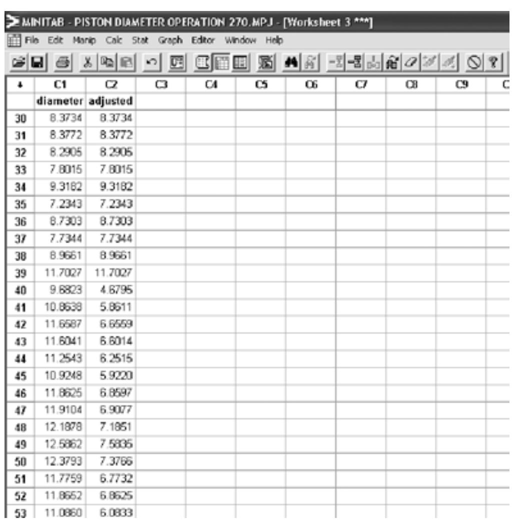
To run the macro from the command line we type:

> *%macro name (including path if necessary) v1 v2 ...*

where v1 and v2 are the characteristics from the current worksheet the macro acts on or characteristics the macros creates. For example, we may issue the command as shown in Figure A.10.

Figure A.10 Running a MINITAB macro.

The result of executing the macro is shown in the data file snapshot in Figure A.11. At observation 39, the observed diameter was 11.7. In the example, this was the first time an adjustment was required. At that time, the cumulative adjustment was set to –5. We can see the effect of the adjustment on subsequent diameters. Later in the data file (not shown in Figure A.11) further adjustments are necessary.



↓	C1 diameter	C2 adjusted	C3	C4	C5	C6	C7	C8	C9
30	8.3734	8.3734							
31	8.3772	8.3772							
32	8.2905	8.2905							
33	7.8015	7.8015							
34	9.3182	9.3182							
35	7.2343	7.2343							
36	8.7303	8.7303							
37	7.7344	7.7344							
38	8.9661	8.9661							
39	11.7027	11.7027							
40	9.6823	4.6795							
41	10.8638	5.8611							
42	11.6587	6.6559							
43	11.6041	6.6014							
44	11.2543	6.2515							
45	10.9248	5.9220							
46	11.8625	6.8597							
47	11.9104	6.9077							
48	12.1878	7.1851							
49	12.5862	7.5835							
50	12.3793	7.3766							
51	11.7759	6.7732							
52	11.8652	6.8625							
53	11.0860	6.0833							

Figure A.11 MINITAB worksheet showing effect of running the macro.

Figure A.12 MINITAB show history dialog button.

To write MINITAB macros, we can use the Show History dialog button that gives the command line interpretation for any MINITAB commands run using the pull-down menus. The History dialog button is the small yellow button with a prompt (>) at the far right at the top of the MINITAB window, as shown in Figure A.12.

Another example macro to simulate the effect of selective fitting is given as part of the solution to Exercise 17.2.

Appendix B
Numerical Summaries

We use a manifold sand scrap investigation to illustrate various numerical data summaries in MINITAB. In the investigation, each manifold was classified as scrap or not and linked (as well as possible) to a number of sand and pour characteristics. We have a total of 17 characteristics measured on 970 castings. The data are given in the file *manifold sand scrap comparison*.

B.1 NUMERICAL SUMMARIES FOR CONTINUOUS CHARACTERISTICS

Numerical summaries appropriate for continuous characteristics are available using:

Stats Æ *Basic Statistics* Æ *Display Descriptive Statistics*

Selecting the input characteristics pour time and temperature gives:

```
Descriptive Statistics: pourtime, temperature

Variable         N       Mean     Median     TrMean      StDev    SE Mean
pourtime       970     4.6990     4.7000     4.6794     0.5132     0.0165
temperature    970     92.446     92.450     92.442      2.751      0.088

Variable       Minimum    Maximum         Q1         Q3
pourtime        3.3000     7.5000     4.4000     5.0000
temperature     86.100     98.500     90.500     94.100
```

To define the summary measures, *Mean* is the sample average and *StDev* is the sample standard deviation. If we rank the values, the smallest value is given by *Minimum*, the largest value by *Maximum*, and the middle value by *Median*. *Q1* and *Q3* define the first and

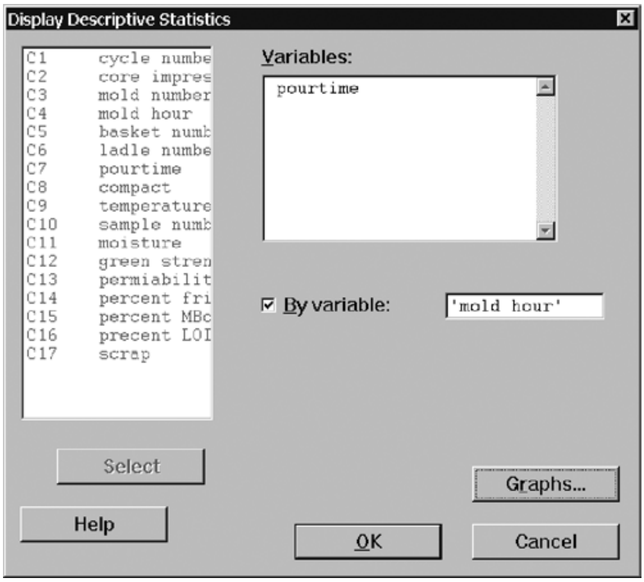


Figure B.1 Display Descriptive Statistics dialog box.

third quartiles—that is, the observation halfway in the ordered list between the minimum and the median, and halfway between the median and the maximum, respectively. The median defines the second quartile.

We can also produce numerical summaries for any characteristic stratified by a discrete characteristic. The dialog box in Figure B.1 shows how to stratify the summary for pour time by the mold hour.

The results are:

Descriptive Statistics: pourtime by mold hour

Variable	mold hour	N	Mean	Median	TrMean	StDev
pourtime	3	166	4.8470	4.7000	4.8320	0.6400
	4	172	4.7488	4.7000	4.7130	0.5893
	5	176	4.5409	4.5000	4.5196	0.5540
	6	180	4.6289	4.6000	4.6228	0.4833
	7	142	4.7423	4.7000	4.7359	0.3190
	8	134	4.7075	4.7000	4.7100	0.2568

Variable	mold hour	SE Mean	Minimum	Maximum	Q1	Q3
pourtime	3	0.0497	3.7000	6.2000	4.3000	5.3000
	4	0.0449	3.8000	7.5000	4.3000	5.1000
	5	0.0418	3.3000	7.3000	4.1250	4.8750
	6	0.0360	3.5000	6.5000	4.3000	4.9000
	7	0.0268	4.1000	5.5000	4.5000	5.0000
	8	0.0222	4.1000	5.2000	4.5000	4.9000

Sometimes we need to save numerical summaries in the worksheet for further analysis. We request storage of some of the descriptive summaries using:

Stats Æ *Basic Statistics* Æ *Store Descriptive Statistics*

From the dialog box (see Figure B.2) selecting *Statistics*, we check off the data summaries we want to save. This is useful in the analysis of multivari investigations that involve a part-to-part family or other family that is expected to have a haphazard effect.

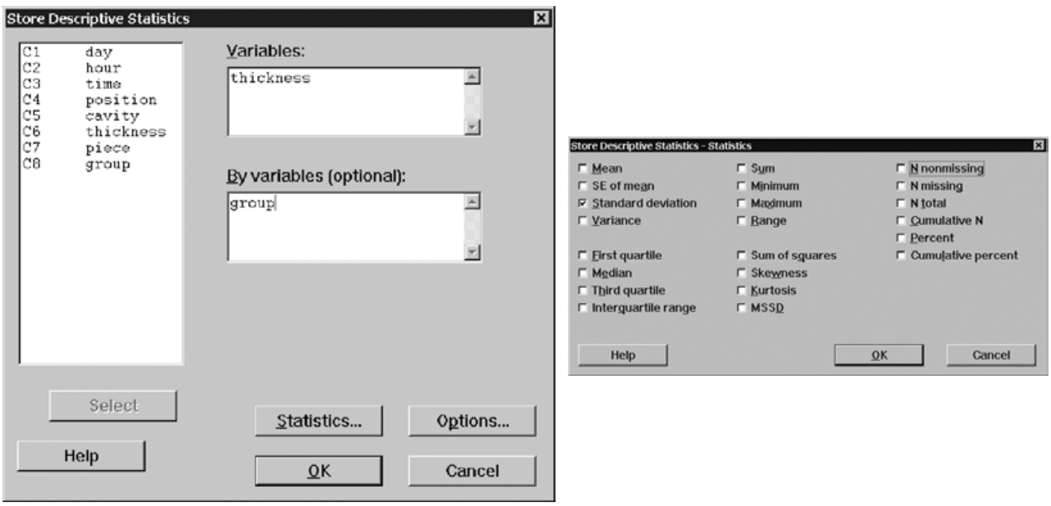


Figure B.2 Store Descriptive Statistics dialog box.

Note that numerical summaries are sometimes given using exponential notation. Exponential notation is convenient for very small or very large numbers. The integer given after the *E* tells us how many positions (to the right, if the integer is positive; to the left, if the integer is negative) to shift the decimal point; for example, 4.49E–03 = 0.00449.

B.2 NUMERICAL SUMMARIES FOR DISCRETE CHARACTERISTICS

For discrete characteristics, count summaries are available through the MINITAB menu selection:

Stat Æ *Tables* Æ *Tally*

Figure B.3 shows the *Tally* dialog box.

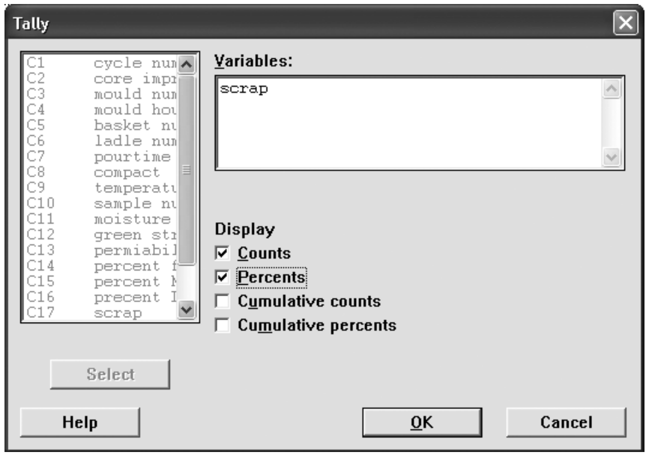


Figure B.3 Tally dialog box.

Using the tally summary for the characteristic scrap, we get

```
Tally for Discrete Variables: scrap

scrap  Count  Percent
    0    853    87.94
    1    117    12.06
   N=    970
```

We see that there are only two possible values for scrap (where 0 represents a passed manifold and 1 a manifold that was scrapped). The proportion of scrap is roughly 12%.

In such situations, we may also be interested in determining whether a relationship exists between two discrete characteristics. We use the MINITAB menu selection:

Stat Æ *Tables* Æ *Cross Tabulation*

For example, as illustrated by Figure B.4, we may be interested in determining if there is a relationship between mold hour and whether the casting is scrapped. The MINITAB results given as follows suggest the scrap rate is lower in hours 7 and 8 though the sample size in each hour is small.

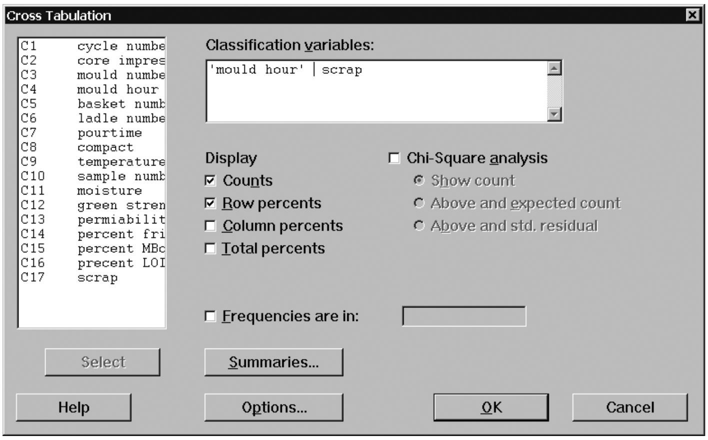


Figure B.4 Cross Tabulation dialog box.

```
Tabulated Statistics: mold hour, scrap

Rows: mold hour     Columns: scrap

           0        1      All

 3       141       25      166
       84.94    15.06   100.00

 4       148       24      172
       86.05    13.95   100.00

 5       151       25      176
       85.80    14.20   100.00

 6       156       24      180
       86.67    13.33   100.00

 7       133        9      142
       93.66     6.34   100.00

 8       124       10      134
       92.54     7.46   100.00

 All     853      117      970
       87.94    12.06   100.00

  Cell Contents --
                  Count
                  % of Row
```

Appendix C

Graphical Summaries

In all investigations, graphical summaries play an essential role in the analysis. Often we use only a graphical summary to draw conclusions.

C.1 HISTOGRAM

A histogram summarizes the distribution of a continuous characteristic in the data set. Histograms are available through the MINITAB menu selection:

Graph Æ *Histogram*

To make comparing histograms easier, we recommend using the *option* button and choosing *percent* rather than the default *frequency* display. In the manifold sand scrap example introduced in Appendix B, we enter the characteristic *pour time* as the graph variable to get the histogram in Figure C.1. The data are given in the file *manifold sand scrap comparison*.

