QUALITY AND PRODUCTIVITY IMPROVEMENT USING REGRESSION ANALYSIS: A CASE STUDY

J. Clifton Young

IIQP Research Report
RR-94-09

July 1994

QUALITY AND PRODUCTIVITY IMPROVEMENT USING REGRESSION ANALYSIS: A CASE STUDY

J. Clifton Young
Institute for Improvement in Quality and Productivity
University of Waterloo

ABSTRACT

Although there is an enormous amount of data being collected in industry today, the unfortunate fact is that little of it is ever analyzed. One effective method of looking at such data is regression analysis. In this example an initial regression analysis revealed a very critical clue as to how to achieve major gains in both quality and productivity. Multiple linear regression of subsequent trial data produced strong evidence of major improvements.

Key Words: Regression analysis; Indicator variables; Productivity improvement

Introduction

The following study was undertaken to increase productivity in a bar mill in a steel plant. A bar mill consists of a series of "passes" or stations at each of which the size of an incoming slab of steel is reduced until the desired final thickness is achieved. The passes wear over time with the result that the final product has poor appearance or is out of specification limits. When this occurs the pass is changed. Changeovers result in lost time and substantial replacement costs.

For confidentiality reasons, the data have been coded and presentation of the results delayed; thus the magnitude of the results is not representative of the steel industry as it exists today. The magnitude of the improvement is accurate.

The example shows how regression analysis can be used to draw conclusions from complex data sets and more specifically, how to use indicator or "dummy" variables to model qualitative factors in a process.

The study also illustrates a situation in which a careful search for an "assignable cause" for outliers resulted in major gains in *both* quality and productivity.

Analysis of the Original Data Set

The initial clue regarding the possibility of a significant improvement in productivity and quality came from an analysis of the relationship between the weight (w_i) of steel with a final bar size or thickness (t_i) that was processed before one or more passes had to be replaced. It should be stated that once the mill is set up to run a specific size, only that size is run until it is necessary to shut the line down to change a pass.

The initial data set, which was collected over an eight month period, is plotted in Figure 1. Since a smaller final size requires more effort on the part of the passes, it was expected that the greater the final thickness the greater the amount of steel processed before a changeover was required. That this is true can be seen in the scatter diagram. The exact nature of the relationship was determined using linear regression to fit a model to the data. The resulting model was

$$w = -134.03 + 1.225\ t.$$

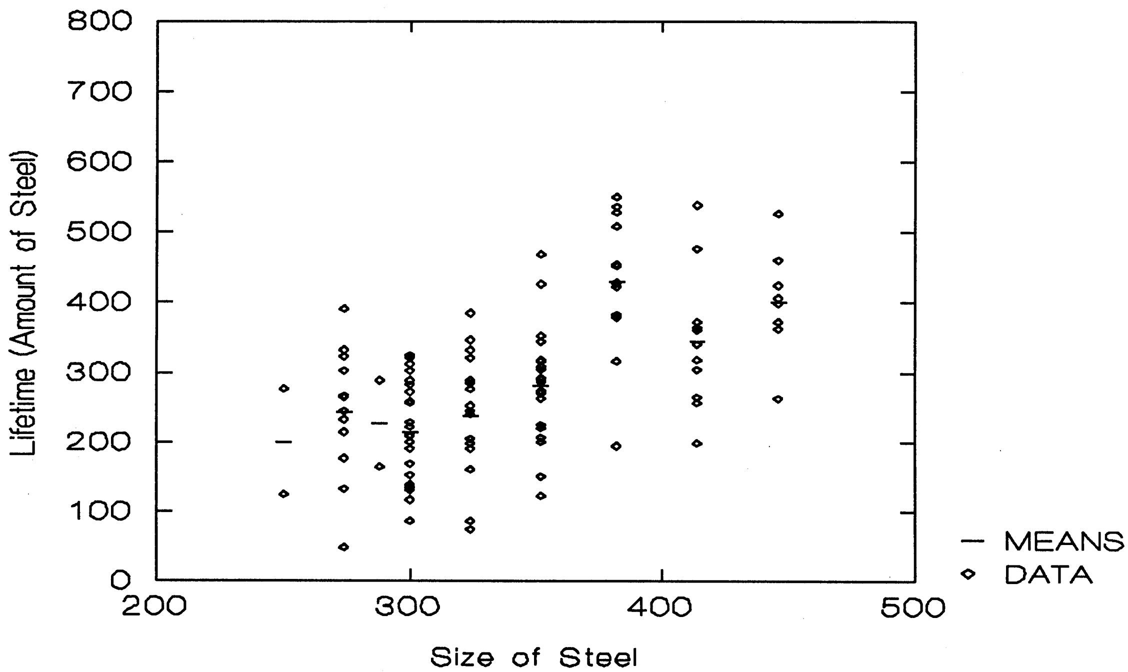


Figure 1: Scatter diagram showing the relationship between the amount of steel processed (lifetime) and steel size under the *original conditions*

Since there are several independent runs at each of the nine sizes, it was possible to check the adequacy of the linear model by means of a standard "lack of fit" test (Draper and Smith, 1981, Section 1.5). The results of this test are given in the analysis of variance

(ANOVA) presented in Table 1.

Table 1: Analysis of variance table showing lack of fit test for the data collected under the original operating conditions.

Source	**DF**	**SS**	**MS**	**F**	**P**
Linear term	1	439546.66	439546.66	61.79	<< .0001
Lack of Fit	7	168701.48	24100.21	3.39	.0027
Pure Error	103	732730.72	7113.89		
Corrected Total	111	1340978.86			

Not surprisingly, when compared to the pure error, there is very strong evidence of a relationship between weight and thickness (a significance level **$p < .0001$**).

The engineers involved were, however, surprised to see the strong evidence of a lack of fit of the linear model (**$p = .0027$**).

In order to determine the nature of the lack of fit, the mean values of weight were calculated for each of the nine thicknesses, and included in Figure 1. It can be seen that, although there is a significant lack of fit, there is no hint of a smooth curve in the mean values. The problem seems to be that the mean amounts processed at **$t = 382$** and possibly **$t = 274$** appear to be well above the line described by the rest of the data. The means at **$t = 288$** and **$t = 250$** are also off the line but since they are means of only two runs each this is much less surprising than the other two which are both means of 12 independent runs. The remaining five means (all from large samples) lie on an almost perfect straight line.

Dropping the data sets at **$t = 382$** and **$t = 274$** results in the model

$$w = -139.13 + 1.185t$$

and the ANOVA in Table 2.

Table 2: ANOVA table showing a test of lack of fit after dropping two outlying data sets. (*t = 382 and t = 274*)

Source	**DF**	**SS**	**MS**	**F**	**P**
Linear term	1	306763.37	306763.37	47.73	<< .0001
Lack of Fit	5	7711.22	1542.24	.24	> .5
Pure Error	81	520573.72	6426.84		
Corrected Total	87	835048.31			

As can be seen, there is no longer any evidence whatsoever of a lack of fit (**p > 0.5**). There is thus very strong evidence that the data at the removed points do not fit the model that fits the rest of the data quite well. What caught the eye of the engineers involved was the fact that the amounts of steel processed at these points was *significantly higher* than expected according to the model that fits the rest of the data. This corresponds to a significantly longer time between replacement of the passes and hence considerable savings.

When the engineers investigated, they found that the steel rolled at these two sizes was for one specific customer who demanded a very high level of finished product quality. In fact, operating personnel disliked running product for that customer because extra changeover time was required to ensure the required level of finished product quality. Everyone "knew" that this would result in fewer tons out the door and, since productivity was measured in tons out the door, reduced productivity. Although the finished product quality level was higher, it was felt to be an unreasonable requirement since no one else asked for it.

What was not realized until this analysis was the fact that the passes lasted longer when special attention was given to the changeover. Thus, if the results of the above preliminary analysis are valid, the special set-up would not only result in better quality but also in increased productivity. In order to test this possibility, the special set-up was tentatively adopted for all sizes of steel.

Analysis of Pass Life After Modifications

The line was then run and data collected using the modified change-over procedure. After about four months further refinements were made and additional data collected for another four months.

The following is a quote of the questions asked by the engineers when they supplied the data from process after the modification (m) and the subsequent refinement (r).

"We would like to determine if the modifications improved the pass life and by how much?

- three sets of data from the same source
- one thing changed from each set (same thing)
- nine groups (bar thickness in thousands of an inch) within each set
- large variations within each group

Questions

- the best method of analyzing the data
- Do I compare sets?
- Do I compare groups within the sets?

- Can I compare three sets or groups simultaneously?
- Which data (if any) can be rejected statistically?
- The best way of reporting the data, simultaneously if possible?
- To compare groups or sets do I need more data? If so, in which groups or sets?"

The three sets of data referred to are the data from the original (o), the modified (m) and the refined (r) process. The data are plotted in Figure 2 using the symbols **o, r and m** for the corresponding conditions. As can be seen, the data sets overlap considerably and it is not at all clear which conditions are best. Figures 3a and 3b show the relationship separately for modified and refined data and include the group means.

