Using Baseline Data in Problem Solving

Stefan H. Steiner and R. Jock MacKay
Business and Industrial Statistics Research Group
Dept. of Statistics and Actuarial Sciences
University of Waterloo
Waterloo, N2L 3G1 Canada

For many processes an improvement goal is to reduce costs and improve quality by reducing variation. 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.

Establishing a baseline is the first step (or one of the first steps) in most problem solving (variation reduction) strategies. For example, it is one of the necessary activities in the Measure stage of DMAIC in Six Sigma (Breyfogle, 2005). It is also the first stage of the Statistical Engineering algorithm (Steiner and MacKay, 2005) illustrated in Figure 1.

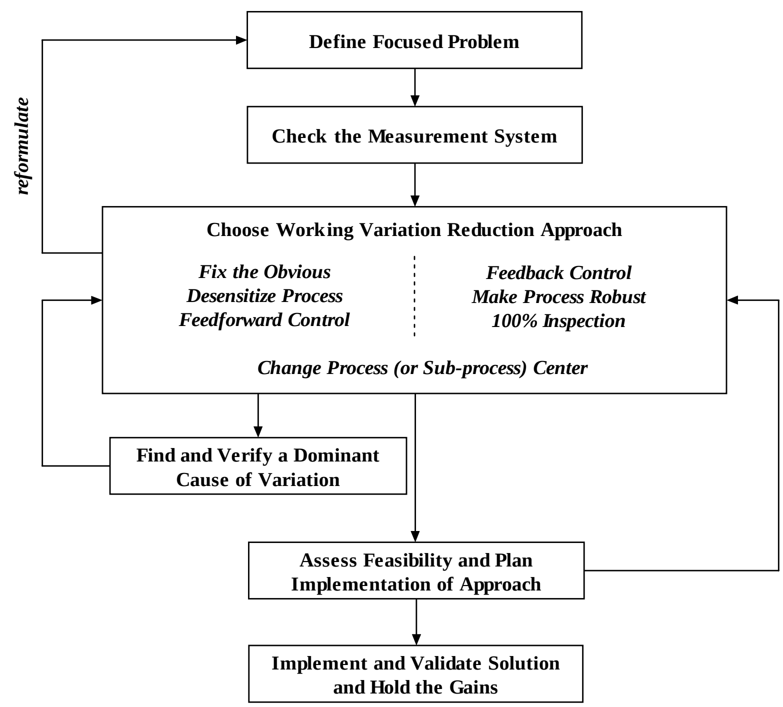

Figure 1: Statistical Engineering Variation Reduction Algorithm

We define the **problem baseline,** also called simply the **baseline,** as a numerical or graphical summary of the current process performance. In other words, the baseline quantifies the size and nature of the variation reduction problem we want to address. In our view, the results of the baseline investigation should be used to:

1. help set the problem goal,
2. allow validation of a potential solution (if and when one is found), and
3. help plan and analyze subsequent investigations when searching for a cause or a solution.

The first two uses are clear, and commonly conducted. However, it is our contention that, unlike most current practice, the information gained in the baseline can be exploited in planning and analyzing subsequent investigations designed to gain the process knowledge necessary to solve the problem. Some may argue this is common sense, as we should always use any prior information as a guide when planning any investigations. However, in our experience, in practice, mistakes and oversights are common. In addition, explicitly acknowledging the direct use of the baseline in problem solving suggests a particular plan for the baseline investigation itself. We give our recommendations in the next section.

Our proposed use of baseline information can be thought of as akin to George Box’s (1999) sequential learning idea. We apply sequential learning to problem solving in general, not just response surface methods. This makes sense since problem solving involves a series of process investigations to learn about the process.

We shall use the Statistical Engineering algorithm (Figure 1) to explicitly illustrate the use of the baseline information in problem solving. The points we make are also relevant for other approached like Six Sigma (Breyfogle, 1999). We hope that by illustrating the potential benefits of baseline information and making suggestions for the planning and analysis of the baseline investigation and subsequent investigations, problem solvers will make more systematic use of the baseline information and achieve better results in less time.

We use a crossbar dimension example to illustrate ideas. In the manufacture of an injection molded plastic base, as shown in Figure 2, there was excessive variation in a key crossbar dimension, measured as the difference from a nominal value. With rescaling, the target dimension was 1.0 and the specifications were 0 to 2.0 thousandths of an inch (thou). In a later assembly process many electronic components are inserted into spaces in the plastic base. Problems occurred due to both breakage when spaces were too small and loose assembly when spaces were too large. The crossbar dimension of the plastic base was used as a surrogate for all the internal dimensions. If crossbar dimension was small (large) the spaces were generally too small (big). The goal was to reduce variation in the crossbar dimension.