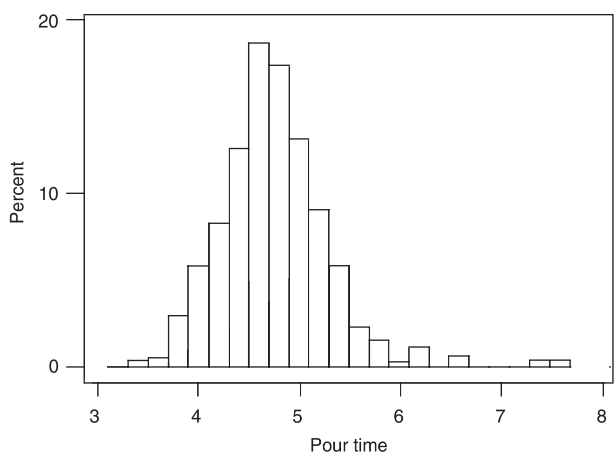


Figure C.1 Histogram of pour times.

The histogram shows the distribution of pour times. In Figure C.1, we see that the average pour time is roughly 4.75 seconds and that there are a few large pour times above 7 seconds. Note that the time order of the data is lost in a histogram.

If there is a large number of observations, it may be difficult to identify outliers (unusually small or large values) in a histogram. This occurs, for instance, in the histogram of alignment pull given in Figure 1.2. Box plots, covered in Section C.3, are a good alternative when the number of observations is large or when we wish to compare distributions.

C.2 RUN CHART

We use run charts (also called time series plots) to look for patterns over time. A run chart is available through:

Graph Æ Time Series Plot

The run chart of pour time in Figure C.2 shows the time order. Figure C.2 suggests the variation in pour times is greater at the beginning of the investigation. At the end, say after casting 700, the variation is noticeably smaller.

The run chart assumes the observations are equally spaced in time. If this is not the case, we use a scatter plot (see Section C.4) where the horizontal axis is defined in terms of time. If the data are collected in subgroups over time, say five units measured each hour, we can also use a multivari plot (see Section C.5).

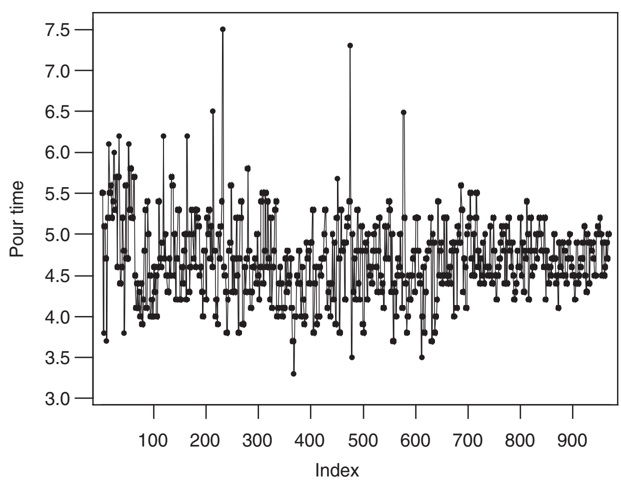


Figure C.2 Run chart of pour times.

C.3 BOX PLOT

A box plot is a simplified histogram useful for comparing data subsets and identifying outliers. Box plots are available using (see Figure C.3):

Graph Æ Boxplot

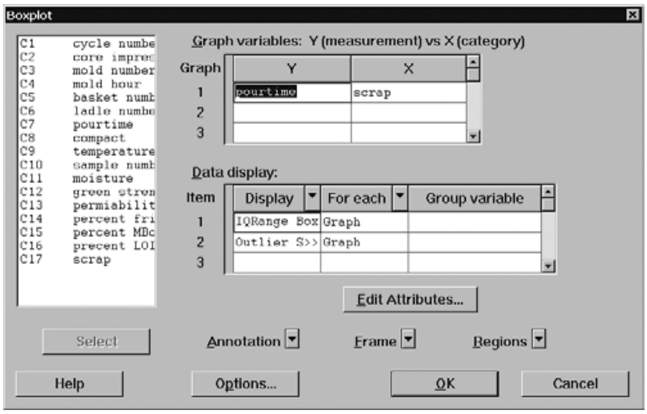


Figure C.3 Boxplot dialog box.

In Figure C.4 we stratify pour time by whether or not the part is scrapped. The horizontal line in the middle of each box gives the *median* value. The upper and lower edges of the rectangle show the first and third quartiles, *Q1* and *Q3*, defined in the descriptive statistics summary in Appendix B. The so-called whiskers are the lines coming out of the central rectangle. The ends of the whiskers identify the range of the data with the exception of unusual values plotted as separate stars. The plotting of unusual values with separate symbols is useful for finding outliers.

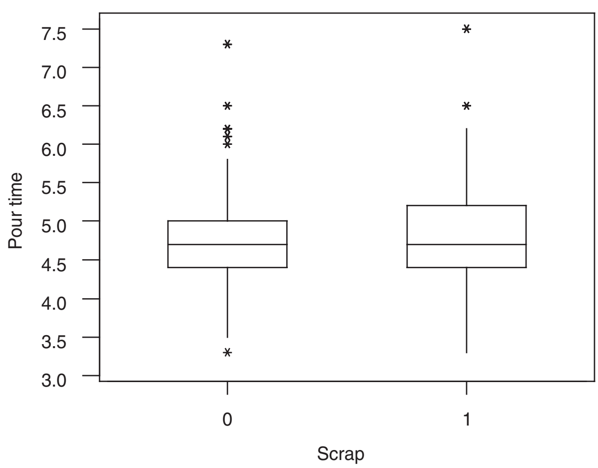


Figure C.4 Box plot of pour time by scrap.

If the number of observations summarized in the box plot is small, we prefer displaying the individual observations rather than the summary statistics. In MINITAB we change the display items to *Individual Symbol*, rather than the default of *IQRange Box* and *Outlier Symbol*. In MINITAB release 14.11, we can plot individual observations using:

Graph Æ Individual Value Plot

For example, Figure C.5 shows measurement error stratified by part number for the camshaft diameter relative bias investigation discussed in Chapter 7. The data come from the file *camshaft journal diameter measurement2*. In Figure C.5, we see the that measurement error is negative in all cases and that there is little difference in the distribution of the errors among the parts.

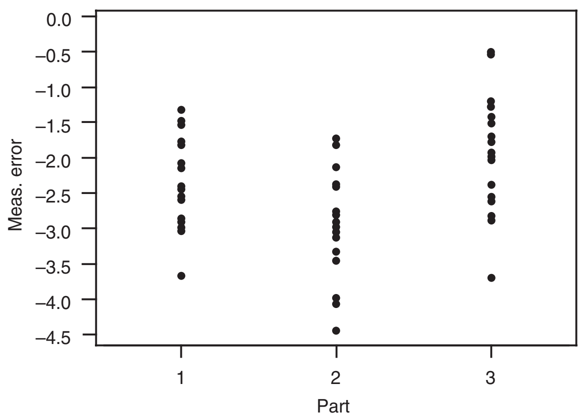


Figure C.5 Box plot (showing individual observations) of measurement error by part number.

C.4 SCATTER PLOT

A scatter plot is a generalization of a run chart where the horizontal axis is defined by any characteristic, not just time. Scatter plots are available through:

Graph Æ Plot [Graph Æ ScatterPlot in release 14.11*]*

Scatter plots are useful for examining the relationship between two characteristics. In the manifold sand scrap example we examine how pour time changes with ladle number (see Figure C.6).

The resulting scatter plot, given in Figure C.7, suggests there is no association between pour time and ladle number.

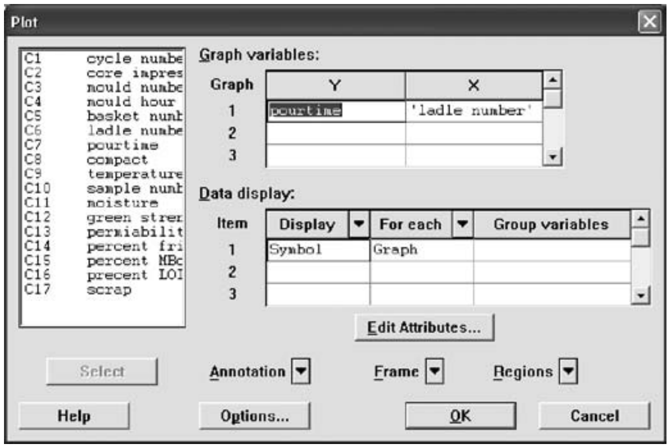


Figure C.6 (Scatter) Plot dialog window.

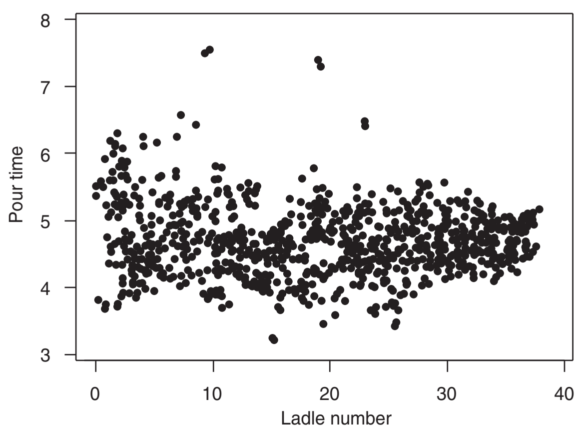


Figure C.7 Scatter plot of pour time versus ladle number.

Adding Jitter

In the crossbar dimension robustness investigation from Chapter 19, eight treatments were used, and each run consisted of five consecutive parts for each treatment. The data are given in the file *crossbar dimension robustness*. The output "burn" could take only four possible values. Plotting burn by treatment, as shown in the left panel of Figure C.8, is helpful but difficult to interpret, since we cannot see the output for all five parts in each run. Many observations are plotted on the same position, called overplotting. We add jitter (small

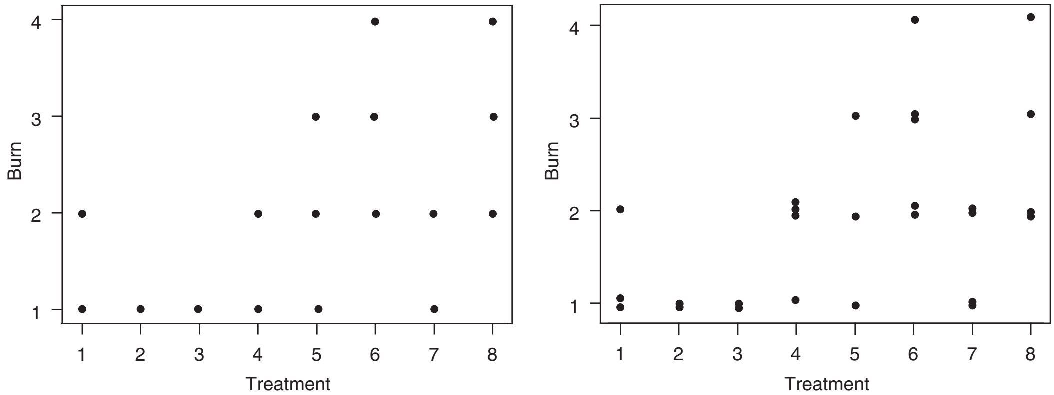


Figure C.8 Scatter plot of burn by treatment—regular on left, with jitter in vertical direction on right.

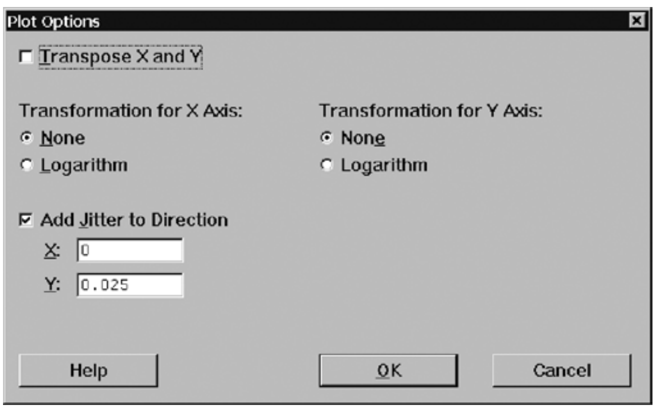


Figure C.9 Plot Options dialog box showing selecting of jitter.

random perturbations) to each observation to reduce the effect of overplotting as given in Figure C.9. Here we illustrate adding jitter in the vertical, or *Y*, direction. The scatter plot with added jitter is shown in the right panel of Figure C.8. The dialog box shown in Figure C.9 is obtained by selecting options in Figure C.6.

Labeling Points

The plotted points can be labeled using the values of any discrete characteristic. Consider the cylinder head scrap example discussed in Chapter 12. The data are given in the file *cylinder head scrap multivari*. The team wanted to see if the relationship between side shift (the output) and time depended on pattern.

They used the dialog box, as shown in Figure C.10, to make the scatter plot given in Figure C.11. Note the change to the data display in the Plot dialog box. There are four different plotting symbols, one for each pattern. Generally, pattern 1 has the largest side shift and pattern 2 the smallest, but there is no dependency on time.