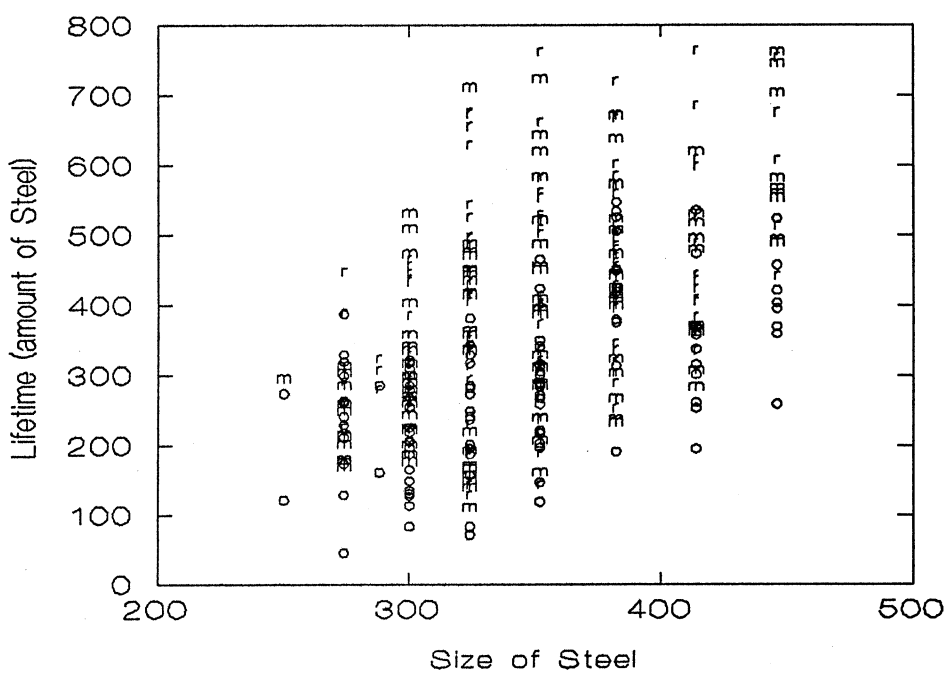


Figure 2: Scatter diagram showing the relationship between the amount of steel processed and the size of steel for all three sets of data. The symbols **o**, **m** and **r** are used for the corresponding conditions.

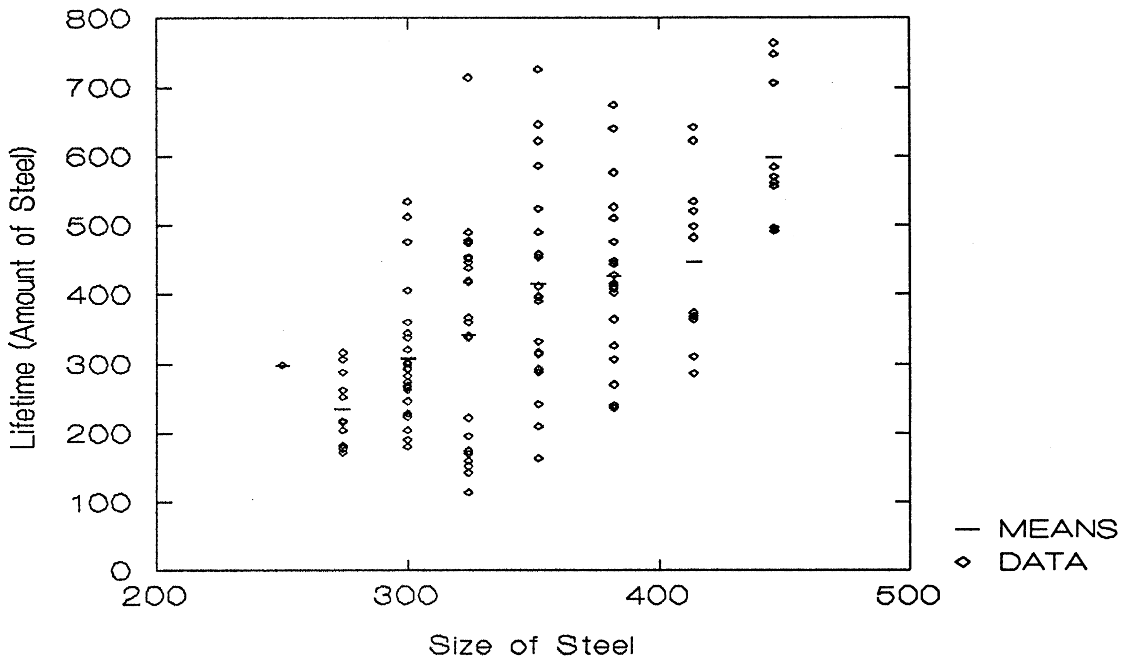


Figure 3a: Scatter diagram showing the relationship between amount and size for the *modified conditions*.

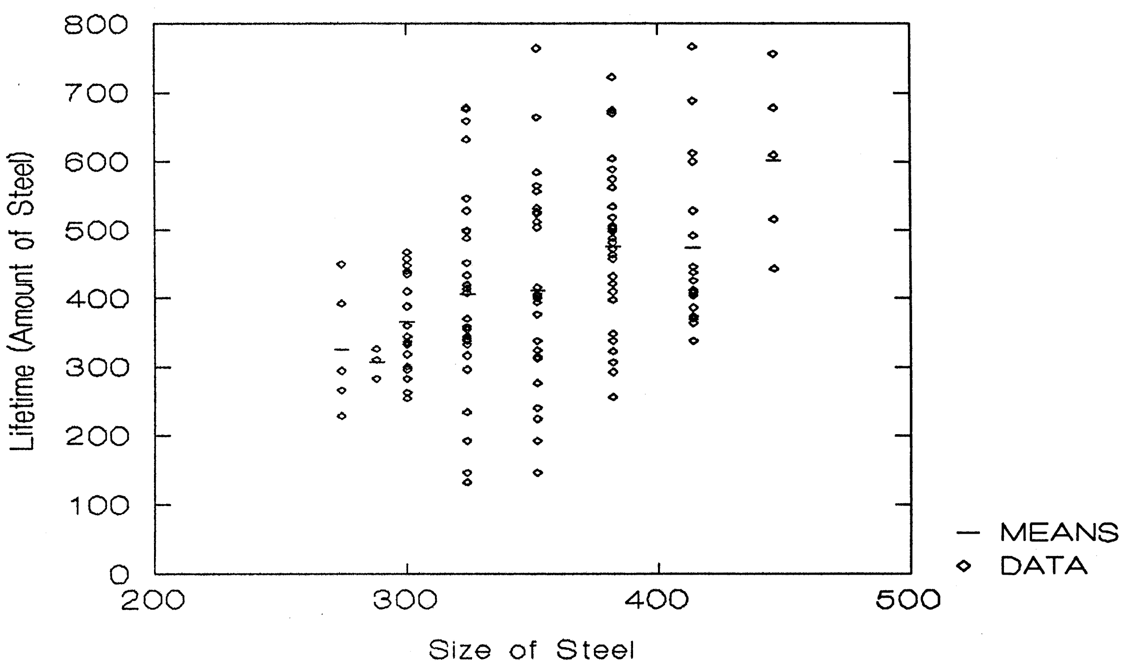


Figure 3b: Scatter diagram showing the relationship between amount and size for the *refined conditions*.

In order to answer the engineer's questions, the following model was set up for analysis using multiple regression with two indicator ("dummy") variables (Draper and Smith, 1981, Section 5.4) to account for the three different methods of running the process. The model used was

$$w_i = b_0 + b_1 M_i + b_2 R_i + b_3 t_i + b_4 M_i t_i + b_5 R_i t_i + e_i \,, \qquad [1]$$

where

w_i is the i[th] observation of weight (amount)

t_i is the thickness (size) corresponding to w_i

M_i is an indicator variable which equals 1 for those observations collected under both modified and refined conditions and 0 otherwise

R_i is an indicator variable which equals 1 for observations collected after the process was refined

$M_i t_i$ is the product of M_i and t_i

$R_i t_i$ is the product of R_i and t_i .

Since the refinement involved a relatively minor revision to the modified process, it was expected to be either similar to the modification or, hopefully, slightly better. In order to reflect this fact, the model was set up such that the indicator M_i treats data from *both* the modified and refined series the same ($M_i = 1$ for both and 0 otherwise) while R_i differentiates between the two ($R_i = 1$ for the refinement and 0, otherwise). Thus the significance of terms involving M_i will indicate that, on average, both the modified and refined conditions differ from the original; the significance of terms involving R_i will indicate a difference between the refinement and the modification.

A more typical approach would be to have $M_i = 1$ for the modified process and zero otherwise and $R_i = 1$ for the refined process and zero otherwise. This approach would work but the testing of the specific hypothesis of interest would be somewhat more difficult so the approach outlined above was used.

Further insight into the interpretation of the model and the corresponding significance tests may be gained by recognition of the fact that, in effect, the equation is that of a straight line with

$$intercept = b_0 + b_1 M_i + b_2 R_i \ . \qquad [2]$$

If neither b_1 nor b_2 are significant that would mean that the same intercept (b_0) adequately models all three process conditions; the significance of b_1 but not b_2 would mean that b_0, the intercept for the original conditions must be modified by an amount b_1 (*to* $b_0 + b_1$) for both the modified and refined conditions ($M_i = 1$). Finally, if b_1 and b_2 are both significant, then three separate and significantly different intercepts are needed (b_0 for original conditions, $b_1 + b_2$ for modified, and $b_0 + b_1 + b_2$ for refined). Since the refinement involves a minor change to the modification, it would be very suprising to find b_2 significant and b_1 not significant.

In a similar manner the slope of the line is given by

$$slope = b_3 + b_4 M_i + b_5 R_i \ . \qquad [3]$$

The same similar argument applies to the slope adjustment terms b_4 and b_5. In effect the model [1] simultaneously fits the three separate lines,

$$w_i = b_0 + b_1 t_i \qquad \text{(original conditions)}$$

$$w_i = (b_0 + b_1) + (b_3 + b_4)t_i \qquad \text{(modified conditions)}$$

and

$$w_i = (b_0 + b_1 + b_2) + (b_3 + b_4 + b_5)t_i \qquad \text{(refined conditions)}$$

simultaneously. Tests for the significance of b_2, b_3, b_4 and b_5 enable the assessment of the significance of differences between the lines.