Figure 2: Plastic Base

Planning and Analysis of the Baseline Investigation

To assess the problem baseline we need an empirical investigation that will allow us to estimate the long term properties (mean, standard deviation, etc.) of the critical process output(s). For the purposes of illustration, we assume an output of interest and a performance measure are given. There are many feasible choices for a performance measure – standard deviation, capability ratio, etc.

In the empirical investigation the sampling scheme is critical. We argue against the standard recommendation of a random sample. Random sampling is often not feasible logistically and does not allow us to accomplish all the goals we set for the baseline investigation. The important point is that the sampling scheme needs to cover a time period long enough to see the full range or extent of variation in the output. We wish to avoid study error, as we want the baseline results to reflect the long term performance of the process.

We propose a plan for the baseline investigation that is designed to help guide our problem solving. Specifically, to accomplish the goals, the baseline investigation should allow us to

- estimate the long-term performance measure
- estimate the full extent of variation in the output
- determine the nature of the output variation over time

Instead of random sampling, we propose a systematic sampling plan that provides information about the time nature of the output variation. This baseline investigation can be thought of as a multivari investigation focused on the time family of variation. See Snee (2001) for more details on multivari investigations. In this light our suggestion for the baseline investigation is similar to the suggestion in Shainin (1993) to start problem solving with a multivari investigation.

In the crossbar dimension example, to quantify the problem, a team planned and executed a baseline investigation where six consecutive parts were selected from the process each hour for five days. This choice was expected to provide ample time for the process output to vary over its normal range, and give a large enough sample size to reasonably estimate the process variation. Numerical and graphical summaries of the 240 observations in the baseline are given below and in Figure 3. We suggest always using both a histogram and some sort of run chart. The right panel in Figure 3 gives a multivari chart that illustrates how crossbar dimension varies over time. The six consecutive values each hour are plotted at the same horizontal location. The vertical dashed lines show the division into the five days.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
crossbar dimension	240	0.8383	0.8300	0.8275	0.4497	0.0290

Variable	Minimum	Maximum	Q1	Q3
crossbar dimension	-0.2500	2.1100	0.6025	1.0900

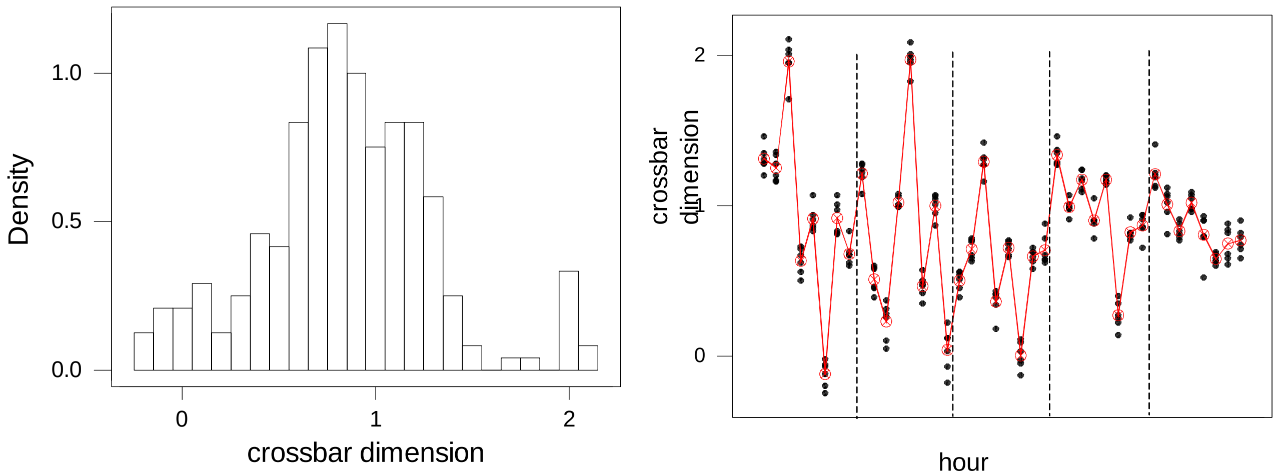

Figure 3: Histogram and Multivari Chart for Crossbar Dimension Baseline Data

We define the full extent of variation as the range within which the vast majority of output values lie. The range (minimum to maximum) defines the full extent of variation when the sample size is reasonably large (i.e. the sample size is in the hundreds) and there are no wild outliers. More generally, for a histogram with a bell-shape, the full extent of variation corresponds to the range of output values given by the average plus or minus three times the standard deviation. This way the full extent of variation covers 99.7% of output values using a Gaussian assumption. To define the full extent of variation we ignore rare outliers. For binary and discrete outputs the full extent of variation is given by all the output values seen in normal production.