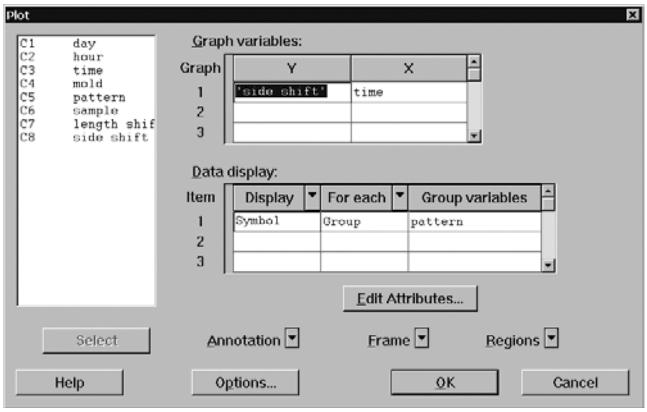


Figure C.10 (Scatter) Plot dialog box showing a group variable.

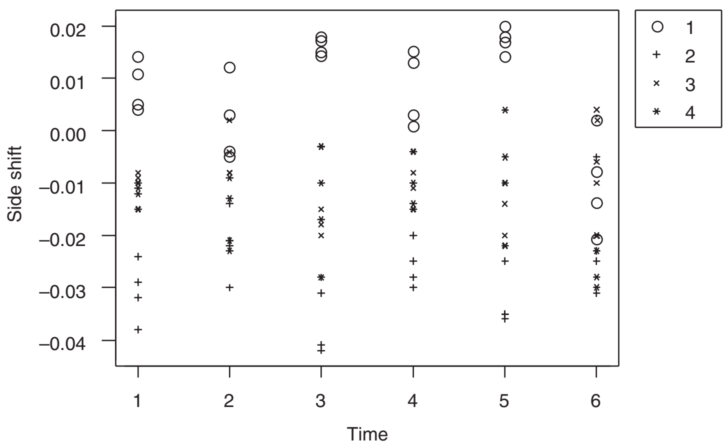


Figure C.11 Scatter plot of side shift by time with labels for different patterns.

Draftsman and Matrix Plots

To create many scatter plots simultaneously, MINITAB provides a way to produce multiple graphs with defined output and input characteristics. A draftsman plot is available through the MINITAB menu selection:

> *Graph* Æ *Plot* Æ *Draftsman Plot [*option *each Y versus each X* under *Graph* Æ *Matrix Plot* in release 14.11*]*

The draftsman plot automatically produces all scatter plots that involve the output characteristic (*Y* variable) and the list of input characteristics (X variables). Figure C.12 shows the dialog box.

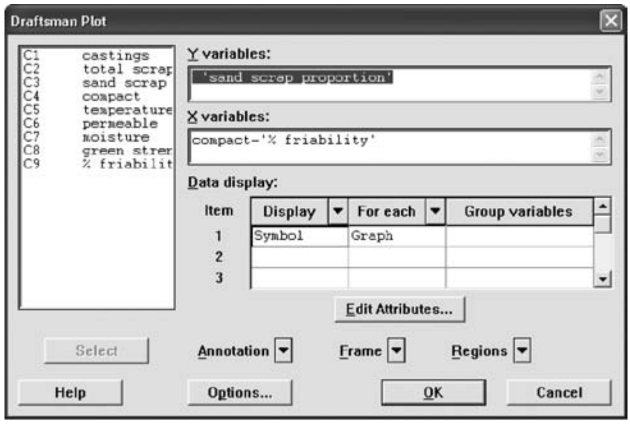


Figure C.12 Draftsman Plot dialog box.

In Figure C.13, we see no clear relationship between the sand scrap proportion and various sand characteristics. Upon closer inspection, however, a quadratic relationship between temperature and sand scrap proportion becomes clear. This example illustrates the point that if too many characteristics are chosen, draftsman plots (and matrix scatter plots, as will be discussed) will be hard to interpret, since each individual plot will be very small. In that case we look at the scatter plots one at a time.

A matrix plot is another way of producing multiple scatter plots simultaneously. Unlike with a draftsman plot, we now make no distinction between inputs and outputs. The window leaks problem discussed in Chapter 12 provides an example, where we also label the

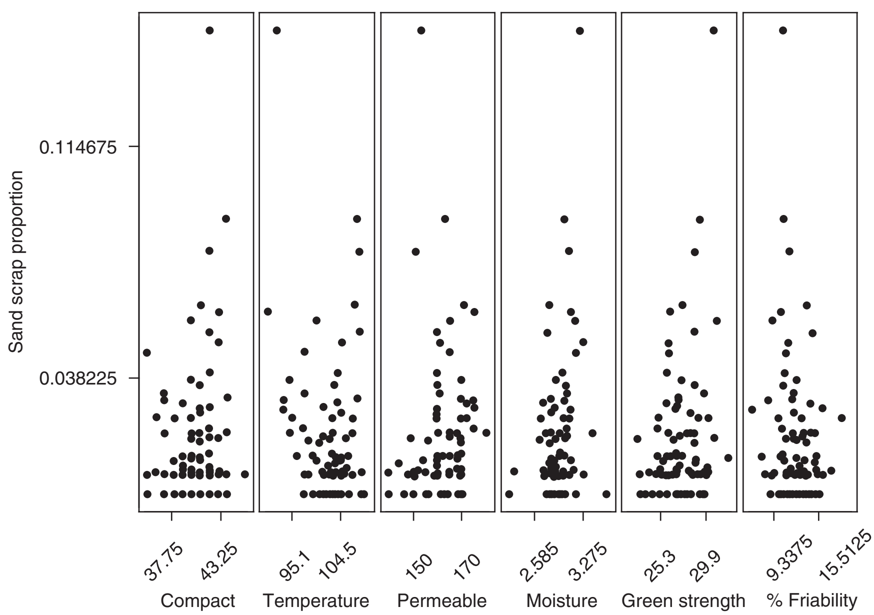


Figure C.13 Draftsman plot.

plotted points to show leakers and nonleakers. The data are given in the file *window leaks comparison*. A matrix scatter plot is available using:

Graph Æ Plot Æ Matrix Plot

Selecting all the inputs and using *class* as a group variable, as shown in Figure C.14, we obtain the matrix scatter plot in Figure C.15.

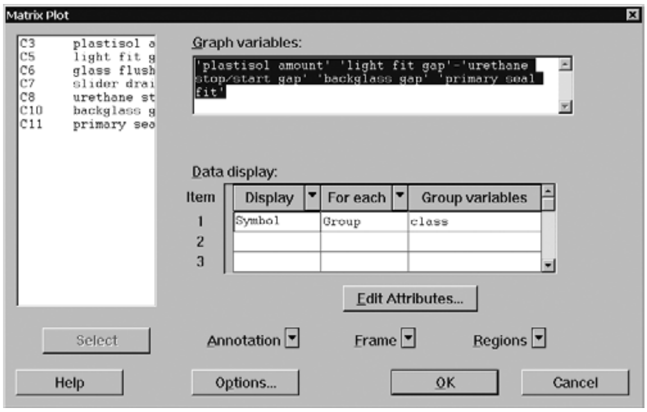


Figure C.14 Matrix (scatter) Plot dialog window.

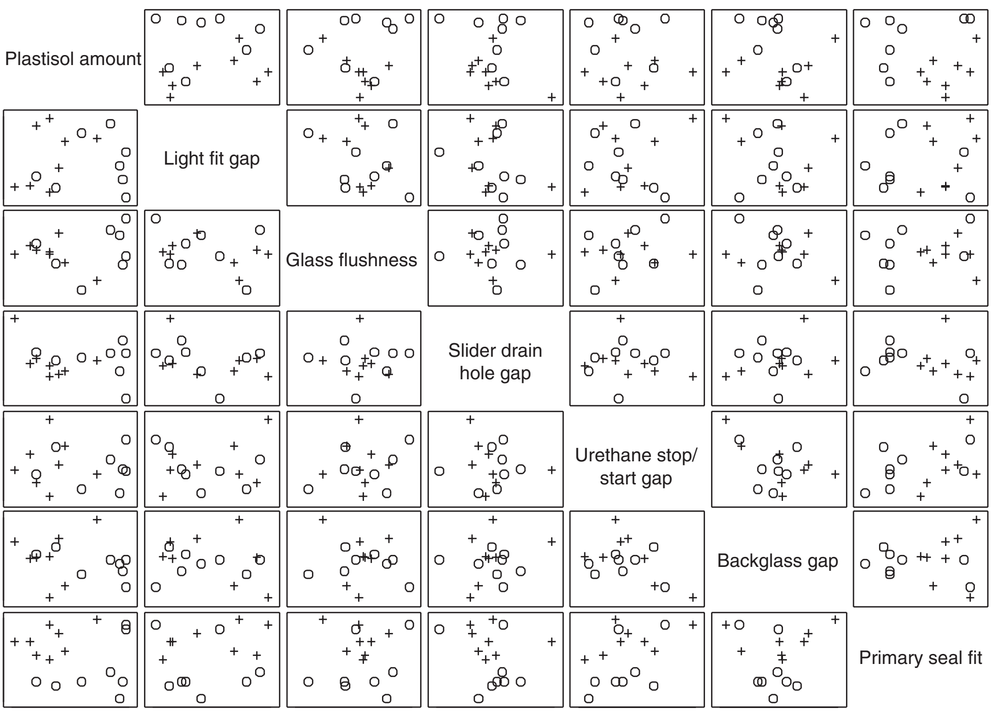


Figure C.15 Matrix scatter plot with labeled points.

Leaking windows are shown with circles, and nonleakers are shown with plus signs. There is a separation between leakers and nonleakers if we look simultaneously at the two inputs: plastisol amount and primary seal fit. The plot in the lower left corner of Figure C.15 is reproduced in a larger format in the right panel of Figure 12.2.

C.5 MULTIVARI CHART

A multivari chart displays the variation in the output due to families labeled by a discrete or categorical input. Multivari charts are available using:

Stat Æ *Quality Tools* Æ *Multi-Vari Chart*

The background and multivari investigation for the cylinder head scrap problem are described in Chapter 11. The side and end shift data for the 96 measured parts are stored in the file *cylinder head scrap multivari*.

To create a multivari chart, specify an output characteristic (called *response* in MINITAB) and up to four inputs (factor 1, factor 2, and so on). The order of the inputs affects the graphical display. Choose either time or the characteristic with the most different levels as the last factor, because this factor defines the labels on the horizontal axis.

Figure C.16 shows the Multi-Vari Chart dialog box. We recommend always using the Options button to check *Display individual data points*.

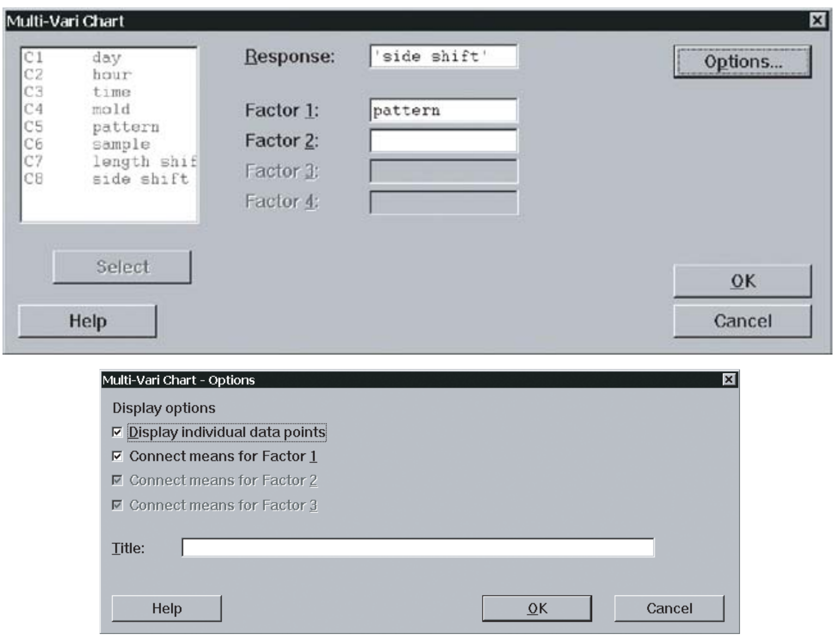


Figure C.16 Multi-Vari Chart dialog box with Options window.

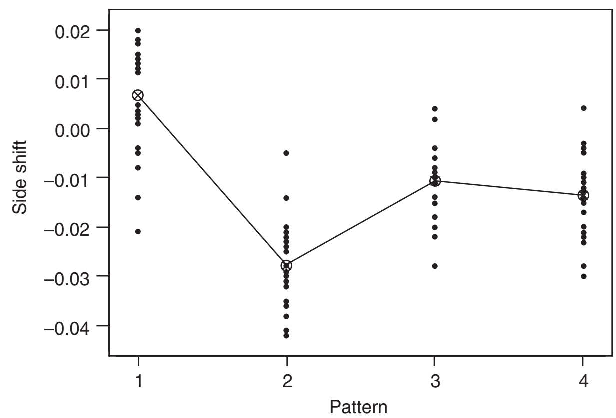


Figure C.17 Multivari chart for cylinder head scrap example.

Using the selections made in Figure C.16 results in the multivari chart given in Figure C.17.

Figure C.17 shows large systematic differences among the patterns (cavities). The variation within each pattern is about 60% of the overall variation.