An analysis of variance table and other details of an analysis of the above model are given in Table 3. The format of the table is typical of the output from any standard multiple regression computer package.

Table 3: ANOVA table for model [1] with tests for individual parameters

Source	**DF**	**SS**	**MS**	**F**	**P**
Model	5	2958369.1938	591673.83877	43.645	<<.0001
Error	337	4568558.293	13556.552798		
Corrected Total	342	7526927.4869			

Root MSE	116.43261	R-square	0.3930
Dep Mean	370.47230	Adj R-sq	0.3840
C.V.	31.42816		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H_0: Parameter=0	Prob > |T|
Intercept	1	-139.127	86.523	-1.608	0.1088
Size	1	1.185	0.249	4.757	0.0001
M	1	-72.052	113.357	-0.636	0.5255
R	1	203.319	110.510	1.840	0.0667
M × Size	1	0.527	0.325	1.623	0.1055
R × Size	1	-0.476	0.313	-1.522	0.1288

The results in the ANOVA table show overwhelming evidence ($p < .0001$) that the model is statistically significant. However, the results in the lower "Parameter Estimates" portion of the table show only the slope (b_1) of size as being strongly significant ($p < .0001$) with R marginally significant. As well, the intercept, M × Size and R × Size might also be regarded as marginally significant (p around .10). It is, however, important to take great care in interpreting these tests. In each of these t-tests the significance of the contribution of the parameter is tested when it is *added* to a model *containing all the other parameters.* Since, with this model, R and R × Size and M and M × Size are very highly correlated ($r = .97$ and .99 respectively), this means that once M (or R) is in the model M × Size (or R × Size) is unlikely to help unless there is a dramatic change in the slope.

The marginal significance of the intercept is of little interest, especially as there is no data anywhere near the origin and there is no reason to assume linearity outside the range of the data. In any case, there is no reason to expect a zero intercept.

In view of the above results, a C_p analysis of all possible models was performed (see Draper and Smith, Chapter 6). The C_p statistic is calculated in such a way that, if no important variables have been left out of the model C_p should be approximately equal to p, where p is the total number of parameters in the model (including the constant term b_0). There is no statistical difference between models with $C_p \approx p$. The use of C_p enables an analyst to see if there are several equally good models. The output from a C_p analysis of these data is provided in Table 4. Note that the basic variable size (*T*) was included in all models.

A study of the information in this table shows that a model with T, R and MT has $C_p \approx p$ ($p = 4$, $C_p = 4.3$). Models with T, MT and RT or with T, M and R have slightly larger C_p values (5.4 and 5.5 with $p=4$) but are also worth considering. In particular, the model containing T, M and R has the advantage of simplicity in that is corresponds to three parallel lines. The adequacy of this model can be tested by finding the difference in the residual sums

Table 4: The results of a C_p analysis of the model.
All models studied contain the linear term in size (T).

Number in Model (p)	C_p	Residual Sum of Squares	Variables in Model
2	88.3	5793922.5	T
Note: The above variable is included in all models to follow.			
3	8.5	4685045.6	MT
3	9.8	4702539.6	M
3	49.4	5239059.8	R
3	51.8	5271278.7	RT
4	4.3	4599982.8	R MT
4	5.4	4614830.4	MT RT
4	5.5	4615622.0	M R
4	6.2	4626287.7	M RT
4	10.5	4684814.2	M MT
4	50.3	5224468.9	R RT
5	4.4	4574035.4	R MT RT
5	6.3	4599979.8	M R MT
5	6.6	4604263.2	M R RT
5	7.3	4614446.3	M MT RT
6	6.0	4568558.3	M R MT RT

of squares between it and the model with all the parameters in it. This is given by the additional sum of squares

$$SS(MT, RT/Constant, T, M,R) = 4615622.0 - 4568558.3 = 47063.7 .$$

This sum of squares has two degrees of freedom.

The significance of the contribution of the two additional terms is given by

$$p = Prob\{F(2,337) > \frac{47063.7/2}{13556.6}\}$$

$$= Prob\{F(2,327) > 1.74\}$$

$$\approx 0.18 .$$

Thus, once the intercept adjustment terms are in the model, the slope adjustment terms are not significant. The resulting model is

$$w = -217.13 + 1.41t + 110.28M + 36.98R . \qquad [6]$$

This means that relative to the original operating conditions, on average, one can expect that about 110 more units steel can be processed using the modified conditions and an additional 37 units using the refined conditions. The differences between the conditions are very significant from a statistical point of view. The results of an analysis of this model are given in Table 5. A test of lack of fit of the model is included. It is not remotely significant.

Realistically, it's very unlikely that the lines are parallel, or even perfectly linear, the table below does, however, model the data adequately from a statistical point of view and, most importantly, it is relatively easy to explain to non-statisticians. As well, it clearly shows the direction, magnitude and significance of the improvement due to the changes.

Although there is very strong evidence to support the conclusions, it will be noticed that $R^2 = .387$. This would be judged to be very small by many analysts. In fact, R^2 can be very misleading (see Weisburg, 1985, Example 3.4). A more meaningful statistic is the Root MSE. In this case, it is 116.7. This says (approximately) that individual data points (amounts of steel processed) are distributed around the predicted values with a standard deviation $s=116.7$. This value is very large considering the magnitude of the numbers. There is a need

Table 5: ANOVA table for the final model with a test of lack of fit.

Source	DF	SS	MS	F	P
Model	3	2911305.4418	970435.14726	70.17	<< .0001
Lack of Fit	19	189864.37	9992.86	.72	> .50
Pure Error	320	4425757.68	13830.49		
Corrected Total	342	7526927.4869			

Root MSE	116.68506	R-square	0.3868
Dep Mean	370.47230	Adj R-sq	0.3814
C.V.	31.49630		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H_0: Parameter= 0	Prob > |T|
Intercept	1	-217.130	47.146	-4.605	0.0001
Size	1	1.412	0.132	10.672	0.0001
M	1	110.279	16.297	6.767	0.0001
R	1	36.983	14.637	2.527	0.0120

for further study of why the amounts of steel processed vary so much for a given size (from146 to 764 units at $t = 352$ for the refined conditions). Note that the assumption of normality of the data around the line is supported by normal plots of the residuals.

It was also found that s was lower for the original conditions ($s = 78.4$) than for the modified ($s = 128.5$) and the refined ($s = 125.4$) conditions. The overall Root MSE is essentially a weighted average of these values. Further efforts should be made to find out why the variation in the amount of steel processed has increased. This does not, of course, detract from the substantial average increases for modified over the original conditions and for refined

Table 5: ANOVA table for the final model with a test of lack of fit.

Source	**DF**	**SS**	**MS**	**F**	**P**
Model	3	2911305.4418	970435.14726	70.17	<< .0001
Lack of Fit	19	189864.37	9992.86	.72	> .50
Pure Error	320	4425757.68	13830.49		
Corrected Total	342	7526927.4869			

Root MSE	116.68506	R-square	0.3868
Dep Mean	370.47230	Adj R-sq	0.3814
C.V.	31.49630		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H_0: Parameter= 0	Prob > |T|
Intercept	1	-217.130	47.146	-4.605	0.0001
Size	1	1.412	0.132	10.672	0.0001
M	1	110.279	16.297	6.767	0.0001
R	1	36.983	14.637	2.527	0.0120

for further study of why the amounts of steel processed vary so much for a given size (from146 to 764 units at $t = 352$ for the refined conditions). Note that the assumption of normality of the data around the line is supported by normal plots of the residuals.

It was also found that s was lower for the original conditions ($s = 78.4$) than for the modified ($s = 128.5$) and the refined ($s = 125.4$) conditions. The overall Root MSE is essentially a weighted average of these values. Further efforts should be made to find out why the variation in the amount of steel processed has increased. This does not, of course, detract from the substantial average increases for modified over the original conditions and for refined

Residual plots indicated that there might be a change in slope for the modified conditions. This was found to be due entirely to the data for size = 484. Removal of this set of data lead to a model

$$w = -176.00 + 1.29\, t + 103.03M + 45.21R \,. \qquad [7]$$

which differs slightly from the previous model [6]. There is, however, essentially no difference in the magnitude of the improvement of moving from original to modified conditions [+103 vs. +110] and from modified to refined [+45 vs. +37]. No reason for the slightly unusual behaviour of this set of runs could be found.

Conclusion

Very substantial savings resulted from moving to the refined changeover procedure and mill operating conditions. This resulted from finding the "assignable cause" for the outliers in the original data set. Although careful study of the appropriate scatter diagrams might lead to this result, the statistical analysis was able to estimate the magnitude of the improvement and the strength of the evidence supporting the conclusions. Although the initial goal was productivity improvement, the result was a substantial improvement in *both* quality and productivity.

References

Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis*. John Wiley & Sons, New York

Weisburg, S. (1985). *Applied Linear Regression.* John Wiley & Sons, New York.

QUALITY AND PRODUCTIVITY IMPROVEMENT USING REGRESSION ANALYSIS: A CASE STUDY

J. Clifton Young
Institute for Improvement in Quality and Productivity
University of Waterloo

ABSTRACT

Although there is an enormous amount of data being collected in industry today, the unfortunate fact is that little of it is ever analyzed. One effective method of looking at such data is regression analysis. In this example an initial regression analysis revealed a very critical clue as to how to achieve major gains in both quality and productivity. Multiple linear regression of subsequent trial data produced strong evidence of major improvements.

Key Words: Regression analysis; Indicator variables; Productivity improvement

Quality Engineering version
- for publication

RR-94-09.