By sampling parts consecutively at regular intervals we are able to distinguish between situations where the output varies quickly (part-to-part) or slowly (say, day-to-day) or somewhere in between. This information is valuable both to help us chose the study population for subsequent investigations and to give us clues about the possible major causes of variation.

From Figure 3 and the numerical summaries, we see that the full extent of crossbar dimension variation is –0.25 to 2.1 thou (indicated by the dashed lines on subsequent plots) and the output variation acts hour-to-hour with some evidence of day-to-day differences. The variation in crossbar dimension for consecutive parts (bases) is small. The standard deviation of the crossbar dimension is 0.45. The team set the goal to reduce the standard deviation to less than 0.25 thou. There was no immediate explanation for the smaller variation in crossbar dimension observed on the fifth day. Note that had there been a large day effect, i.e. had day averages been very different, the baseline investigation was (probably) not conducted over enough days to capture the long-term performance. In that case the team should collect data over some additional days before drawing conclusions.

Due to the time nature of the crossbar dimension variation, the team concluded that the study population for further observational investigations should be hours and days. We expect to see the full extent of variation in the output over that time frame. Investigations conducted over a shorter time frame, say only an hour, would not see the full extent of output variation and thus not reflect the long term behavior of the process and thus not provide clues about the major causes of output variation.

Next, we illustrate the use of the baseline information in subsequent investigations needed at various stages of the Statistical Engineering algorithm.

Using the Baseline to Help Check the Measurement System

After establishing the baseline, the next step in problem solving (See Figure 1) is to assess the measurement system for the output. The goal of the measurement investigation is to compare the size of the measurement variation and the process variation. We want to check if the measurement system is a large source of variation and whether it is adequate to support further process investigations. If the measurement variation is large, improving the measurement system is necessary before proceeding with problem solving and may solve the original problem.

A generic plan for measurement assessment is to measure the same parts repeatedly over a variety of conditions and times. We plan to use the baseline estimate of the overall variation (i.e. the combined effect of the process and measurement) to improve the precision of the conclusion about the measurement variation. If we assume independence, i.e. the part dimension does not effect the measurement variation, we have $\sigma_{overall} = \sqrt{\sigma^2_{process} + \sigma^2_{measurement}}$. The measurement investigation will provide an estimate for $\sigma_{measurement}$, combining that with the estimate for $\sigma_{overall}$ given by the baseline allows us to solve for $\sigma_{process}$.

In the measurement system assessment investigation, we suggest selecting three parts chosen (from the baseline) to cover the full extent of variation observed in the baseline. We select one large, one small and one intermediate sized part. The benefits of choosing extreme parts are explored in more detail in Browne et al. (2009a, 2009b) where a more complicated analysis that incorporates the initial dimension used to select the parts is presented. Note the difference from the usual suggestion in gage R&R investigations for 10 randomly selected parts (AIAG, 2003). The traditional gage R&R estimates both $\sigma_{measurement}$ and $\sigma_{process}$ using only the measurement investigation data.

In the crossbar dimension example, three parts were measured five times each on two separate days. Based on what we observed in the baseline, we expect to see the full extent of output variation within two days. The results are shown graphically in Figure 4 and the one-way analysis of variance (ANOVA) numerical results that follow.

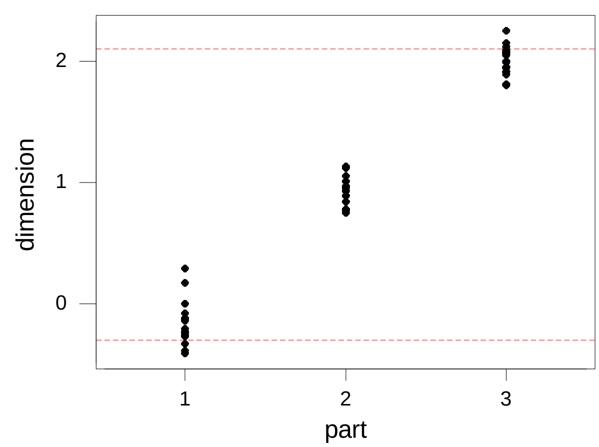

Figure 4: Crossbar Dimension Measurement Investigation Results
dashed horizontal lines show the full extent of variation in the baseline

One-way ANOVA: dimension versus part

```
Analysis of Variance for dimension
Source     DF        SS        MS        F        P
part        2   42.5111   21.2556  1064.87    0.000
Error      51    1.0180    0.0200
Total      53   43.5291
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ----+---------+---------+---------+--
1          18   -0.1722    0.1797   (*
2          18    0.9222    0.1166                (*)
3          18    2.0011    0.1183                               (*)
                                   ----+---------+---------+---------+--
Pooled StDev =   0.1413              0.00      0.70      1.40      2.10
```