The dialog box in Figure C.18 creates a multivari chart involving two inputs—time and pattern. From the resulting chart in Figure C.19, we conclude that the dominant cause acts only in the pattern-to-pattern family and does not involve the time-to-time family.

A multivari chart with three or four inputs can be difficult to interpret. Using a variety of different charts and orders for the inputs may help.

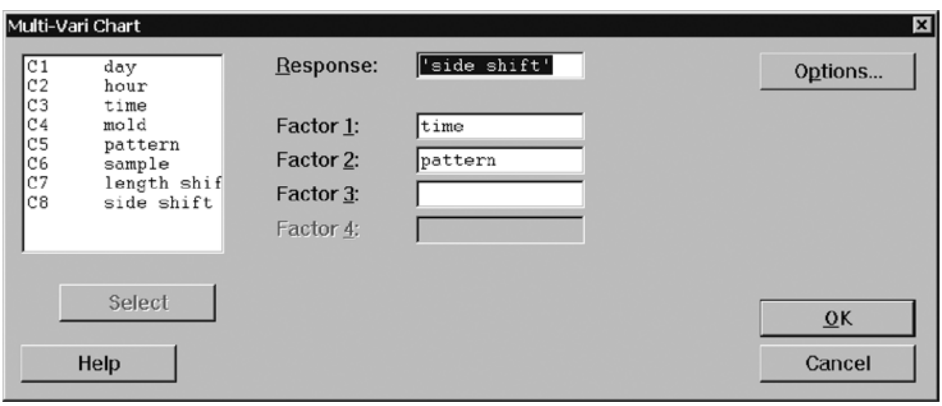


Figure C.18 Multi-Vari Chart dialog box showing two inputs.

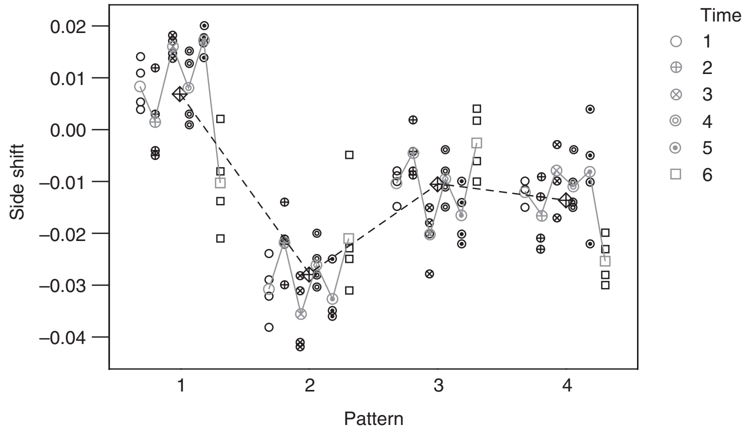


Figure C.19 Multivari chart of side shift versus pattern and time.

Exploring Haphazard Families

Consider the casting thickness example introduced in the supplement to Chapter 11, where the casting-to-casting family has haphazard effect. We want to determine if changing one of the other inputs results in changes in variation within different levels of the haphazard family.

To conduct the analysis, we first define a new output based on the standard deviation of original output within each level of group, where group = (time – 1)*24 + (cavity – 1)*4 + position. The variation within each group is due to causes acting in the casting-to-casting family. We use the MINITAB calculator, described in Appendix A, to define group. Next, we define a new output characteristic using:

Stat Æ Basic Statistics Æ Store Descriptive Statistics

The dialog box is shown in Figure C.20. This creates two new columns in the MINITAB worksheet. The first column (labeled “ByVar1”) gives the group number, while the second column gives the within-group standard deviation. To ease interpretation of the subsequent analysis, we relabel the second new column “group stdev.” Part of the resulting MINITAB worksheet is shown in Figure C.21.

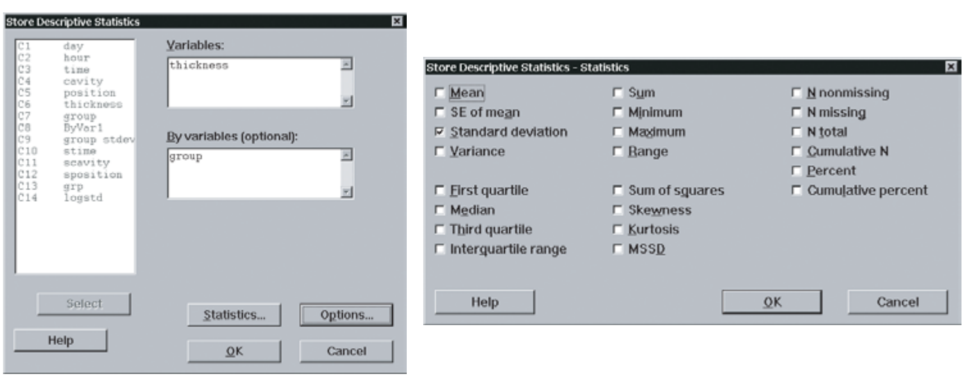
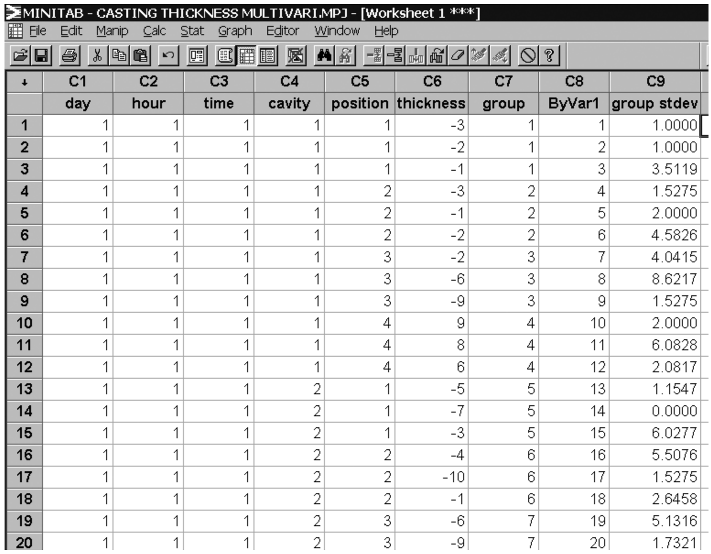


Figure C.20 MINITAB dialog box to store group standard deviation.



MINITAB - CASTING THICKNESS MULTIVARI.MPJ - [Worksheet 1 ***]

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9
	day	hour	time	cavity	position	thickness	group	ByVar1	group stdev
1	1	1	1	1	1	-3	1	1	1.0000
2	1	1	1	1	1	-2	1	2	1.0000
3	1	1	1	1	1	-1	1	3	3.5119
4	1	1	1	1	2	-3	2	4	1.5275
5	1	1	1	1	2	-1	2	5	2.0000
6	1	1	1	1	2	-2	2	6	4.5826
7	1	1	1	1	3	-2	3	7	4.0415
8	1	1	1	1	3	-6	3	8	8.6217
9	1	1	1	1	3	-9	3	9	1.5275
10	1	1	1	1	4	9	4	10	2.0000
11	1	1	1	1	4	8	4	11	6.0828
12	1	1	1	1	4	6	4	12	2.0817
13	1	1	1	2	1	-5	5	13	1.1547
14	1	1	1	2	1	-7	5	14	0.0000
15	1	1	1	2	1	-3	5	15	6.0277
16	1	1	1	2	2	-4	6	16	5.5076
17	1	1	1	2	2	-10	6	17	1.5275
18	1	1	1	2	2	-1	6	18	2.6458
19	1	1	1	2	3	-6	7	19	5.1316
20	1	1	1	2	3	-9	7	20	1.7321

Figure C.21 MINITAB worksheet showing group standard deviation.

To create the multivari charts for the group standard deviation, we also need to define new input columns that give the values of the original inputs for each level of the characteristic group. This is best accomplished using:

Manip Æ Code Æ Use Conversion Table

We define a new input column for cavity as shown in Figure C.22. We label the new input column "scavity," since it tells us the value of the cavity that corresponds to the group standard deviation column. We similarly define "stime" and "sposition."

With this preliminary work, we can now create the desired multivari charts for *group stdev*.

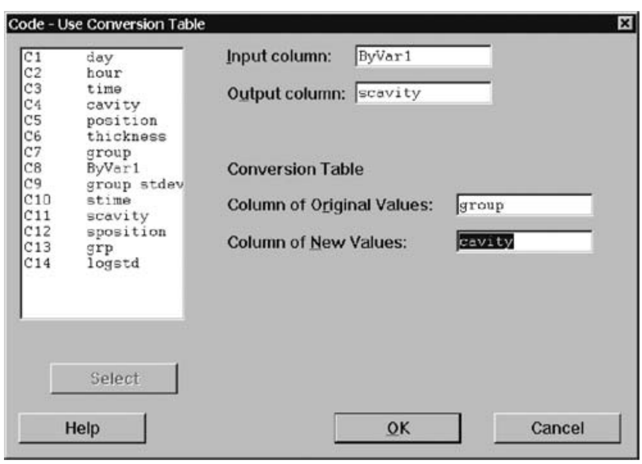


Figure C.22 Defining a new input column for cavity.

C.6 EXPONENTIAL SMOOTHING

When determining feasibility and deciding how to implement a feedback controller, we can fit an EWMA (exponential smoother) to an equally spaced time series of output values. See Chapter 18 and its supplement for details.

As shown in Figure C.23, to fit an EWMA we use:

Stat Æ *Time Series* Æ *Single Exp Smoothing*

To illustrate, we use the fascia film build example explored in Chapter 18. The data are given in the file *fascia film build baseline*. A plot of the results of the EWMA smoothing is given in Figure C.24.

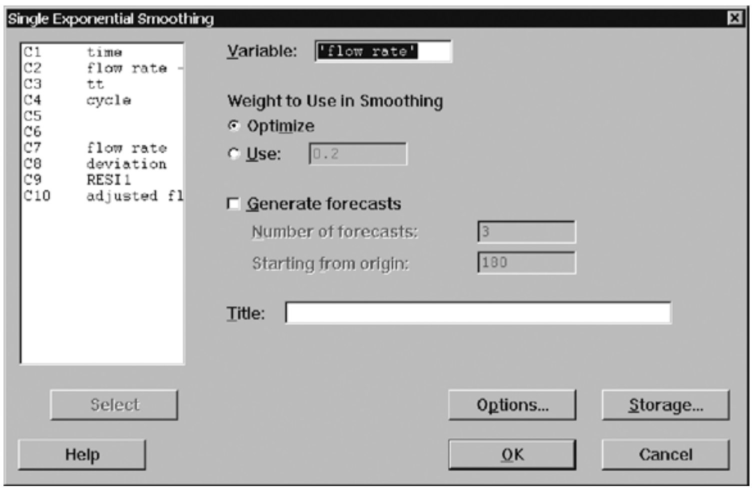


Figure C.23 MINITAB Single Exponential Smoothing dialog box.

We see that the time-to-time variation is captured well by the exponential smoothing. The smoothing constant (alpha) is 0.17, and the standard deviation of the smoothed series is given by the square root of the mean squared deviation (MSD). In the example, we have $\sqrt{3.83} = 1.96$.

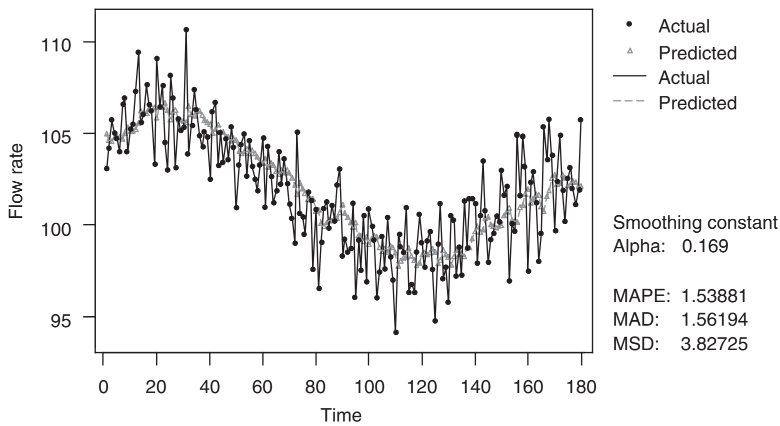


Figure C.24 EWMA smoothing of flow rate.

We can also use MINITAB to determine how the adjusted series will look (assuming no adjustment errors) by calculating the output minus the target value and fitting the single exponential smoother to the translated output. Saving the fitted values (using the Storage button from Figure C.23), we can determine the adjusted series by subtracting the fitted values from the original film build values.

Appendix D
Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is the main formal numerical analysis tool associated with the search for a dominant cause of variation. While in most investigations we draw conclusions from graphical displays, we can use ANOVA to:

- Assess a measurement system (Chapter 7).
- Compare contributions of two families to output variation (Chapter 10).
- Supplement the analysis of multivari investigations (Chapter 11).
- Set a new goal for a reformulated problem (Chapter 14).

ANOVA is applicable when the output is continuous and we have one or more discrete inputs. See Box et al. (1978) and Neter et al. (1996). Here we describe a few forms of ANOVA that we may need when applying the Statistical Engineering algorithm.

ANOVA partitions the overall variation (as quantified by the total sum of squares) into components attributable to various inputs or families. We recommend using ANOVA as a supplement to graphical displays if the results are unclear.