About the Author:

J. Clifton Young was the founding Director of the Institute for Improvement in Quality and Productivity at the University of Waterloo. He is an associate professor in the Department of Statistics and Actuarial Science. Clif received a B.A.Sc. in Electrical Engineering from the University of Toronto and a Ph.D. in Statistics from the University of Edinburgh. He continues to consult on the use of statistical methods for process improvement with a wide range of industries.

Introduction

The following study was undertaken to increase productivity in a bar mill in a steel plant. A bar mill consists of a series of "passes" or stations at each of which the size of an incoming slab of steel is reduced until the desired final thickness is achieved. The passes wear over time with the result that the final product has poor appearance or is out of specification limits. When this occurs the pass is changed. Changeovers result in lost time and substantial replacement costs.

For confidentiality reasons, the data have been coded and presentation of the results delayed; thus the magnitude of the results is not representative of the steel industry as it exists today. The magnitude of the improvement is accurate.

The example shows how regression analysis can be used to draw conclusions from complex data sets and more specifically, how to use indicator or "dummy" variables to model qualitative factors in a process.

The study also illustrates a situation in which a careful search for an "assignable cause" for outliers resulted in major gains in *both* quality and productivity.

Analysis of the Original Data Set

The initial clue regarding the possibility of a significant improvement in productivity and quality came from an analysis of the relationship between the weight (w_i) of steel with a final bar size or thickness (t_i) that was processed before one or more passes had to be replaced. It should be stated that once the mill is set up to run a specific size, only that size is run until it is necessary to shut the line down to change a pass.

The initial data set, which was collected over an eight month period, is plotted in Figure 1. Since a smaller final size requires more effort on the part of the passes, it was expected that the greater the final thickness the greater the amount of steel processed before a changeover was required. That this is true can be seen in the scatter diagram. The exact nature of the relationship was determined using linear regression to fit a model to the data. The resulting model was

$$w = -134.03 + 1.225\ t.$$

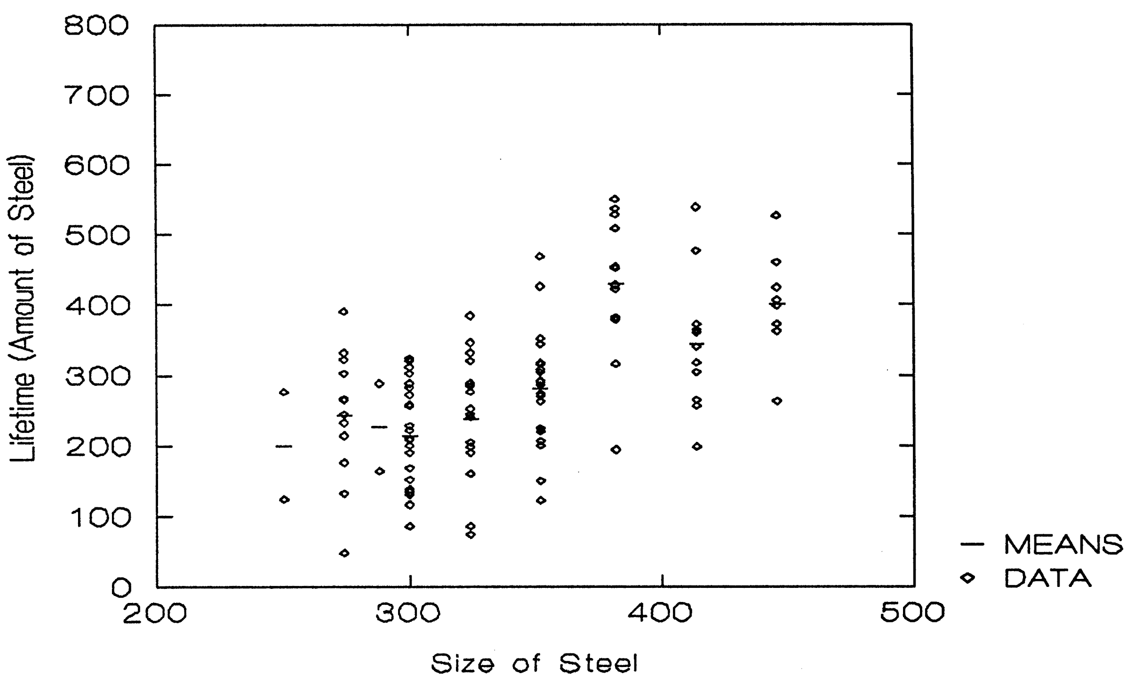


Figure 1: Scatterdiagram showing the relationship between the amount of stee processed (lifetime) and steel size under the *original conditions*

Since there are several independent runs at each of the nine sizes, it was possible to check the adequacy of the linear model by means of a standard "lack of fit" test (Draper and

Smith, 1981, Section 1.5). The results of this test are given in the analysis of variance (ANOVA) presented in Table 1.

Table 1: Analysis of variance table showing lack of fit test for the data collected under the original operating conditions.

Source	**DF**	**SS**	**MS**	**F**	**P**
Linear term	1	439546.66	439546.66	61.79	<< .0001
Lack of Fit	7	168701.48	24100.21	3.39	.0027
Pure Error	103	732730.72	7113.89		
Corrected Total	111	1340978.86			

Not surprisingly, when compared to the pure error, there is very strong evidence of a relationship between weight and thickness (a significance level **p < .0001**).

The engineers involved were, however, surprised to see the strong evidence of a lack of fit of the linear model (**p = .0027**).

In order to determine the nature of the lack of fit, the mean values of weight were calculated for each of the nine thicknesses, and included in Figure 1. It can be seen that, although there is a significant lack of fit, there is no hint of a smooth curve in the mean values. The problem seems to be that the mean amounts processed at **t = 382** and possibly **t = 274** appear to be well above the line described by the rest of the data. The means at **t = 288** and **t = 250** are also off the line but since they are means of only two runs each this is much less surprising than the other two which are both means of 12 independent runs. The remaining five means (all from large samples) lie on an almost perfect straight line.

Dropping the data sets at **t = 382** and **t = 274** results in the model

$$w = -139.13 + 1.185t$$

and the ANOVA in Table 2.

Table 2: ANOVA table showing a test of lack of fit after dropping two outlying data sets. (*t = 382 and t = 274*)

Source	DF	SS	MS	F	P
Linear term	1	306763.37	306763.37	47.73	<< .0001
Lack of Fit	5	7711.22	1542.24	.24	> .5
Pure Error	81	520573.72	6426.84		
Corrected Total	87	835048.31			

As can be seen, there is no longer any evidence whatsoever of a lack of fit (**p > 0.5**). There is thus very strong evidence that the data at the removed points do not fit the model that fits the rest of the data quite well. What caught the eye of the engineers involved was the fact that the amounts of steel processed at these points was *significantly higher* than expected according to the model that fits the rest of the data. This corresponds to a significantly longer time between replacement of the passes and hence considerable savings.

When the engineers investigated, they found that the steel rolled at these two sizes was for one specific customer who demanded a very high level of finished product quality. In fact, operating personnel disliked running product for that customer because extra changeover time was required to ensure the required level of finished product quality. Everyone "knew" that this would result in fewer tons out the door and, since productivity was measured in tons out the door, reduced productivity. Although the finished product quality level was higher, it was felt to be an unreasonable requirement since no one else asked for it.

What was not realized until this analysis was the fact that the passes lasted longer when special attention was given to the changeover. Thus, if the results of the above preliminary analysis are valid, the special set-up would not only result in better quality but also in increased productivity. In order to test this possibility, the special set-up was tentatively adopted for all sizes of steel.

Analysis of Pass Life After Modifications

The line was then run and data collected using the modified change-over procedure. After about four months further refinements were made and additional data collected for another four months.

The following is a quote of the questions asked by the engineers when they supplied the data from process after the modification (m) and the subsequent refinement (r).

"We would like to determine if the modifications improved the pass life and by how much?

- three sets of data from the same source
- one thing changed from each set (same thing)
- nine groups (bar thickness in thousands of an inch) within each set
- large variations within each group

Questions

- the best method of analyzing the data
- Do I compare sets?

- Do I compare groups within the sets?
- Can I compare three sets or groups simultaneously?
- Which data (if any) can be rejected statistically?
- The best way of reporting the data, simultaneously if possible?
- To compare groups or sets do I need more data? If so, in which groups or sets?"

The three sets of data referred to are the data from the original (o), the modified (m) and the refined (r) process. The data are plotted in Figure 2 using the symbols **o, r and m** for the corresponding conditions. As can be seen, the data sets overlap considerably and it is not at all clear which conditions are best. Figures 3a and 3b show the relationship separately for modified and refined data and include the group means.