In Figure 4 we have added horizontal dashed lines to show the full extent of output variation (-0.3 to 2.1) seen in the baseline. From the ANOVA results we estimate $\sigma_{measurement}$ = 0.14. The baseline standard deviation was 0.45. Thus, we estimate $\sigma_{process}$ = $\sqrt{0.45^2 - 0.14^2}$ = 0.43. To draw conclusions, Steiner and MacKay (2005) suggest estimating the discrimination ratio $\sigma_{process}/\sigma_{measurement}$; we obtain 3.07 and conclude that the measurement system is adequate. We are clearly able to distinguish between the three parts and the measurement variation is small.

Using the Baseline to Help Search for a Dominant Cause

Often the next step in problem solving is to search for a dominant cause of output variation (Juran and Gryna, 1980). If the dominant cause can be identified we hope to be able to use this knowledge to find a way to reduce output variation. A dominant cause is a process input that, if held fixed, would substantially reduce the variation in the output. Assuming independence, and denoting the standard deviation as “sd”, we have

$$\text{sd}_{\text{overall}} = \sqrt{\text{sd}^2_{\text{due to spccific cause}} + \text{sd}^2_{\text{due to all other causes}}}$$

A special case of this formula was discussed in the measurement assessment section. The notion of a dominant cause uses the Pareto principle applied to causes (Juran and Gryna, 1980). For a dominant cause, the residual variation, i.e. $\text{sd}_{\text{due to all other causes}}$, must be relatively small. Figure 5 shows the percent reduction in the overall variation possible if we eliminate the contribution due to a specific cause. There is little improvement unless we reduce the contribution of a cause that is dominant.

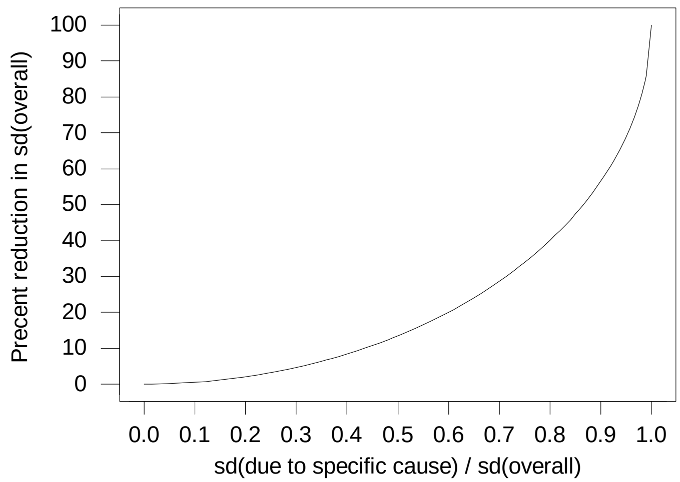

Figure 5: Reduction in Variation if We Remove a Cause
Contributing a Given Proportion of the Overall Variation

In searching for a dominant cause we can use the baseline information in many ways. First, the baseline results themselves can be used to eliminate many inputs as suspect dominant causes. For instance, if the output varies slowly, any quickly varying cause can not be a dominant cause. Second, as with planning for the measurement assessment investigation, the time nature of the output variation suggests a reasonable choice for the study population time frame. To provide clues about the dominant cause using an observational investigation, we want to be sure the dominant cause has acted during the investigation. An alternative is to use leverage and specially select parts with extreme output values for our investigation, as we did in the measurement assessment investigation. Finally, we can use the full extent of variation to *check* that the dominant cause has acted during the investigation. There is no sense in finding causes that "explain" only a small part of the output variation. If the output variation in an investigation does not closely matched the full extent of variation seen in the baseline we conclude that the dominant cause did not act. Then, it is not possible to generate clues about the identity of the dominant cause using the investigation results.

In the crossbar dimension example, what clues about the dominant cause are provided by the baseline and measurement investigations? First, we know the dominant cause must vary the same way over time as the output. Thus, from the baseline, the dominant cause is not an input that varies quickly, say part-to-part. Otherwise, we would not have seen the time pattern of variation in crossbar dimension in the right panel of Figure 3. Second, we showed the dominant cause does not act in the measurement system.

To search for a dominant cause, the team planned, what we call, an input/output investigation where they measured five inputs and the crossbar dimension on 40 parts haphazardly selected over a two day period. The five inputs were all though to be possible substantial causes and all matched the pattern observed in the baseline, i.e. all five inputs varied slowly over time. There is no sense in considering inputs that vary quickly as possible large causes. The investigation was conducted over two days since the baseline results suggest we should see the full range of values of the dominant cause(s) within that time.