D.1 ONE-WAY ANOVA

A one-way ANOVA is appropriate for a continuous output and a single discrete input. We use the MINITAB menu selection:

Stat Æ *ANOVA* Æ *One-way*

Consider the cylinder head scrap example discussed in Chapter 11. The data are given in the file *cylinder head scrap multivari*. Figure D.1 shows a plot of the side shift (the output) versus pattern (an input). With ANOVA, we quantify how much of the variation in the side shift can be explained by differences among the four mold patterns.

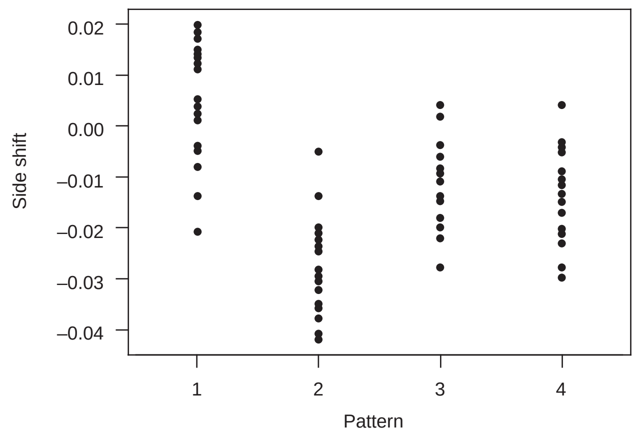


Figure D.1 Plot of side shift versus pattern.

Using the dialog box shown in Figure D.2, we get the ANOVA results:

```
One-way ANOVA: side shift versus pattern

Analysis of Variance for side shift
Source     DF        SS        MS        F        P
pattern     3 0.0146454 0.0048818    60.56    0.000
Error      92 0.0074168 0.0000806
Total      95 0.0220622
```

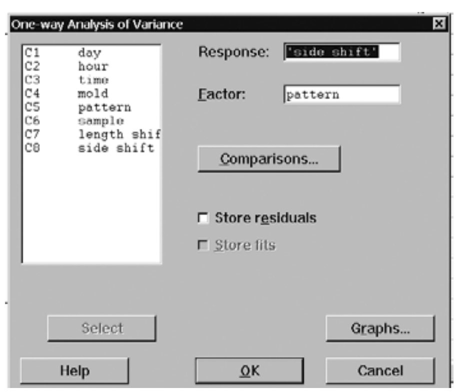


Figure D.2 MINITAB One-way ANOVA dialog box.

```
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  --------+---------+---------+---------
1          24  0.006833  0.011001(--*--)
2          24  -2.8E-02  0.008371  (--*--)
3          24  -1.1E-02  0.007812                (--*--)
4          24  -1.4E-02  0.008387              (--*--)
                                   -------+---------+---------+---------
Pooled StDev = 0.008979                -0.024    -0.012    -0.000
```

Pooled StDev provides an estimate of the variation within each pattern pooled across all patterns. That is, Pooled StDev provides an estimate of remaining output variation if we could eliminate all pattern-to-pattern differences in the average output. It estimates the variation attributable to all families other than the pattern-to-pattern family.

We can also use a one-way ANOVA model to estimate the variation due to the measurement system. In Chapter 7, we described a measurement investigation for the camshaft journal diameter example. The data are given in the file *camshaft journal diameter measurement*. A one-way ANOVA by part is:

```
One-way ANOVA: diameter versus part

Analysis of Variance for diameter
Source     DF        SS        MS        F        P
part        2  5008.014  2504.007  4383.00    0.000
Error      51    29.136     0.571
Total      53  5037.151
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  -------+---------+---------+---------
1          18   -10.901     0.638   *)
2          18     1.204     0.796                  (*
3          18    12.685     0.820                                  *)
                                   -------+---------+---------+---------
Pooled StDev =     0.756                -7.0       0.0       7.0
```

The measurement variation is estimated as 0.756, as given by Pooled StDev. This matches the result obtained in Chapter 7.

In interpreting the ANOVA results, we ignore the *p*-value and confidence intervals (CIs), which are formal statistical procedures. See the supplement to Chapter 10 for our reasons.

D.2 ANOVA FOR TWO OR MORE INPUTS

To fit an ANOVA model with two or more inputs we use the MINITAB menu selection:

Stat Æ *ANOVA* Æ *Balanced ANOVA*

To illustrate, we consider the casting thickness example previously discussed in the supplement to Chapter 11. The data are given in the file *casting thickness multivari*.

The MINITAB Balanced ANOVA dialog box is shown in Figure D.3. The notation shown in Figure D.3, using the symbol "|" between inputs, requests a model with all possible interaction terms. In the casting thickness example, this includes up to the three-way interaction involving all of time, cavity, and position.

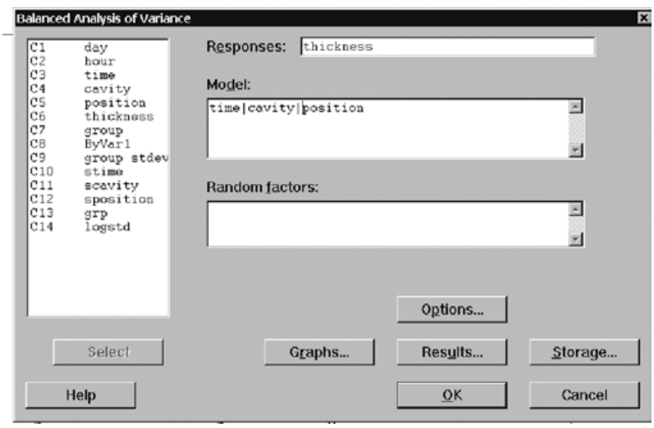


Figure D.3 MINITAB Balanced ANOVA dialog box.

The MINITAB results are:

```
ANOVA: thickness versus time, cavity, position

Factor     Type Levels Values
time      fixed     12      1      2      3      4      5      6      7
                            8      9     10     11     12
cavity    fixed      6      1      2      3      4      5      6
position  fixed      4      1      2      3      4

Analysis of Variance for thickness

Source                   DF         SS         MS       F      P
time                     11    9008.82     818.98   31.81  0.000
cavity                    5   16994.99    3399.00  132.00  0.000
position                  3   16697.24    5565.75  216.15  0.000
time*cavity              55    1544.02      28.07    1.09  0.311
time*position            33    7580.19     229.70    8.92  0.000
cavity*position          15    5363.43     357.56   13.89  0.000
time*cavity*position    165    1634.06       9.90    0.38  1.000
Error                   576   14832.00      25.75
Total                   863   73654.75
```

To compare the relative sizes of the families, we look at the sum of squares (SS) column. We see large, roughly equal-sized effects due to position, cavity, and error. The error sum of squares includes the effect of the casting-to-casting family and all interactions between the casting-to-casting family and the other families. The calculated sum of squares tells us approximately how much we could expect to reduce the total sum of squares if we could eliminate all variation due to the given family. It is complicated to translate these sums of squares into estimates for the standard deviation attributable to each family, but there is a rough correspondence. See the supplement to Chapter 11 for a complete discussion of the casting thickness example.

If some data are lost, the assumption of balance (equal number of observations at each level of each input) is violated and the balanced ANOVA analysis is no longer appropriate. For unbalanced data, we use a general linear model:

Stat Æ *ANOVA* Æ *General Linear Model*

Appendix E
Regression Models and Analysis

With regression, we model the relationship between an output characteristic and one or more inputs. We find regression models useful for the following tasks:

- Investigate variation transmission (chapters 10 and 11).
- Investigate the relationship between the output and inputs (Chapter 12).
- Set the goal for a reformulated problem (Chapter 14).
- Find a prediction equation for a feedforward controller (Chapter 17).
- Calibrate an adjuster (Chapter 18).

There are many good references. See Box et al. (1978), Ryan (1989), and Montgomery et al. (2001).

E.1 REGRESSION WITH A SINGLE INPUT

A regression model relating the output to a single input is:

$$output = b_0 + b_1 input + residual$$

The term $b_0 + b_1 input$ captures the effect of the input, and *residual* describes the variation in the output due to all other inputs. The regression analysis consists of estimating (also called *fitting*) the unknown constants b_0 and b_1 and the residual variation.

In MINITAB, regression analysis is available using:

Stat Æ *Regression* Æ *Regression*

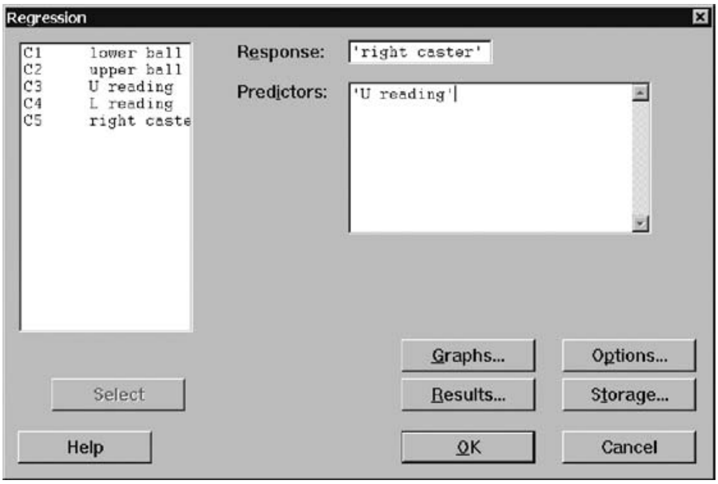


Figure E.1 Regression dialog box.

To illustrate, we use the truck alignment example introduced in Chapter 1 and discussed in Chapter 12. The data are given in the file *truck pull input-output*. Figure E.1 is the MINITAB Regression dialog box, used to fit a regression model with output right caster and input U-reading.

The corresponding MINITAB regression results are:

```
Regression Analysis: right caster versus U reading

The regression equation is
right caster = 4.55 + 0.171 U reading

Predictor        Coef     SE Coef         T        P
Constant      4.55157     0.03544    128.44    0.000
U reading     0.17115     0.05309      3.22    0.003

S = 0.1926      R-Sq = 27.1%     R-Sq(adj) = 24.5%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         1     0.38566     0.38566     10.39     0.003
Residual Error    28     1.03893     0.03710
Total             29     1.42459

Unusual Observations
Obs   U readin   right ca          Fit      SE Fit    Residual     St Resid
 22       2.09     5.1663       4.9089      0.1121      0.2574        1.64 X

X denotes an observation whose X value gives it large influence.
```

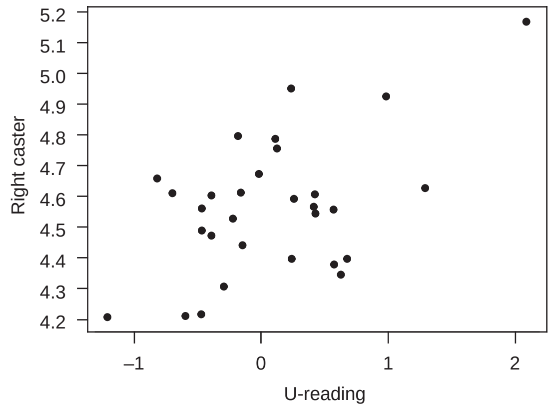


Figure E.2 Scatter plot of right caster versus U-reading.

The regression equation, right caster = 4.55 + 0.17*U-reading, provides a summary of the relationship between right caster and U-reading. In other words, the model suggest that a one-unit increase in U-reading would, on average, result in a 0.17-unit increase in right caster. However, the scatter plot in Figure E.2 suggests the regression model is not a very good summary. The value of $S = 0.193$ provides an estimate of the residual variation, the standard deviation in right caster if we could hold U-reading fixed. Holding the input U-reading fixed, we could reduce the variation in caster from around 0.22 (the baseline variation and, as it happens, also the caster variation in the regression data) to 0.19. We conclude that U-reading is not a dominant cause.

In the MINITAB regression results, the list of unusual observations flags outliers, observations that are either not close to the fitted regression line or that have a large influence on the estimated regression parameters. In this case, the one observation in the upper right-hand corner of the plot in Figure E.2 has been flagged because it has a large influence on the slope of the fitted line.

Fitted Line Plot

To superimpose the fitted regression equation on a scatter plot of the output versus the input, we use:

Stat Æ *Regression* Æ *Fitted Line Plot*

We illustrate using the crossbar dimension example covered in Chapter 12. The data from the investigation are given in the file *crossbar dimension input-output*. The left panel in Figure E.3 shows the resulting graph and numerical regression model summary when fitting a model for dimension as a linear function of barrel temperature. We see that the linear model fits quite well and (since the dimension variation in the regression data matches the baseline) that barrel temperature is a dominant cause. However, we also notice that the relationship between dimension and barrel temperature seems to be nonlinear.