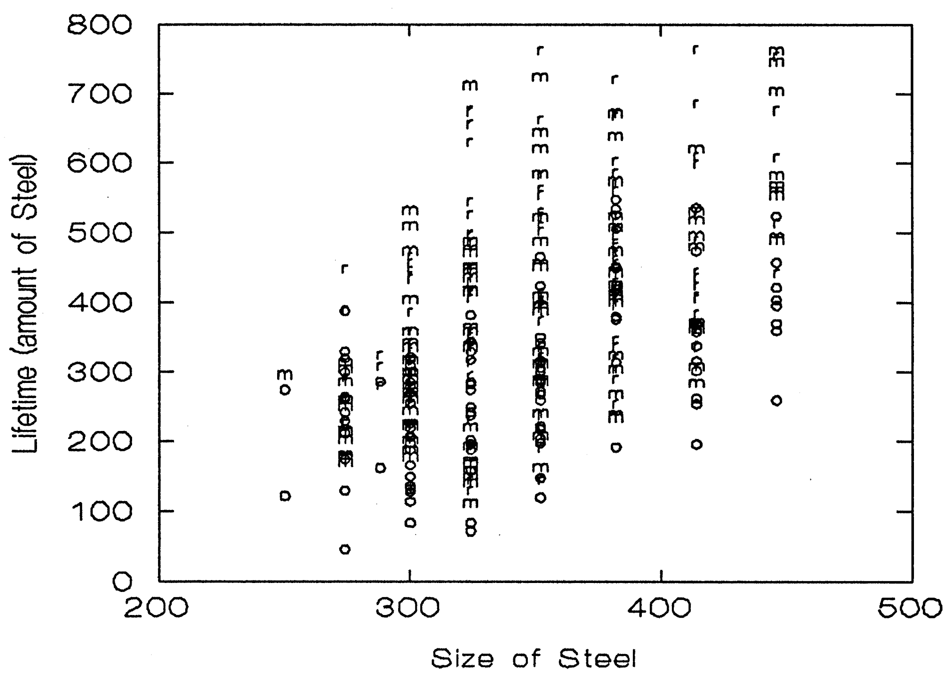


Figure 2: Scatter diagram showing the relationship between the amount of steel processed and the size of steel for all three sets of data. The symbols **o, m** and **r** are used for the corresponding conditions.

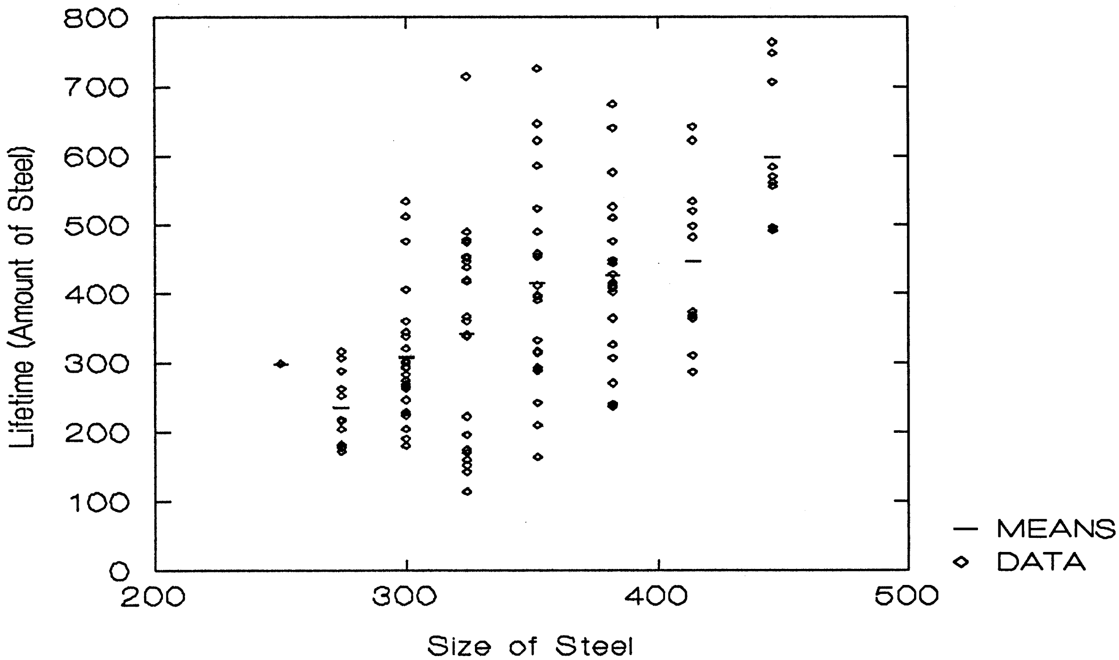


Figure 3a: Scatter diagram showing the relationship between amount and size for the *modified conditions*.

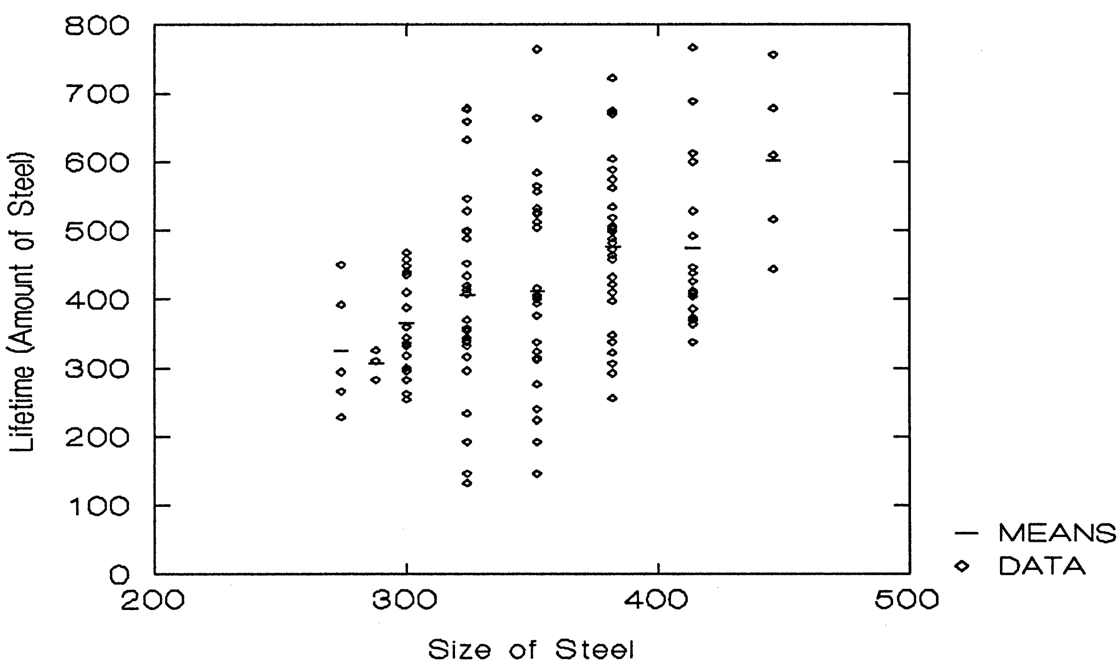


Figure 3b: Scatter diagram showing the relationship between amount and size for the *refined conditions*.

In order to answer the engineer's questions, the following model was set up for analysis using multiple regression with two indicator ("dummy") variables (Draper and Smith, 1981, Section 5.4) to account for the three different methods of running the process. The model used was

$$w_i = b_0 + b_1 M_i + b_2 R_i + b_3 t_i + b_4 M_i t_i + b_5 R_i t_i + e_i \, , \qquad [1]$$

where

w_i is the i^{th} observation of weight (amount)

t_i is the thickness (size) corresponding to w_i

M_i is an indicator variable which equals 1 for those observations collected under both modified and refined conditions and 0 otherwise

R_i is an indicator variable which equals 1 for observations collected after the process was refined

$M_i t_i$ is the product of M_i and t_i

$R_i t_i$ is the product of R_i and t_i .

Since the refinement involved a relatively minor revision to the modified process, it was expected to be either similar to the modification or, hopefully, slightly better. In order to reflect this fact, the model was set up such that the indicator M_i treats data from *both* the modified and refined series the same ($M_i = 1$ for both and 0 otherwise) while R_i differentiates between the two ($R_i = 1$ for the refinement and 0, otherwise). Thus the significance of terms involving M_i will indicate that, on average, both the modified and refined conditions differ from the original; the significance of terms involving R_i will indicate a difference between the refinement and the modification.

A more typical approach would be to have $M_i = 1$ for the modified process and zero otherwise and $R_i = 1$ for the refined process and zero otherwise. This approach would work but the testing of the specific hypothesis of interest would be somewhat more difficult so the approach outlined above was used.

Further insight into the interpretation of the model and the corresponding significance tests may be gained by recognition of the fact that, in effect, the equation is that of a straight line with

$$intercept = b_0 + b_1 M_i + b_2 R_i \, . \qquad [2]$$

If neither b_1 nor b_2 are significant that would mean that the same intercept (b_0) adequately models all three process conditions; the significance of b_1 but not b_2 would mean that b_0, the intercept for the original conditions must be modified by an amount b_1 (*to* $b_0 + b_1$) for both the modified and refined conditions ($M_i = 1$). Finally, if b_1 and b_2 are both significant, then three separate and significantly different intercepts are needed (b_0 for original conditions, $b_1 + b_2$ for modified, and $b_0 + b_1 + b_2$ for refined). Since the refinement involves a minor change to the modification, it would be very suprising to find b_2 significant and b_1 not significant.

In a similar manner the slope of the line is given by

$$slope = b_3 + b_4 M_i + b_5 R_i \, . \qquad [3]$$

The same similar argument applies to the slope adjustment terms b_4 and b_5. In effect the model [1] simultaneously fits the three separate lines,

$$w_i = b_0 + b_1 t_i \qquad \text{(original conditions)}$$

$$w_i = (b_0 + b_1) + (b_3 + b_4)t_i \qquad \text{(modified conditions)}$$

and

$$w_i = (b_0 + b_1 + b_2) + (b_3 + b_4 + b_5)t_i \qquad \text{(refined conditions)}$$

simultaneously. Tests for the significance of b_2, b_3, b_4 and b_5 enable the assessment of the significance of differences between the lines.

An analysis of variance table and other details of an analysis of the above model are given in Table 3. The format of the table is typical of the output from any standard multiple regression computer package.