The input/output investigation results are summarized using the two scatterplots of an input versus an output given in Figure 6. The plots for the remaining three inputs showed no pattern, i.e. looked similar to the left panel of Figure 6. In the scatterplots, the horizontal dashed lines give the full extent of variation seen in the baseline. We conclude, first of all, that the dominant cause acted in the investigation since the crossbar dimension values seen in the investigation span close to the full extent of variation. Second, we conclude that barrel temperature is a strong suspect for the dominant cause. If we could hold barrel temperature fixed, (it appears) there would be much less variation in the crossbar dimension. The other four inputs are eliminated as possible dominant causes. Note that at this point we could fail to find a dominant cause if it is measured with large measurement variation. We should ideally have checked the measurement systems for all inputs (suspected dominant causes).

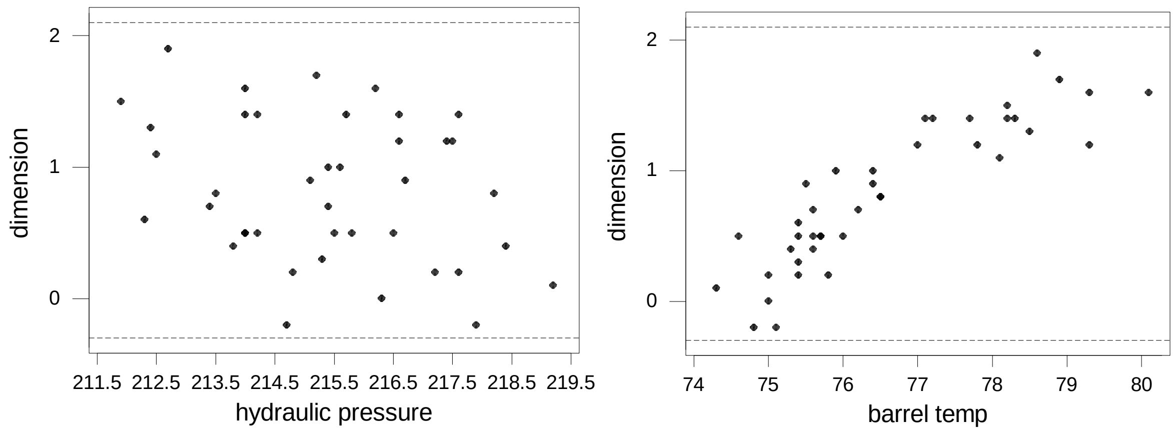

Figure 6: Scatterplots of Crossbar Dimension by Hydraulic Pressure and Barrel Temperature
dashed horizontal lines show the full extent of variation in the baseline

Using the Baseline to Help Verify a Dominant Cause

We want to be sure that the suspected cause, here called a **suspect**, is dominant before moving on to trying to use this information to help solve the original problem. We need to verify the suspect because in the search for the dominant cause, we might have inadvertently ruled out a family of causes that contains the dominant cause or been mislead by confounding. To verify that a suspect is a dominant cause, we use an experimental plan where the value of the suspect is deliberately manipulated and we observe the effect on the output. Since we suggest first searching for a dominant cause primarily using observational plans and the method of elimination (Steiner and MacKay, 2005) the verification experiment should not need to consider many suspects. It is used only to verify clues previously attained.

We can use the baseline information to help plan and analyze the verification experiment. The time nature of the output variation in the baseline helps us define an experimental run, and determine the importance of replication (i.e. choosing the number of runs) and randomly assigning the order of the runs to reduce the risk of misleading results due to confounding. To draw conclusions, we compare the output variation observed in the verification experiment to that seen in the baseline. Note that we are not primarily concerned with statistical significance. The range of values for the suspect seen in regular production should generate (close to) the full extent of variation in the output if it is a dominant cause. We first illustrate these ideas using our motivating example and then try to draw general conclusions about how to use the baseline information when verifying a dominant cause.

In the crossbar dimension example, the team concluded that barrel temperature was a suspect based on the results of the input/output investigation. They were confident of the results from the baseline investigation, i.e. they believed the dominant cause acted hour-to-hour, but decided that verification was necessary because it was possible that, in the observational input/output investigation, barrel temperature may have been confounded with the real dominant cause (that was not measured).