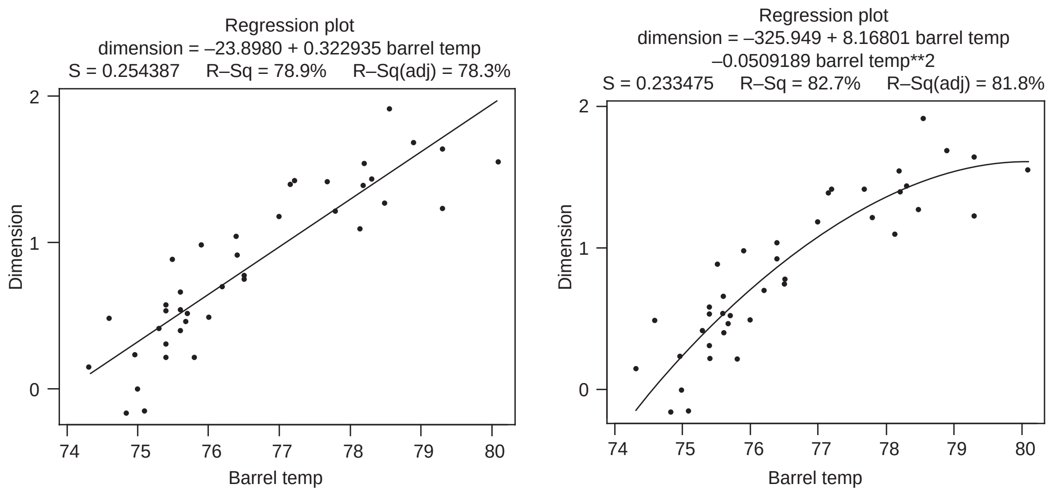


Figure E.3 Regression plots with fitted lines.

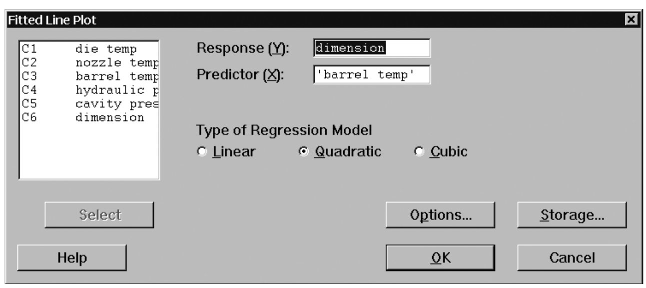


Figure E.4 Fitted Line Plot dialog box.

Using the fitted line plot dialog box as shown in Figure E.4, where we request a quadratic model, we get the plot given in the right panel of Figure E.3. In this way, we can explore the relationship between the input and output. Using the Options button we can also request a log transformation of the input, output, or both.

Best Subsets Regression

To quickly fit many possible models, we use automatic model building. See MINITAB help or Neter et al. (1996) for more details.

In investigations where many inputs are measured, the Best Subsets Regression dialog box, as shown in Figure E.5, can quickly summarize the results from fitting regression models with all possible single inputs.

Best subsets regression is available using:

Stat Æ Regression Æ Best Subsets

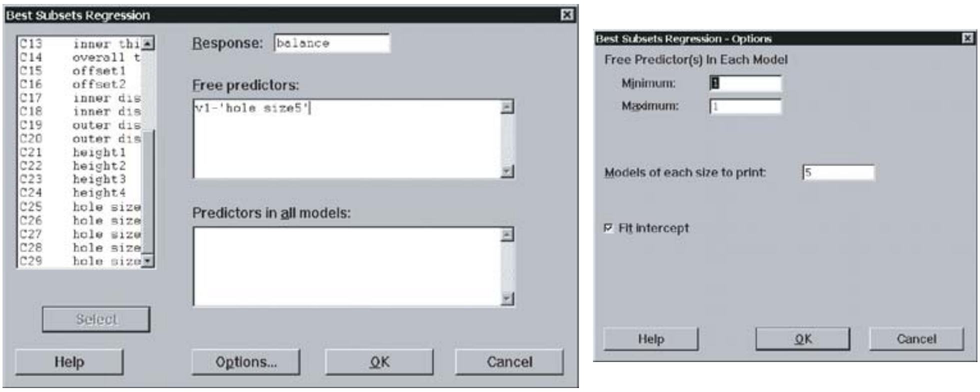


Figure E.5 Best Subsets Regression dialog box.

We use the best subsets regression options, as shown in the right panel of Figure E.5, to request a minimum and maximum of one free predictor (that is, input) and to display five models of each size to print. To illustrate, we use the brake rotor balance example discussed in one of the case studies. In a group comparison investigation, 26 foundry-related inputs were measured on 30 balance rejects and 30 well-balanced brake rotors. The data are given in the file *brake rotor balance comparison*. Previous analysis for this data was based on the binary output: balanced or not balanced. However, since the balance weight required for each rotor was also recorded, we can also analyze these data using regression models. Filling in the Best Subsets Regression dialog box as in Figure E.5 gives the following (edited) MINITAB results:

```
Best Subsets Regression: balance versus v1, v2, ...

Response is balance

                                              o i     i i o o       h h h h h
                                              u n o o o n n u u h h h h o o o o o
                                              t n v f f n n t t e e e e l l l l l
                                              e e e f f e e e e i i i i e e e e e
                                              r r r s s r r r r g g g g
                                                  a e e         h h h h s s s s s
                              v v v v v v v v t t l t t d d d d t t t t i i i i i
R-Sq R-Sq(adj) C-p        S   1 2 3 4 5 6 7 8 h h l 1 2 i i i i 1 2 3 4 z z z z z

62.6   61.9    2.4  0.30960                       X
48.2   47.3   24.9  0.36435                   X
44.1   43.1   31.3  0.37863                         X
34.7   33.6   45.9  0.40900                     X
 5.5    3.9   91.5  0.49213             X
```

The results summarize the five best regression models that involve only a single input. The models are ranked by largest R-Sq value (or equivalently, since all models have the same number of parameters, by the smallest value for *S*). Overall thickness variation is the single input that explains the most variation in the output. We would then look at the scatter plots that correspond to the best inputs. To determine the regression equation for any particular input, we need to fit the individual regression model.

Using the best subsets regression routine, we can avoid examining all scatter plots (and fitting all the corresponding regression models) involving the output and all possible single inputs. There are 26 such plots (models) in this case. We do not recommend the use of the best subsets regression routine unless the number of inputs is large. Looking at the individual plots is preferred because the regression summary can miss patterns like nonlinear relationships and the effect of outliers.

E.2 REGRESSION WITH MULTIPLE INPUTS

Fitting regression models with more than one input can be useful:

- As a screening tool to analyze the results of investigations to look at the relationship between inputs and the output (Chapter 12)
- For fitting a prediction equation for a feedforward controller (Chapter 17)

In cases where we have many input characteristics the regression model is extended to

$$output = b_0 + b_1 input_1 + ... + b_k \mathrm{input}_k + residual$$

Regression analysis with multiple inputs is available using:

Stat Æ *Regression* Æ *Regression*

To illustrate, consider the truck pull feedforward investigation discussed in Chapter 17. The goal of the investigation was to find a prediction equation for caster that could be used to build a feedforward controller. The data are given in the file *truck pull feedforward*. We fit a regression model to describe the relationship between the output left caster and the inputs given by the truck frame geometry using the Regression dialog box, as shown in Figure E.6.

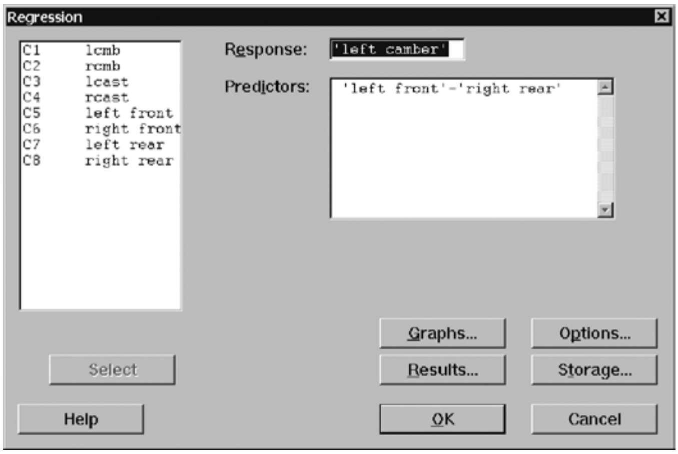


Figure E.6 Regression dialog box showing multiple inputs.

The MINITAB results are:

```
The regression equation is
left caster =  -18.6 + 1.24 left front + 0.677 right front + 0.140 left rear +
0.156 right rear

Predictor        Coef     SE Coef         T        P
Constant     -18.6097      0.6100    -30.51    0.000
left front    1.24243     0.04018     30.92    0.000
right front   0.67669     0.03585     18.88    0.000
left rear     0.13953     0.02763      5.05    0.000
right rear    0.15624     0.04162      3.75    0.000

S = 0.1760      R-Sq = 96.4%     R-Sq(adj) = 96.2%

Analysis of Variance

Source           DF         SS         MS         F        P
Regression        4     77.827     19.457    627.82    0.000
Residual Error   95      2.944      0.031
Total            99     80.771
```

In this example, the regression equation summarizes the relationship between the left caster and the four truck frame geometry inputs. The value of *S* in the regression results gives an estimate of the remaining variation in left caster (quantified in terms of standard deviation) if we hold *all* of the four truck frame inputs fixed or if we could perfectly compensate for their effects using a feedforward controller.

For regression models with many inputs, we need to use caution when trying to interpret the estimated regression parameters (given by the "Coef" column in the MINITAB regression results). We may interpret the parameter estimate corresponding to the input "left front" (1.24) as the expected change in the left caster for a unit change in left caster if we *hold all the other inputs fixed*. If the inputs used in the regression analysis do not vary independently, it may make no physical sense to think of changing one input while holding all the others fixed. In the truck pull example, where we want a prediction equation, we do not interpret the individual regression parameter estimates, so this is not a major concern.

Appendix F

Planning and Analysis of Designed Experiments

Designed experiments are used in the Statistical Engineering algorithm to:

- Verify a dominant cause (Chapter 13).
- Find adjusters that move the process center (Chapter 15).
- Find candidate settings that desensitize the process to variation in a dominant cause (Chapter 16).
- Find candidate settings that make the process robust (Chapter 19).

There is an extensive literature on designed experiments. See, for example, Box et al. (1978) and Montgomery (2001). In this appendix we show how to set up and analyze the types of experiments used in this book.

F.1 SETTING UP THE EXPERIMENTAL DESIGN

In this section, we consider the design of factorial experiments with two or more levels for each input. In MINITAB, to create a factorial design we use the menu selection:

Stat Æ *DOE* Æ *Factorial* Æ *Create Factorial Design*

To illustrate, we use the brake rotor balance verification experiment discussed in Chapter 13 and in one of the case studies. In the experiment, two levels for each of the three suspects—tooling, position, and thickness variation—were chosen to capture their full range of variation. The team decided to make eight rotors for each of the eight treatments. That is, there were eight runs with eight repeats. No treatment was replicated. Figure F.1 shows how the experimental plan was entered into MINITAB. In MINITAB, the inputs that are changed in an experiment are called *factors.*

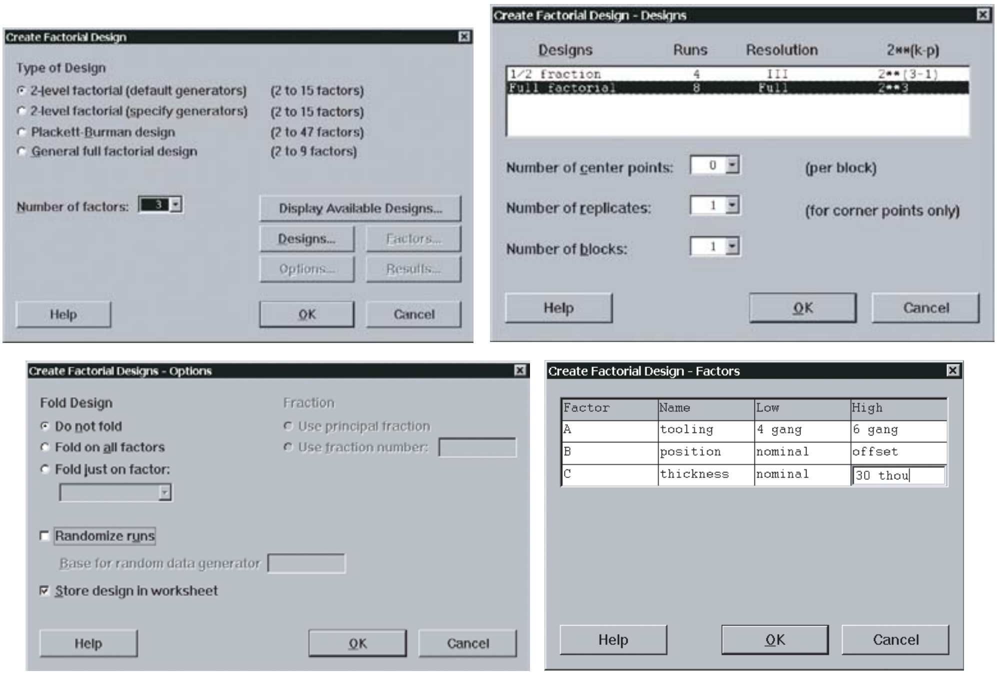


Figure F.1 Create Factorial Design dialog box.

For the choices shown in Figure F.1, MINITAB gives the summary:

```
Factorial Design

Full Factorial Design

Factors:       3   Base Design:         3, 8
Runs:          8   Replicates:             1
Blocks:     none   Center pts (total):     0

All terms are free from aliasing
```

The plan and the collected data are given in the file *brake rotor balance verification*. The worksheet with the data is shown in Figure F.2. Note that MINITAB has created four special columns in the worksheet: StdOrder, RunOrder, CenterPt, and Blocks. These columns are necessary to use any of the MINITAB subcommands under Stat Æ DOE Æ Factorial once the design has been created.