Table 3: ANOVA table for model [1] with tests for individual parameters

Source	**DF**	**SS**	**MS**	**F**	**P**
Model	5	2958369.1938	591673.83877	43.645	<<.0001
Error	337	4568558.293	13556.552798		
Corrected Total	342	7526927.4869			

Root MSE	116.43261	R-square	0.3930
Dep Mean	370.47230	Adj R-sq	0.3840
C.V.	31.42816		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H_0: Parameter=0	Prob > |T|
Intercept	1	-139.127	86.523	-1.608	0.1088
Size	1	1.185	0.249	4.757	0.0001
M	1	-72.052	113.357	-0.636	0.5255
R	1	203.319	110.510	1.840	0.0667
M × Size	1	0.527	0.325	1.623	0.1055
R × Size	1	-0.476	0.313	-1.522	0.1288

The results in the ANOVA table show overwhelming evidence ($p < .0001$) that the model is statistically significant. However, the results in the lower "Parameter Estimates" portion of the table show only the slope *(b_1)* of size as being strongly significant ($p < .0001$) with R marginally significant. As well, the intercept, M × Size and R × Size might also be regarded as marginally significant (p around .10). It is, however, important to take great care in interpreting these tests. In each of these t-tests the significance of the contribution of the parameter is tested when it is *added* to a model *containing all the other parameters*. Since, with this model, R and R × Size and M and M × Size are very highly correlated ($r = .97$ and .99 respectively), this means that once M (or R) is in the model M × Size (or R × Size) is unlikely to help unless there is a dramatic change in the slope.

The marginal significance of the intercept is of little interest, especially as there is no data anywhere near the origin and there is no reason to assume linearity outside the range of the data. In any case, there is no reason to expect a zero intercept.

In view of the above results, a C_p analysis of all possible models was performed (see Draper and Smith, Chapter 6). The C_p statistic is calculated in such a way that, if no important variables have been left out of the model C_p should be approximately equal to p, where p is the total number of parameters in the model (including the constant term b_0). There is no statistical difference between models with $C_p \approx p$. The use of C_p enables an analyst to see if there are several equally good models. The output from a C_p analysis of these data is provided in Table 4. Note that the basic variable size (*T*) was included in all models.

A study of the information in this table shows that a model with T, R and MT has $C_p \approx p$ ($p = 4$, $C_p = 4.3$). Models with T, MT and RT or with T, M and R have slightly larger C_p values (5.4 and 5.5 with p=4) but are also worth considering. In particular, the model containing T, M and R has the advantage of simplicity in that is corresponds to three parallel lines. The adequacy of this model can be tested by finding the difference in the residual sums

Table 4: The results of a C_p analysis of the model.
All models studied contains the linear term in size.

Number in Model (p)	C_p	Residual Sum of Squares	Variables in Model
2	88.3	5793922.5	T
Note: The above variable is included in all models to follow.			
3	8.5	4685045.6	MT
3	9.8	4702539.6	M
3	49.4	5239059.8	R
3	51.8	5271278.7	RT
4	4.3	4599982.8	R MT
4	5.4	4614830.4	MT RT
4	5.5	4615622.0	M R
4	6.2	4626287.7	M RT
4	10.5	4684814.2	M MT
4	50.3	5224468.9	R RT
5	4.4	4574035.4	R MT RT
5	6.3	4599979.8	M R MT
5	6.6	4604263.2	M R RT
5	7.3	4614446.3	M MT RT
6	6.0	4568558.3	M R MT RT

of squares between it and the model with all the parameters in it. This is given by the additional sum of squares

$$SS(MT, RT/Constant, T, M,R) = 4615622.0 - 4568558.3 = 47063.7 .$$

This sum of squares has two degrees of freedom.

The significance of the contribution of the two additional terms is given by

$$p = Prob\{F(2,337) > \frac{47063.7}{13556.6}\}$$

$$= Prob\{F(2,327) > 1.74\}$$

$$\approx 0.18 .$$

Thus, once the intercept adjustment terms are in the model, the slope adjustment terms are not significant. The resulting model is

$$w = -217.13 + 1.41t + 110.28M + 36.98R . \qquad [6]$$

This means that relative to the original operating conditions, on average, one can expect that about 110 more units steel can be processed using the modified conditions and an additional 37 units using the refined conditions. The differences between the conditions are very significant from a statistical point of view. The results of an analysis of this model are given in Table 5. A test of lack of fit of the model is included. It is not remotely significant.

Realistically, it's very unlikely that the lines are parallel, or even perfectly linear, the table below does, however, model the data adequately from a statistical point of view and, most importantly, it is relatively easy to explain to non-statisticians. As well, it clearly shows the direction, magnitude and significance of the improvement due to the changes.

Although there is very strong evidence to support the conclusions, it will be noticed that $R^2 = .387$. This would be judged to be very small by many analysts. In fact, R^2 can be very misleading (see Weisburg, 1985, Example 3.4). A more meaningful statistic is the Root MSE. In this case, it is 116.7. This says (approximately) that individual data points (amounts of steel processed) are distributed around the predicted values with a standard deviation $s=116.7$. This value is very large considering the magnitude of the numbers. There is a need

Table 5: ANOVA table for the final model with a test of lack of fit.

Source	DF	SS	MS	F	P
Model	3	2911305.4418	970435.14726	70.17	<< .0001
Lack of Fit	19	189864.37	9992.86	.72	> .50
Pure Error	320	4425757.68	13830.49		
Corrected Total	342	7526927.4869			

Root MSE	116.68506	R-square	0.3868
Dep Mean	370.47230	Adj R-sq	0.3814
C.V.	31.49630		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H_0: Parameter=0	Prob > |T|
Intercept	1	-217.130	47.146	-4.605	0.0001
Size	1	1.412	0.132	10.672	0.0001
M	1	110.279	16.297	6.767	0.0001
R	1	36.983	14.637	2.527	0.0120

for further study of why the amounts of steel processed vary so much for a given size (from146 to 764 units at $t = 352$ for the refined conditions). Note that the assumption of normality of the data around the line is supported by normal plots of the residuals.

It was also found that s was lower for the original conditions ($s = 78.4$) than for the modified ($s = 128.5$) and the refined ($s = 125.4$) conditions. The overall Root MSE is essentially a weighted average of these values. Further efforts should be made to find out why the variation in the amount of steel processed has increased. This does not, of course, detract from the substantial average increases for modified over the original conditions and for refined over modified.

Residual plots indicated that there might be a change in slope for the modified conditions. This was found to be due entirely to the data for size = 484. Removal of this set of data lead to a model

$$w = -176.00 + 1.29\,t + 103.03M + 45.21R\ . \qquad [7]$$

which differs slightly from the previous model [6]. There is, however, essentially no difference in the magnitude of the improvement of moving from original to modified conditions [+103 vs. +110] and from modified to refined [+45 vs. +37]. No reason for the slightly unusual behaviour of this set of runs could be found.

Conclusion

Very substantial savings resulted from moving to the refined changeover procedure and mill operating conditions. This resulted from finding the "assignable cause" for the outliers in the original data set. Although careful study of the appropriate scatter diagrams might lead to this result, the statistical analysis was able to estimate the magnitude of the improvement and the strength of the evidence supporting the conclusions. Although the initial goal was productivity improvement, the result was a substantial improvement in *both* quality and productivity.

References

Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis.* John Wiley & Sons, New York

Weisburg, S. (1985). *Applied Linear Regression.* John Wiley & Sons, New York.

APPENDIX

Table A1: Coded values for the amount of steel processed under the *original operating conditions* before a pass had to be changed.

Coded Bar Size t

446	414	382	352	324	300	288	274	250
362	538	316	224	332	134	288	266	276
262	198	422	150	204	222	164	232	124
406	318	452	270	74	228		244	
460	364	536	426	252	282		132	
398	304	194	292	240	272		322	
526	256	550	318	284	168		302	
372	372	428	122	204	138		214	
424	264	528	308	384	116		390	
	476	508	304	252	130		176	
	360	382	386	190	302		264	
	340	378	290	276	320		48	
		454	352	198	152		332	
			206	160	200			
			220	86	208			
			316	244	288			
			262	320	324			
			262	288	258			
			200	346	210			
			274		256			
			468		312			
			344		168			
					168			
					86			
					200			
					190			

APPENDIX

Table A2: Coded values for the amount of steel processed under the *modified operating conditions* before a pass had to be changed.

Coded Bar Size t

446	414	382	352	324	300	274	250
492	498	408	396	474	268	252	298
584	482	326	332	366	262	204	
496	534	364	412	418	360	172	
556	310	510	210	360	190	288	
748	622	526	288	490	300	262	
764	364	402	454	478	224	182	
706	520	306	726	446	180	316	
496	364	476	398	452	300	216	
562	642	412	390	420	294	218	
570	368	448	316	338	246	178	
	286	576	458	340	406	306	
	374	270	210	438	204		
		240	292	454	294		
		674	164	174	534		
		444	242	142	320		
		640	292	196	344		
		236	490	160	338		
		428	314	222	292		
		416	808	114	282		
			524	152	476		
			646	170	302		
			622	714	266		
			586		228		
					274		
					512		

APPENDIX

Table A3: Coded values for the amount of steel processed under the *refined operating conditions* before a pass had to be changed.

Coded Bar Size *t*

446	414	382	352	324	300	288
756	688	506	764	414	440	326
516	612	348	416	452	318	310
678	446	498	324	408	262	282
444	766	292	504	632	468	
610	438	518	564	678	410	
	600	422	404	658	336	
	338	534	526	546	388	
	386	432	664	344	344	
	426	604	532	340	282	
	528	472	146	434	360	
	492	458	224	676	458	
	374	482	556	338	332	
	408	588	240	408	300	
	412	464	192	370	436	
	370	488	314	234	254	
	364	398	312	498	448	
	404	502	406	528	458	
		256	376	132	296	
		410	394	354		
		306	276	316		
		562	240	500		
		722	584	332		
		670	400	332		
		674	524	420		
		322	406	146		
		338	512	192		
				296		
				488		
				358		
				358		

January 16, 1994

Mr. Thomas Pyzdek
Quality Publishing Inc.
2405 N. Avenida Sorgo
Tuscon, Arizona
85749-9305

Dear Mr. Pyzdek:

Thank you for your prompt and very helpful suggestions regarding the paper "Quality and Productivity Improvement Using Regression Analysis: A Case Study" that I submitted for publication in Quality Engineering. I have enclosed a revised copy of the paper that incorporates your suggestions.