To verify barrel temperature as the dominant cause, the team planned a simple two-level experiment. They chose the low and high levels for barrel temperature as 75° and 80° to cover the range of barrel temperatures seen in the input/output investigation. Barrel temperature was difficult to hold fixed in normal production but could be controlled for an experiment. The verification experiment was conducted with only two runs one at each of the selected barrel temperatures. For each run, the barrel temperature was set, 25 parts were made to ensure the temperature had stabilized and the next 10 parts were selected and measured. Then, barrel temperature was changed as quickly as possible for the second run. Using design of experiments terminology the investigation consisted of two runs with 10 repeats per run and no replication.

We see, from the results in Figure 7, that barrel temperature has a large effect on crossbar dimension relative to the baseline variation. The team concluded that they had verified barrel temperature as a dominant cause of crossbar dimension variation. The small number of runs and lack of randomization was not a major concern. Previous investigations had shown that the dominant cause acted in the hour-to-hour family, and thus over the 30 minutes needed to conduct the verification experiment, the team felt it was very unlikely they would have seen the full extent of variation in the crossbar dimension unless barrel temperature was a dominant cause. In other words they concluded there was insufficient time for other causes in the hour-to-hour family to change substantially during the experiment. This suggests barrel temperature could not have been confounded with any other reasonable suspect during the verification experiment.

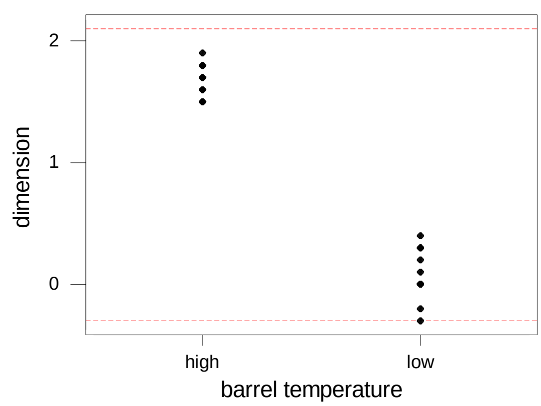

Figure 7: Crossbar Dimension versus Barrel Temperature
dashed horizontal lines show the full extent of variation in the baseline

Assuming the verification experiment can be conducted in a short time, if the dominant cause acts over a long time, as in the crossbar dimension example, we need not worry too much about confounding in the verification experiment. Other causes in the same time family as the suspect will not have time to vary substantially during the verification experiment. As a result, the experimental principles of replication and random assignment are not that critical. On the other hand, if the dominant cause acts over a short time we do need to worry about possible confounding, in the verification experiment, between the suspect and other inputs in the same time family. Then, we need a verification experiment that utilizes sufficient replication (i.e. many runs at each of the two levels of the suspect) and random ordering to try to control the risk of confounding.

Using the Baseline to Help Assess the Feasibility of a Variation Reduction Approach

As suggested in Figure 1, there are seven possible approaches to reducing variation. We can use the time nature of the output variation from the baseline to make an initial assessment of the feasibility of some of the variation reduction approaches. For instance, if the full extent of output variation is seen over a short time, feedback control is not feasible since any observed output values provide only a poor prediction for future values. Also, in any case, there would be little time to adjust the process if needed.

The baseline information is also useful to help plan and analyze subsequent investigations designed to determine if an approach is feasible and/or how to implement a particular approach. The time nature of the output variation seen in the baseline can help define a run. Generally, for experiments conducted to check the feasibility of a variation reduction approach we want each run to resemble a mini baseline investigation, that is, we want it to provide an estimate of the long term behavior of the process with the process changes specified by the factor levels in the run. This suggests, for instance, that if the full extent of output variation is seen over a long time, the robustness approach (as defined in Steiner and MacKay, 2005 – see also upcoming example) is likely not feasible since each run in a robustness experiment would need to be conducted over a long time frame.

In the crossbar dimension example, the team noticed in the right panel of Figure 6, the nonlinear relationship between barrel temperature and crossbar dimension. They decided to raise the barrel temperature set point (average) to make the process less sensitive to variation in barrel temperature, the dominant cause. However, when validating the solution, they discovered that while the variation in the crossbar dimension was reduced substantially, with the new settings there was an increase in the frequency of a mold defect called "burn." They decided to attack the burn defect as a new problem. Using a multivari investigation (see Snee, 2001 for more on multivari charts), they showed 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 due to variation in filling of the mold. Next, since the team felt that the dominant cause would not easily be controlled, they decided to try to make the process robust to the unknown dominant cause(s).

The team planned an experiment with four factors that are normally fixed inputs, that we call candidates. The factors: injection speed, injection pressure, back pressure and screw rpm, were selected because of their influence on fill speed and other potential dominant causes in the part-to-part family. They selected two levels for each factor as given in Table 1. Just for the experiment, the team planned to classify "burn" on each part into one of four categories of increasing severity. Levels 1 and 2 were acceptable, while levels 3 and 4 resulted in scrap. Using a single rater and boundary samples, the team felt this measurement system would add little variation. A full baseline investigation with the new burn classification system was not conducted, but burn levels 1 through 4 had been seen in regular production.