In Figure F.2, the columns r1, r2, ..., r8 give the balance weights for the eight rotors produced for each treatment. Much of the analysis in the brake rotor example uses the performance measure *average weight*. The average weight for each run is calculated as described in Appendix A. For experiments with a number of repeats for each run, such as the brake rotor balance verification experiment, we need to store the data in a different way

↓	C1	C2	C3	C4	C5-T	C6-T	C7-T	C8	C9	C10	C11	C12	C13	C14
	StdOrder	RunOrder	CenterPt	Blocks	Tooling	Position	Thickness	r1	r2	r3	r4	r5	r6	r7
1	1	8	1	1	4 gang	offset	30 thou	0.46	0.51	0.39	0.68	0.86	0.61	0.5
2	5	1	1	1	4 gang	offset	nominal	0.16	0.18	0.16	0.22	0.10	0.24	0.0
3	3	3	1	1	4 gang	nominal	30 thou	0.00	0.97	0.42	0.38	0.00	0.65	0.6
4	7	7	1	1	4 gang	nominal	nominal	0.11	0.09	0.09	0.09	0.04	0.08	0.0
5	2	2	1	1	6 gang	offset	30 thou	1.56	1.95	1.27	1.40	1.48	1.70	1.3
6	6	5	1	1	6 gang	offset	nominal	0.16	0.44	0.39	0.42	0.47	0.22	0.4
7	4	4	1	1	6 gang	nominal	30 thou	0.83	1.61	1.44	1.49	1.21	0.91	1.3
8	8	6	1	1	6 gang	nominal	nominal	0.04	0.03	0.04	0.03	0.03	0.05	0.0
9														
10														

Figure F.2 Experimental design setup in MINITAB worksheet.

to plot the output by treatment. Putting all the output values (that is, repeats) in a single column (use the stack columns command as described in Section A.5), we get the worksheet in Figure A.9.

Custom Designs

In some cases, the experiment we wish to conduct is not a standard two-level factorial design, or we already have the experimental data stored. In such cases, we can set up the experiment in MINITAB using:

Stat Æ *DOE* Æ *Factorial* Æ *Define Custom Factorial Design*

For instance, in the oil pan scrap example discussed in Chapter 16, the experiment had four inputs, with one at three levels and the other three at two levels each. We enter the design as in Figure F.3. For two-level factorial designs, we need to also use the Low/High dialog box to tell MINITAB what codes correspond to the high and low level of each input.

Entering the design as shown in Figure F.3 adds the columns StdOrder, RunOrder, CenterPt, and Blocks to the existing data as shown in Figure F.4. The data are available in the file *oil pan scrap desensitization*.

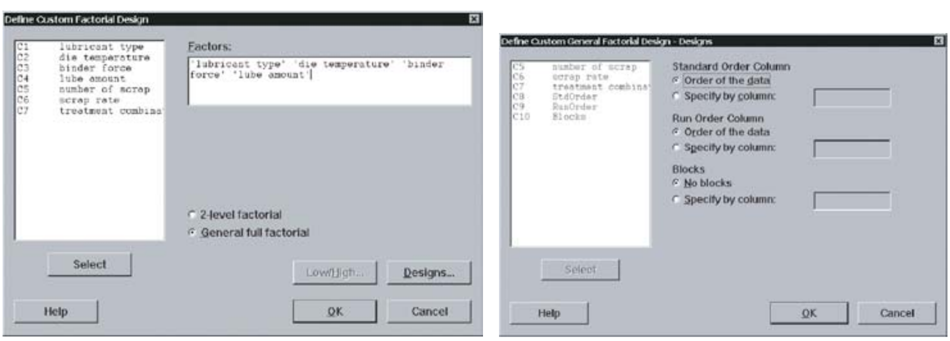


Figure F.3 Define Custom Factorial Design dialog box.

Worksheet 1 ***

↓	C1-T	C2-T	C3-T	C4-T	C5	C6	C7	C8	C9	C10
	lubricant type	die temperature	binder force	lube amount	number of scrap	scrap rate	treatment combination	StdOrder	RunOrder	Blocks
1	A	low	low	low	6	0.07	1	1	1	1
2	A	low	low	high	60	0.75	1	2	2	1
3	A	low	high	low	11	0.14	2	3	3	1
4	A	low	high	high	13	0.16	2	4	4	1
5	A	high	low	low	5	0.06	3	5	5	1
6	A	high	low	high	26	0.33	3	6	6	1
7	A	high	high	low	12	0.15	4	7	7	1
8	A	high	high	high	35	0.44	4	8	8	1
9	B	low	low	low	3	0.04	5	9	9	1
10	B	low	low	high	0	0.00	5	10	10	1
11	B	low	high	low	8	0.10	6	11	11	1
12	B	low	high	high	9	0.11	6	12	12	1
13	B	high	low	low	2	0.03	7	13	13	1
14	B	high	low	high	12	0.15	7	14	14	1
15	B	high	high	low	7	0.09	8	15	15	1
16	B	high	high	high	11	0.14	8	16	16	1
17	C	low	low	low	2	0.03	9	17	17	1
18	C	low	low	high	4	0.05	9	18	18	1
19	C	low	high	low	0	0.00	10	19	19	1
20	C	low	high	high	0	0.00	10	20	20	1
21	C	high	low	low	3	0.04	11	21	21	1
22	C	high	low	high	3	0.04	11	22	22	1
23	C	high	high	low	1	0.01	12	23	23	1
24	C	high	high	high	6	0.07	12	24	24	1
25										

Figure F.4 Results from custom design setup.

For a proper interpretation of this experiment, recall that this is a desensitization experiment. Three of the inputs—namely, lubricant type, die temperature, and binder force—are candidates, while lube amount is the dominant cause. As shown in Figure F.4, we number the treatments based on combinations of the candidates.

Fractional Factorial Designs

Designs for fractional factorial experiments (where only a subset of all possible treatment combinations are used—see the supplement to Chapter 15) are set up in MINITAB with the same command as full factorial designs, namely,

Stat Æ *DOE* Æ *Factorial* Æ *Create Factorial Design*

To plan a fractional factorial experiment, useful information about all possible 2^k factorial designs is given in MINITAB using the Display Available Designs button, as shown in Figure F.5.

To illustrate a fractional factorial experiment, we use the paint film build example discussed in the Chapter 19 exercises. The data are given in the file *paint film build robustness*. The goal was to explore the effect of five inputs on the variation in film build (as measured by the log standard deviation) over five consecutive panels. There were resources for 16 runs, a half fraction. We choose the design in MINITAB as shown in Figure F.6. We also used the Factors button to enter input names and levels.

Figure F.7 gives the resulting worksheet after adding appropriate labels, entering the experimental results, and calculating some performance measures. Performance measures were determined using the calculator function in MINITAB (see Appendix A).

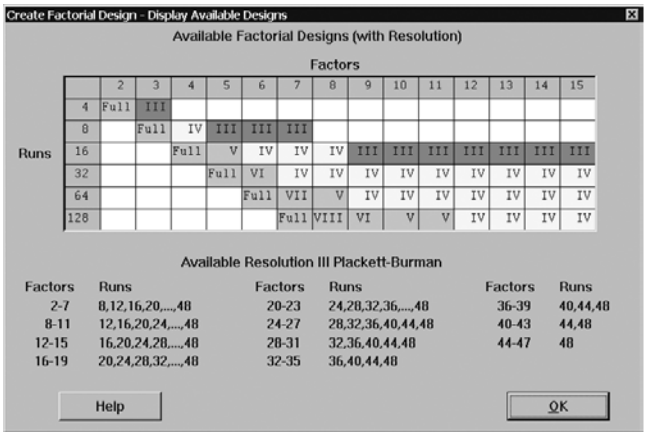


Figure F.5 Two-level factorial experiments available in MINITAB.

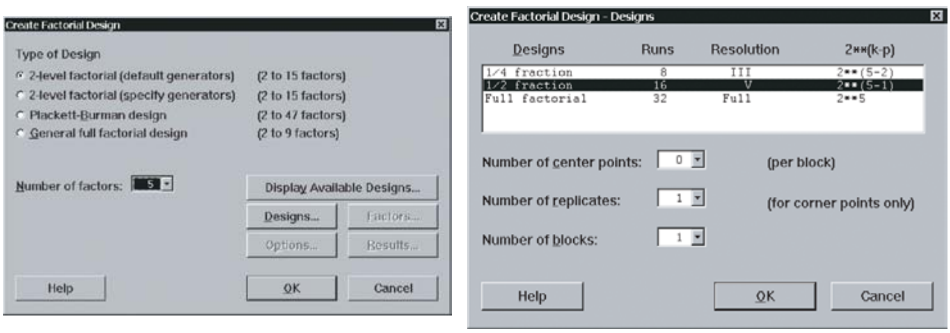
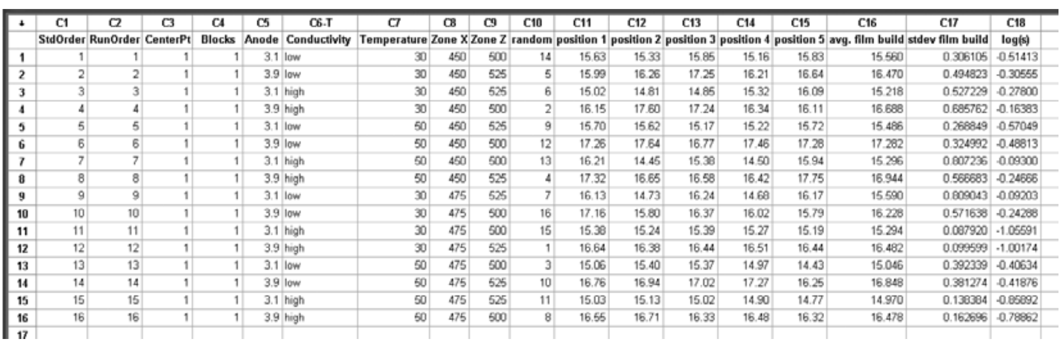


Figure F.6 Design setup for the paint film build robustness experiment.



↓	C1	C2	C3	C4	C5	C6-T	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
	StdOrder	RunOrder	CenterPt	Blocks	Anode	Conductivity	Temperature	Zone X	Zone Z	random	position 1	position 2	position 3	position 4	position 5	avg. film build	stdev film build	log(s)
1	1	1	1	1	3.1	low	30	450	500	14	15.63	15.33	15.85	15.16	15.83	15.560	0.306105	-0.51413
2	2	2	1	1	3.9	low	30	450	525	5	15.99	16.26	17.25	16.21	16.64	16.470	0.494823	-0.30555
3	3	3	1	1	3.1	high	30	450	525	6	15.02	14.81	14.85	15.32	16.09	15.218	0.527229	-0.27800
4	4	4	1	1	3.9	high	30	450	500	2	16.15	17.60	17.24	16.34	16.11	16.688	0.685762	-0.16383
5	5	5	1	1	3.1	low	50	450	525	9	15.70	15.62	15.17	15.22	15.72	15.486	0.268849	-0.57049
6	6	6	1	1	3.9	low	50	450	500	12	17.26	17.64	16.77	17.46	17.28	17.282	0.324992	-0.48813
7	7	7	1	1	3.1	high	50	450	500	13	16.21	14.45	15.38	14.50	15.94	15.296	0.807236	-0.09300
8	8	8	1	1	3.9	high	50	450	525	4	17.32	16.65	16.58	16.42	17.75	16.944	0.566683	-0.24666
9	9	9	1	1	3.1	low	30	475	525	7	16.13	14.73	16.24	14.68	16.17	15.590	0.809043	-0.09203
10	10	10	1	1	3.9	low	30	475	500	16	17.16	15.80	16.37	16.02	15.79	16.228	0.571638	-0.24288
11	11	11	1	1	3.1	high	30	475	500	15	15.38	15.24	15.39	15.27	15.19	15.294	0.087920	-1.05591
12	12	12	1	1	3.9	high	30	475	525	1	16.64	16.38	16.44	16.51	16.44	16.482	0.099599	-1.00174
13	13	13	1	1	3.1	low	50	475	500	3	15.06	15.40	15.37	14.97	14.43	15.046	0.392339	-0.40634
14	14	14	1	1	3.9	low	50	475	525	10	16.76	16.94	17.02	17.27	16.25	16.848	0.381274	-0.41876
15	15	15	1	1	3.1	high	50	475	525	11	15.03	15.13	15.02	14.90	14.77	14.970	0.138384	-0.85892
16	16	16	1	1	3.9	high	50	475	500	8	16.55	16.71	16.33	16.48	16.32	16.478	0.162696	-0.78862
17																		

Figure F.7 Fractional factorial design and output for paint film build experiment.

F.2 ANALYSIS OF THE EXPERIMENTAL RESULTS

To start the analysis, we plot the output by treatment number (see Appendix C). For the paint film build robustness experiment discussed in Section F.1, we get Figure F.8. We look for outliers and differences in the average or variation across the treatments. In some cases, the output by treatment plot is sufficient to draw appropriate conclusions. Note that if the multiple observations for each treatment come from repeats in the experiment (rather than replicates), the variation within each treatment likely underestimates the long-term variation for that treatment.