Unfortunately, I do not have access to the actual dollar value of the improvements. I have, however added Tables 2 and 7 in which I quantify the magnitude of the increase in pass life. The benefit of the combination of an average 52.3% increase in pass life and improved finished product quality should be clear to readers.

As far as the "real world" steps involved in moving from the original to the refined conditions, I have now explained on pages 8 and 11/12 that the refined (r) procedure is in fact a refinement of the modified procedure and thus, that the process can be both refined and modified at the same time. I have also pointed out that changes were made to both operating and changeover procedures. This was an oversight in the original version. In my role as consultant, I was asked to analyze the data and was not privy to actual details of the changes.

As requested, I have included the additional figures you requested (Figure 2(a) and 2(b)). Tables 2 and 7 should satisfy the suggestion your made regarding a bar graph of the improvements.

I hope my additional discussion of the modified and refined procedures and the discussion on page 11/12 will clarify your concern regarding the dummy variables.

I have also enclosed large laser printed versions of all the diagrams in case they are needed by the printers.

Hopefully these changes will satisfy our concerns.

Yours truly,

J. Clifton Young

QUALITY AND PRODUCTIVITY IMPROVEMENT USING REGRESSION ANALYSIS: A CASE STUDY

J. Clifton Young
Institute for Improvement in Quality and Productivity
University of Waterloo

ABSTRACT

Although there is an enormous amount of data being collected in industry today, the unfortunate fact is that little of it is ever analyzed. One effective method of looking at such data is regression analysis. In this example an initial regression analysis revealed a very critical clue as to how to achieve major gains in both quality and productivity. Multiple linear regression of subsequent trial data produced strong evidence of major improvements.

Key Words: Regression analysis; Indicator variables; Productivity improvement; Dummy variables; Multiple linear regression.

About the Author:

J. Clifton Young was the founding Director of the Institute for Improvement in Quality and Productivity at the University of Waterloo. He is an associate professor in the Department of Statistics and Actuarial Science. Clif received a B.A.Sc. in Electrical Engineering from the University of Toronto and a Ph.D. in Statistics from the University of Edinburgh. He continues to consult on the use of statistical methods for process improvement with a wide range of industries.

Introduction

The following study was undertaken to increase productivity of a bar mill in a steel plant. A bar mill consists of a series of "passes" or stations at each of which the size of an incoming slab of steel is reduced until the desired final thickness (gauge) is achieved. The passes wear over time with the result that the final product has poor appearance or is out of specification limits. When this starts to occur the pass is changed. Changeovers result in lost time and substantial replacement costs.

For confidentiality reasons, the data have been coded and presentation of the results delayed; thus the magnitude of the results is not representative of the steel industry as it exists today. The magnitude of the improvement is accurate.

The example shows how regression analysis can be used to draw conclusions from complex data sets and more specifically, how to use indicator or "dummy" variables to model qualitative factors in a process.

The study also illustrates a situation in which a careful search for an "assignable cause" for outliers resulted in major gains in *both* quality and productivity.

Analysis of the Original Data Set

The initial clue regarding the possibility of a significant improvement in productivity and quality came from an analysis of the relationship between the weight (w) of steel with a final bar size or thickness (t) that was processed before one or more passes had to be replaced. It should be stated that once the mill is set up to run a specific size, only that size is run until it is necessary to shut the line down to change a pass.

The initial data set, which was collected over an eight month period, is plotted in Figure 1. Since a smaller final size requires more effort on the part of the passes, it was expected that the greater the final thickness the greater the amount of steel processed before a changeover was required. That this is true can be seen in the scatter diagram. The exact nature of the relationship was determined using linear regression to fit a model to the data. The resulting model was

$$w = -134.03 + 1.225\, t. \qquad [1]$$

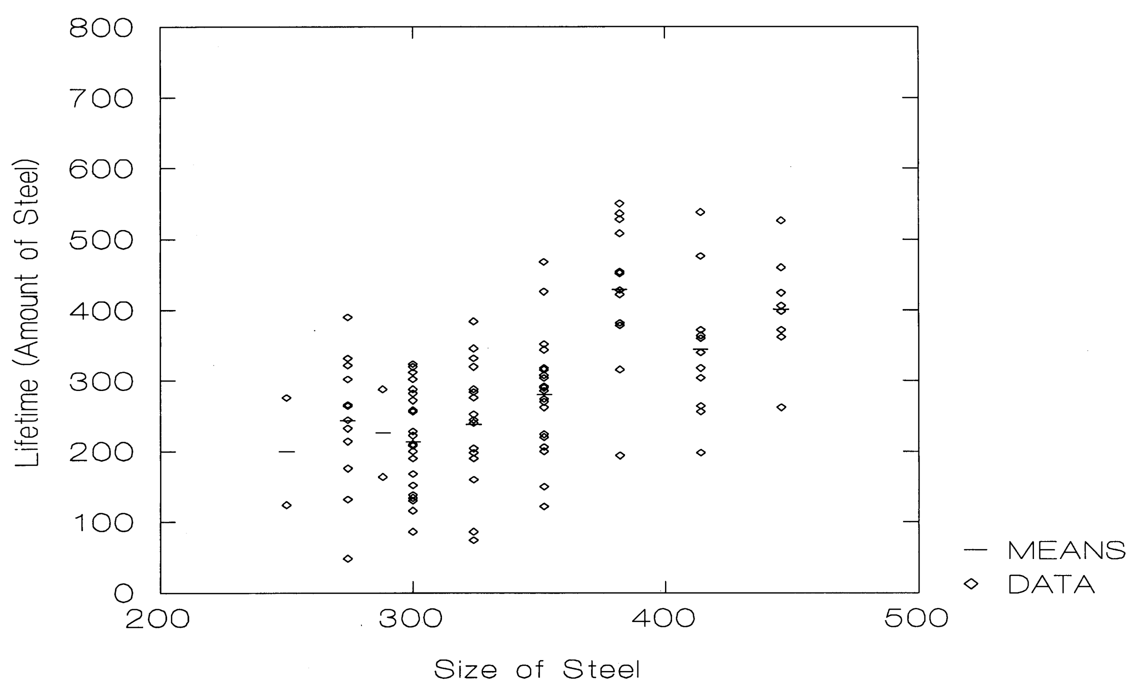


Figure 1: Scatter diagram showing the relationship between the amount of steel processed (lifetime) and steel size under the *original conditions*

Since there are several independent runs at each of the nine sizes, it was possible to check the adequacy of the linear model by means of a standard "lack of fit" test (Draper and

Smith, 1981, Section 1.5). The results of this test are given in the analysis of variance (ANOVA) presented in Table 1.

Table 1: Analysis of variance table showing lack of fit test for the data collected under the original operating conditions.

Source	DF	SS	MS	F	P
Linear term	1	439546.66	439546.66	61.79	<< .0001
Lack of Fit	7	168701.48	24100.21	3.39	.0027
Pure Error	103	732730.72	7113.89		
Corrected Total	111	1340978.86			

Not surprisingly, when compared to the pure error, there is very strong evidence of a relationship between weight and thickness (a significance level **$p < .0001$**).

The engineers involved were, however, surprised to see the strong evidence of a lack of fit of the linear model (**$p = .0027$**).

In order to determine the nature of the lack of fit, the mean values of weight were calculated for each of the nine thicknesses, and included in Figure 1. It can be seen that, although there is a significant lack of fit, there is no hint of a smooth curve in the mean values. This can be seen most clearly in Figure 2(a) which, for clarity, shows only the mean values and the regression line. Note that the line has been calculated using all the data points not just the means. The problem seems to be that the mean amounts processed at **$t = 382$** and possibly **$t = 274$** appear to be well above the line described by the rest of the data. The means at **$t = 288$** and **$t = 250$** are also off the line but, since they are means of only two runs each, this is much less surprising than the other two which are both means of 12 independent runs.

The remaining five means (all from large samples) lie on an almost perfect straight line.

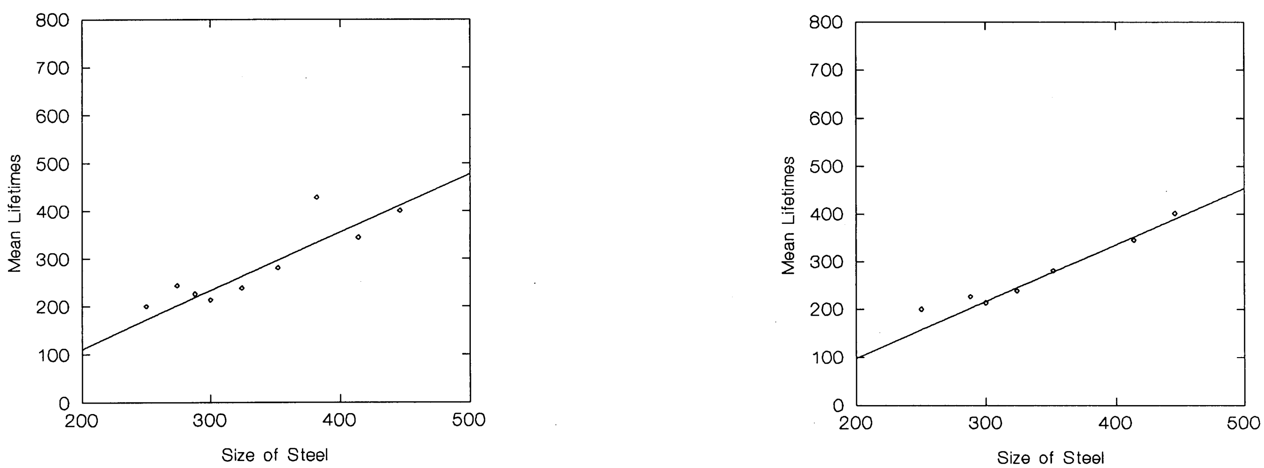


Figure 2(a): All Sizes

Figure 2(b) t=382 and t=274 Removed

Figure 2: Scatter diagrams showing the mean amount of steel processed at each steel size and the corresponding regression line.