Table 1: Factors and Levels for "Burn" Robustness Experiment
level in current process given by *

Factors	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 selected a fractional factorial experiment with eight runs as given in the Table 2. Since there was no proper baseline investigation for the new burn problem, the team assigned the labels A, B, C and D to the factors so that one of the treatments (Treatment 5) corresponded to the current process settings. In the resolution IV design, pairs of two factor interactions are confounded, as given by the following aliasing structure:

```
Alias Structure
A + BCD
B + ACD
C + ABD
D + ABC
AB + CD
AC + BD
AD + BC
```

The team defined a run as five consecutive parts. Since they knew from the baseline that the family of variation containing the dominant cause (of burn) was part-to-part, they hoped the dominant cause would act within each run. Choosing only five parts for each run was a risk. Having more parts would have made it more likely that each run would reflect the long term behavior of the process, but would have cost more time and money. Each run was carried out once the process stabilized after changing the values of the factors. The order of the runs was randomized. The results from this robustness experiment are given in Table 2.

Table 2: Experimental Plan and Data for "Burn" Robustness Experiment
Treatment #5 uses the current process levels

Treatment	Order	Injection speed	Injection pressure	Back pressure	Screw rpm	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

We plot the burn scores against treatment number in Figure 8. Because the data are discrete, we add jitter in the vertical direction. Examining the results, we see that treatments 2 and 3 are promising and look much better than the existing process performance as given by treatment number 5 and our knowledge that in the existing process we see burn scores covering the full range of values 1-4.

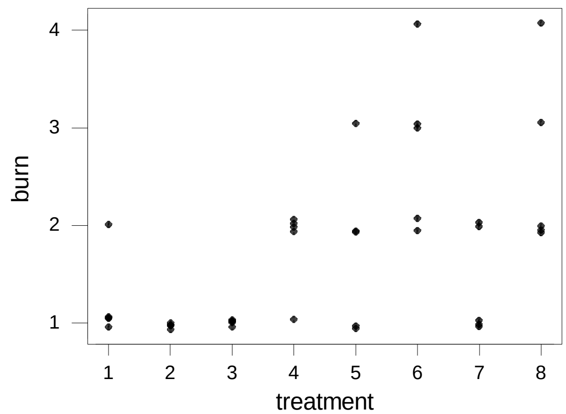

Figure 8: Burn by Treatment With Added Vertical Jitter

The team used the average burn as the performance measure for the formal analysis. We look for process settings that make the performance measure as small as possible. We can think of this as reducing variation in the burn score about the ideal score of zero. Fitting a full model with all possible effects (4 main and 3 two-way interactions) we get the Pareto plot of the effects for the average burn score in Figure 9. We see that only Factor C (back pressure) has a large effect. In drawing this conclusion the team assumed the three input interaction (ABD) aliased with C was negligible. Checking Table 2 we see that low level of back pressure gives less burn on average. The team decided to address the burn defect problem by reducing the back pressure to 75 and leave the other fixed inputs at their original values.

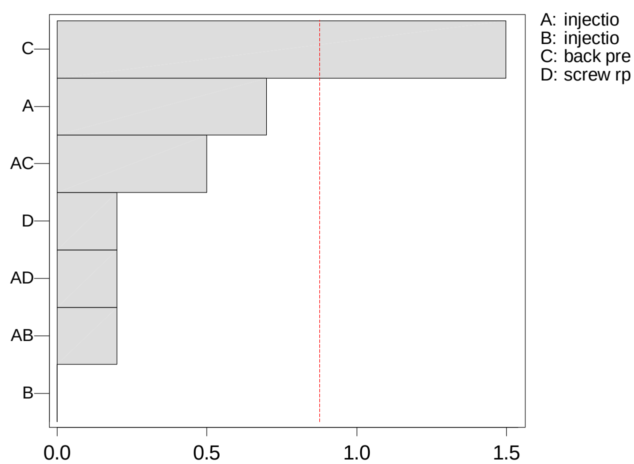

Figure 9: Pareto Plot of Input Effects on Average Burn Score

To finish the project, the team conducted a validation investigation, with the new process settings, by running 300 parts and measuring both the crossbar dimension and the burn defect score. The standard deviation of the crossbar dimension was 0.23 thou and only two parts were scrapped for the burn defect. The team recommended the new settings for the fixed inputs that resulted from investigating the two problems.