To analyze the experimental results more fully we use:

Stat Æ *DOE* Æ *Factorial* Æ *Analyze Factorial Design*

The resulting dialog box is given in Figure F.9. We specify the output (response) to analyze and choose the form of the model using the Terms button.

In this example, as shown in Figure F.9, we choose the performance measure "log(s)" as the output (upper left panel of Figure F.9) and select a model with all possible main effects and interactions (upper right panel of Figure F.9). Here we include interactions up to fifth order, since there are five inputs in the experiment.

From the Graphs dialog box (lower left panel in Figure F.9) we select a Pareto effects plot. In some cases, we may also want to have a numerical summary of some of the important effects. We can use the Results dialog box (lower right panel of Figure F.9) to request a display of the estimated average output for different levels of the input or inputs.

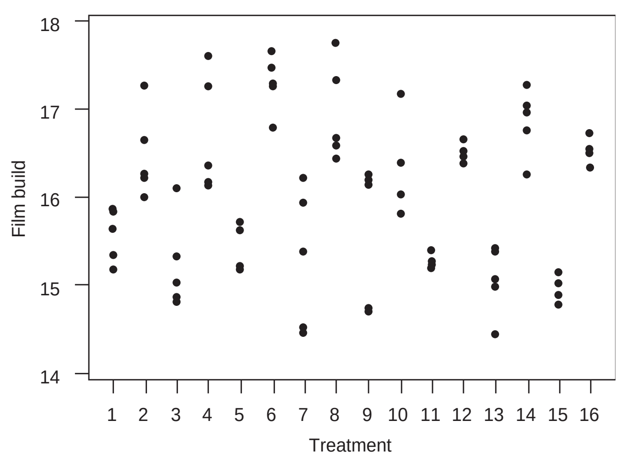


Figure F.8 Film build by treatment.

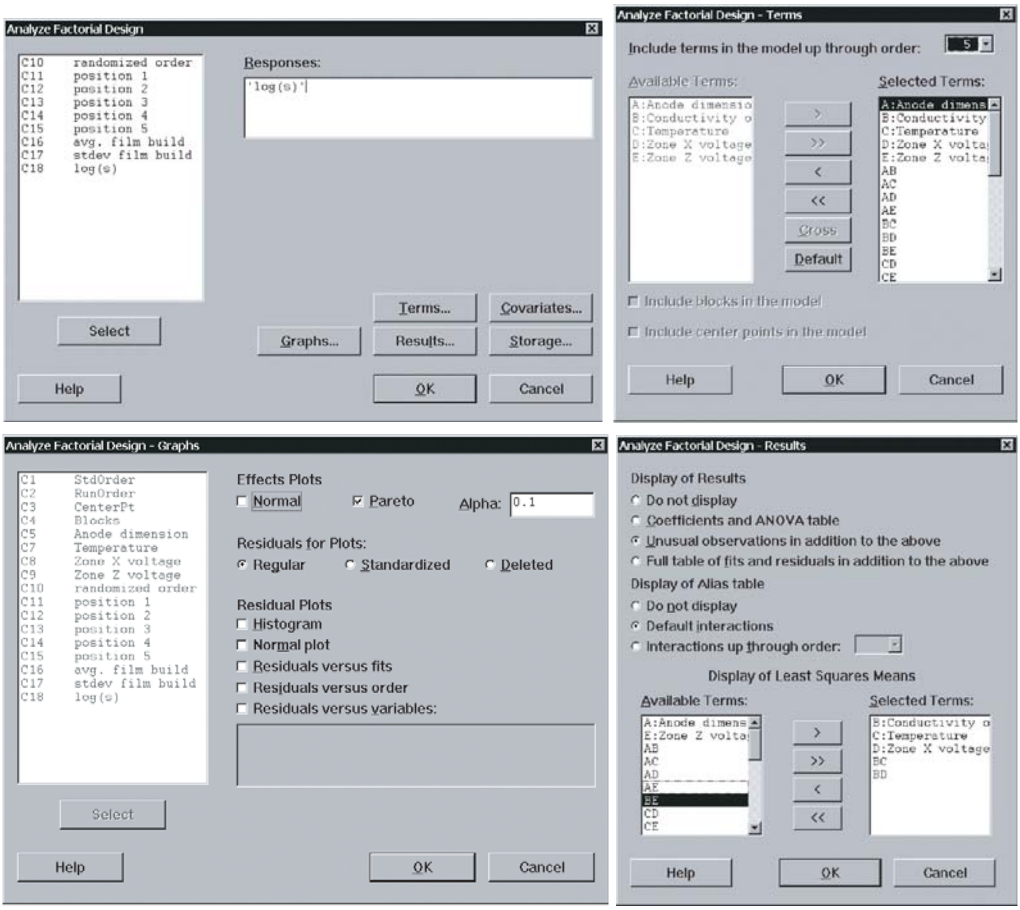


Figure F.9 MINITAB Analyze Factorial Design dialog box.

For the paint film build example, the resulting Pareto effects plot, which ranks the unsigned effects, is given in Figure F.10. In this plot any large effects (relative to the other effects) should be clearly evident. We see that the effects due to zone X voltage, conductivity, and the interaction between these two candidates are large.

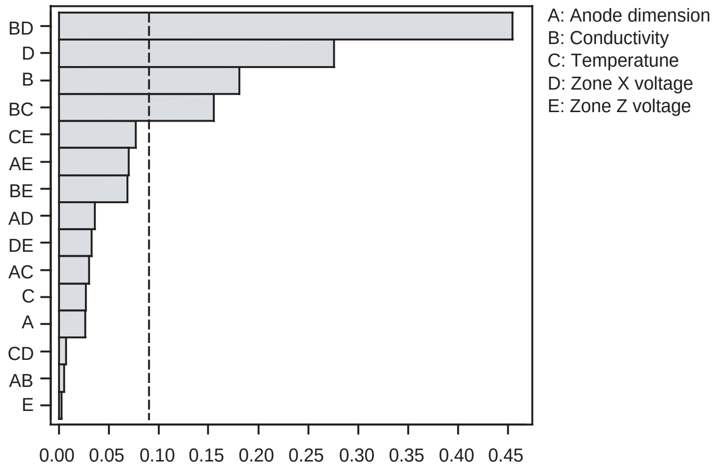


Figure F.10 Pareto plot of the effects for paint film build robustness experiment.

To make further sense of these experimental results, we consider factorial plots obtained using the MINITAB menu selection:

Stat Æ DOE Æ Factorial Æ Factorial Plots

We can request main effects, two-way interaction, and three-way interaction plots. Figure F.11 shows how we request main effects plots based on the performance measure log(s).

In the paint film build example, choosing main effects and interaction plots with all five candidates, we get the plots given in figures F.12 and F.13.

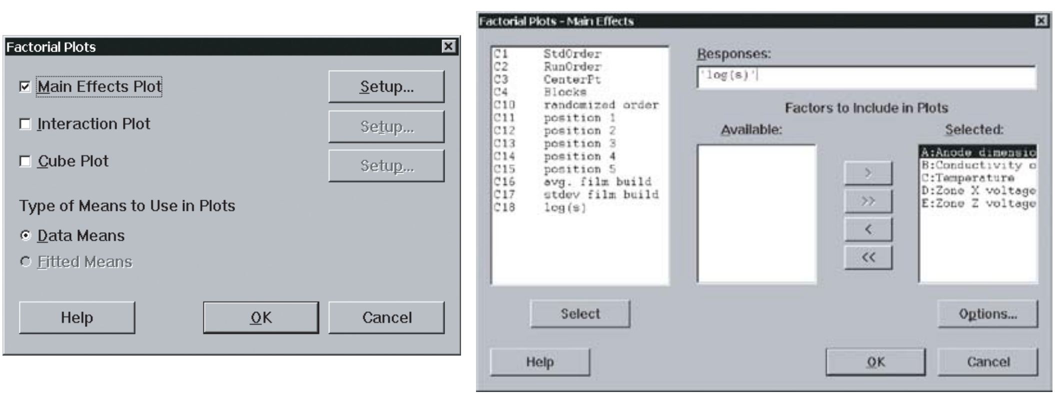


Figure F.11 Factorial Plots dialog box.

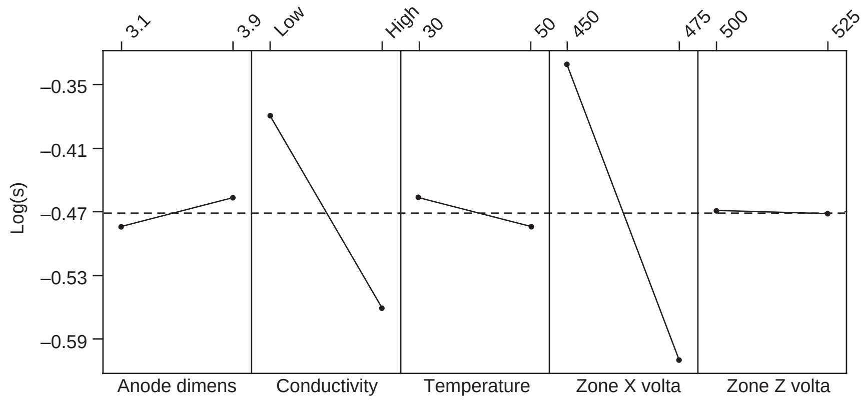


Figure F.12 Main effects plot (in terms of log(s)) for film build robustness experiment.

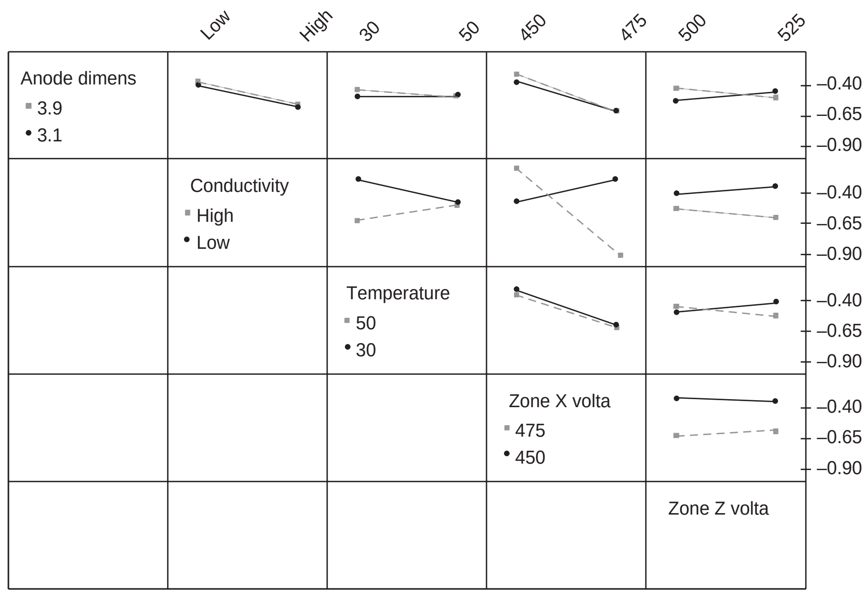


Figure F.13 Interactions plot (in terms of log(s)) for film build robustness experiment.

Note that in the paint film build robustness experiment, the design was resolution V; thus, the main effects and interactions can be estimated separately. If the design is resolution IV, some two-way interactions are aliased with other two-way interactions. In particular, for the standard half fraction resolution IV designs, half the individual interaction plots show the same information. Similarly, in resolution III design, the main effects are confounded with interactions. With resolution III designs, we need only consider the main effects plots, since the interaction plots will not provide any new information.

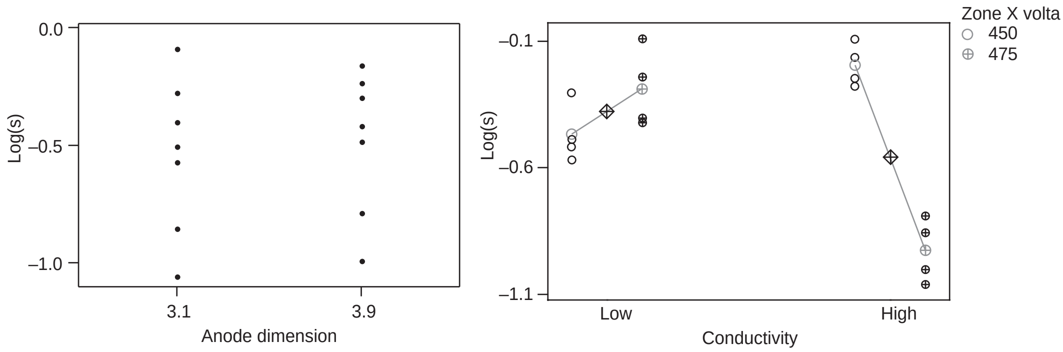


Figure F.14 Main effects and two-way interaction plots showing individual observations.

From figures F.12 and F.13, we conclude there are large effects due to conductivity, zone *X* voltage, and the interactions between conductivity and zone *X* voltage and between conductivity and temperature. Recall that smaller log(*s*) is better. Since there are large interactions, we draw conclusions based on the interaction plot. Figure F.13 suggests high zone *X* voltage, high conductivity, and low temperature are best. We are fortunate that high conductivity is best in both large interactions.

Note that the main effects and interaction effects plots produced by MINITAB show only averages. We can use box plots (showing individual output values) and multivari plots (see Appendix C) to create alternative displays of main effects and two-way interactions that show individual observations, as illustrated in Figure F.14.