Summary and Discussion

In the context of variation reduction for an existing process, we propose a baseline investigation whose results will be valuable to help plan and analyze subsequent process investigations needed to solve the problem, in addition to the usual goal of allowing validation of a solution. In the baseline investigation, we recommend sampling parts from the current process systematically over time and measure the output for each part. From the baseline data we quantify the magnitude of the problem, determine the full extent of variation in the output, and the time nature of the output variation. We summarize the purpose and conduct of our proposed baseline investigation below.

Baseline Investigation Summary

Question

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

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
- determine the time family of the output variation

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. Other inputs should be included if they are available cheaply.
- 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.

Analysis

- Summarize the data using an appropriate sample performance measure(s). For:
  - a continuous output: use an average, standard deviation, histogram and run chart
  - a binary output: use a proportion and 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.
- Determine the time family of output variation

Conclusion

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

The baseline knowledge is helpful when:

- assessing the measurement system,
- searching for and verifying a dominant cause,
- assessing a variation reduction approach (i.e. searching for a solution),
- validating a proposed solution.

The time nature of the output variation is valuable information to help plan subsequent investigations. It can be used to:

- choose an appropriate study population and the time frame,
- generate clues about the dominant cause of variation,
- help define a run and determine the importance of the experimental principles of replication and random assignment in an experimental plan, and
- rule out some variation reduction approaches as not feasible.

The estimated performance measure and full extent of output variation are useful when analyzing the results of any subsequent process investigation. We recommend adding lines showing the full extent of variation in the baseline to all plots that show individual output values. The explicit use of the full extent of output variation (as seen in the baseline) in the analysis of subsequent investigations forces problem solvers to address the important difference between statistical and practical significance. In problem solving, practical significance is what matters. Comparing results to the baseline full extent of variation gives problem solvers a direct way to determine if any observed effects are large relative to the baseline variation. Small effects can be statistical significance while being unimportant. When searching for causes we want to find the dominant cause, i.e. an input that explains a lot of the output variation, not one that is only statistically significant. Of course in many cases a large effect will also be statistically significant, but the opposite is not necessarily true. When we use experiments to look for a solution the issue of practical versus statistical significance is even more critical. We want to find new process settings that are better than the current process rather than better than other treatments used in the experiment. One method to alleviate the concern somewhat is to always include a treatment with the current setting for each of the fixed inputs in the experiment.

Knowing the full extent of variation allows us to directly see:

- whether the dominant cause has acted in an observational investigation, and
- how the process variation compares to the baseline variation in an experimental investigation.

Our proposal for using the information gained in the baseline investigation to help plan and analyze subsequent investigation is one illustration of applying the sequential learning idea. We should ideally use sequential learning throughout problem solving. The results of each investigation provides insight into the process and should be used to help us decide what to do next. Another good example of applying sequential learning, not often employed in practice, is the use of the method of elimination in the search for a dominant cause of output variation (Steiner and MacKay, 2005).

Sequential learning should be used in all problem solving that involves a series of investigation such as for example Six Sigma projects. However, in our experience, Six Sigma books and training material make little connection between the stages of DMAIC. There is no explicit use of information from previous stages to help complete the current stage. For instance, in the well known Six Sigma book by Breyfogle (1999) very few of the examples refer to anything learned in a previous stage of DMAIC. This is especially strange when moving from the Analysis to the Improvement stage; you would think that knowing the cause would be very helpful when looking for a solution.

References

Automotive Industry Action Group (AIAG) (2003), *Measurement Systems Analysis*, 3rd edition. Southfield, MI.

Box G.E.P (1999), "Statistics as a Catalyst to Learning by Scientific Method Part II – A Discussion," *Journal of Quality Technology*, 31, 16-29.

Breyfogle F.W. III (1999), *Implementing Six Sigma: Smarter Solutions Using Statistical Methods*, John Wiley and Sons, New York.

Browne R., MacKay, R.J., Steiner S.H. "Leveraged Measurement System Assessment," accepted by the *Journal of Quality Technology*, Feb. 2009a.

Browne R., MacKay, R.J., Steiner S.H. "Two Stage Leveraged Measurement System Assessment," accepted by *Technometrics*, Jan. 2009b.

De Mast J., Roes K.C.B. and Does R.J.M.M. (2001), "The Multi-Vari Chart: A Systematic Approach," *Quality Engineering*, 13, 437-448.

Juran J.M. and Gryna F.M. (1980), *Quality Planning and Analysis*, 2nd edition, McGraw-Hill, New York.

Shainin R.D. (1993), "Strategies for Technical Problem Solving," *Quality Engineering*, 5(3), 433-448.

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.

Steiner S.H. and MacKay R.J. (2005), Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes, ASQ Quality Press, Milwaukee WI.