Dropping the data sets at **t = 382** and **t = 274** results in the model

$$w = -139.13 + 1.185t \qquad [2]$$

and the ANOVA in Table 2. This new model and the remaining seven means are shown in Figure 2(b).

Table 2: ANOVA table showing a test of lack of fit after dropping two outlying data sets. ($t = 382$ *and* $t = 274$)

Source	**DF**	**SS**	**MS**	**F**	**P**
Linear term	1	306763.37	306763.37	47.73	<< .0001
Lack of Fit	5	7711.22	1542.24	.24	> .5
Pure Error	81	520573.72	6426.84		
Corrected Total	87	835048.31			

As can be seen in Table 2, there is no longer any evidence whatsoever of a lack of fit (**p > 0.5**). There is thus very strong evidence that the data at the removed points do not fit the model that fits the rest of the data quite well. What caught the eye of the engineers involved was the fact that the amounts of steel processed at these points was *significantly higher* than expected according to the model that fits the rest of the data. This corresponds to a significantly longer time between replacement of the passes and hence considerable savings.

When the engineers investigated, they found that the steel rolled at these two sizes was for one specific customer who demanded a very high level of finished product quality. In fact, operating personnel disliked running product for that customer because extra changeover time and a special operating procedure were required to ensure the required level of finished product quality. Everyone "knew" that this would result in fewer tons out the door and, since productivity was measured in tons out the door, reduced productivity. Although the finished product quality level was higher, it was felt to be an unreasonable requirement since no one else asked for it.

What was not realized until this analysis was the fact that the passes lasted longer when special attention was given to the operation. An estimate of the increase in pass life is given by comparing the predicted weight using model [2] with the mean value of the observed lifetimes at **t=382** and **t=274**. The relevant values are given in Table 3. If the results summarized in Table 3 are valid, the special set-up should not only result in better quality but also in an increase of about 40% in the time between shutdowns for pass changes and thus in a corresponding increase in productivity. In order to test this possibility, the special set-up was tentatively adopted for all sizes of steel.

Table 3: Predicted mean lifetimes using model [2] and the corresponding observed mean lifetimes at **t=274** and **t=382**.

Size (t)	Predicted Mean w=-139.13+1.185t	Observed Mean Gain	Percent Increase
274	185.56	274	+47.7%
382	313.54	429	+36.8%

Analysis of Pass Life After Modifications

The line was then run and data collected using the modified operating and changeover procedure. After about four months, further refinements were made and additional data collected for another four months. Note that, since the refinements were made to the modified procedure, the third stage involves *both* the modification *and* the refinement (refined-modified).

The following is a direct quote of the questions asked by the engineers when they supplied the data from the process after the modification (m) and the subsequent refinement (r).

"We would like to determine if the modifications improved the pass life and by how much?

- three sets of data from the same source
- one thing changed from each set (same thing)
- nine groups (bar thickness in thousands of an inch) within each set
- large variations within each group

Questions

- The best method of analyzing the data

Questions

- The best method of analyzing the data
- Do I compare sets?
- Do I compare groups within the sets?
- Can I compare three sets or groups simultaneously?
- Which data (if any) can be rejected statistically?
- The best way of reporting the data, simultaneously if possible?
- To compare groups or sets do I need more data? If so, in which groups or sets?"

The three sets of data referred to are the data from the original (o), the modified (m) and the refined-modified (r) process. The data are plotted in Figure 3 using the symbols **o, r** and **m** for the corresponding conditions. As can be seen, the data sets overlap considerably and it is not at all clear which conditions are best. Figures 4a and 4b show the relationship separately for modified and refined data and include the group means.

In order to answer the engineer's questions, the following model was set up for analysis using multiple regression with two indicator ("dummy") variables (Draper and Smith, 1981, Section 5.4) to account for the three different methods of running the process. The model used was

$$w_i = b_0 + b_1 M_i + b_2 R_i + b_3 t_i + b_4 M_i t_i + b_5 R_i t_i + e_i , \qquad [3]$$

$$= [b_0 + b_1 M_i + b_2 R_i] + [b_3 + b_4 M_i + b_5 R_i] t_i + e_i. \qquad [3a]$$

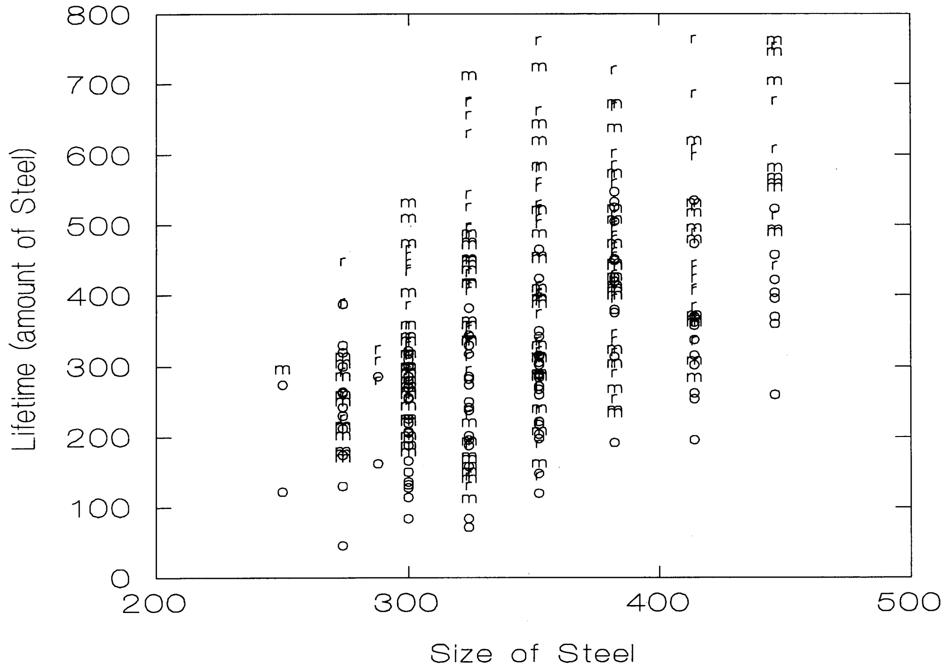


Figure 3: Scatter diagram showing the relationship between the amount of steel processed and the size of steel for all three sets of data. The symbols **o, m** and **r** are used for the corresponding conditions.

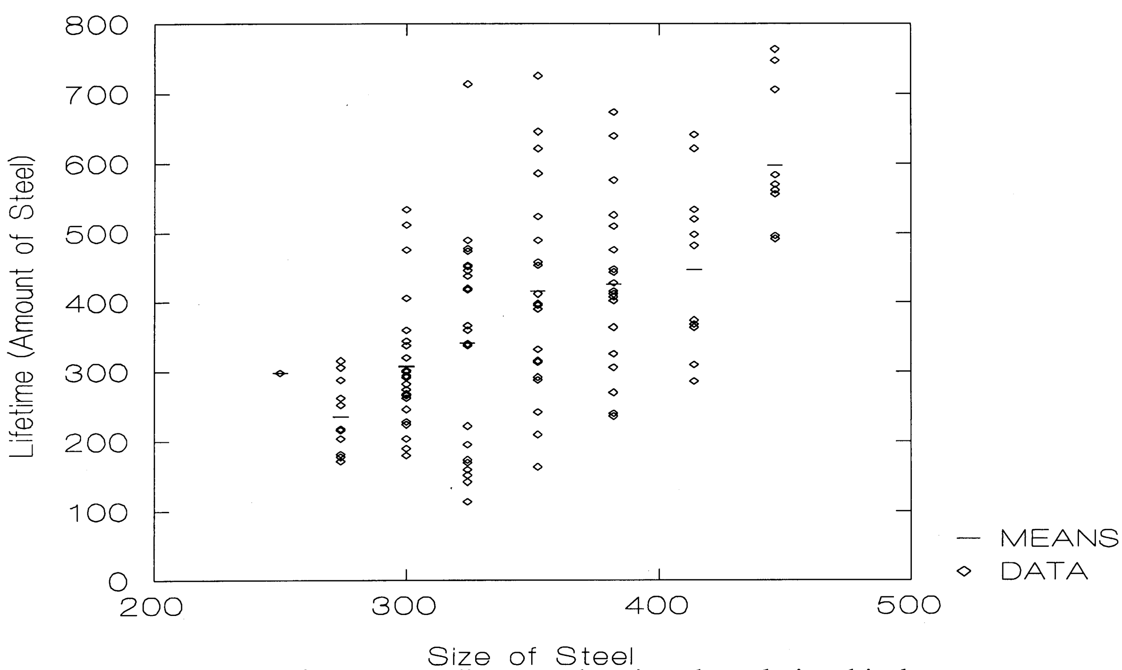


Figure 4a: Scatter diagram showing the relationship between amount and size for the *modified conditions*.

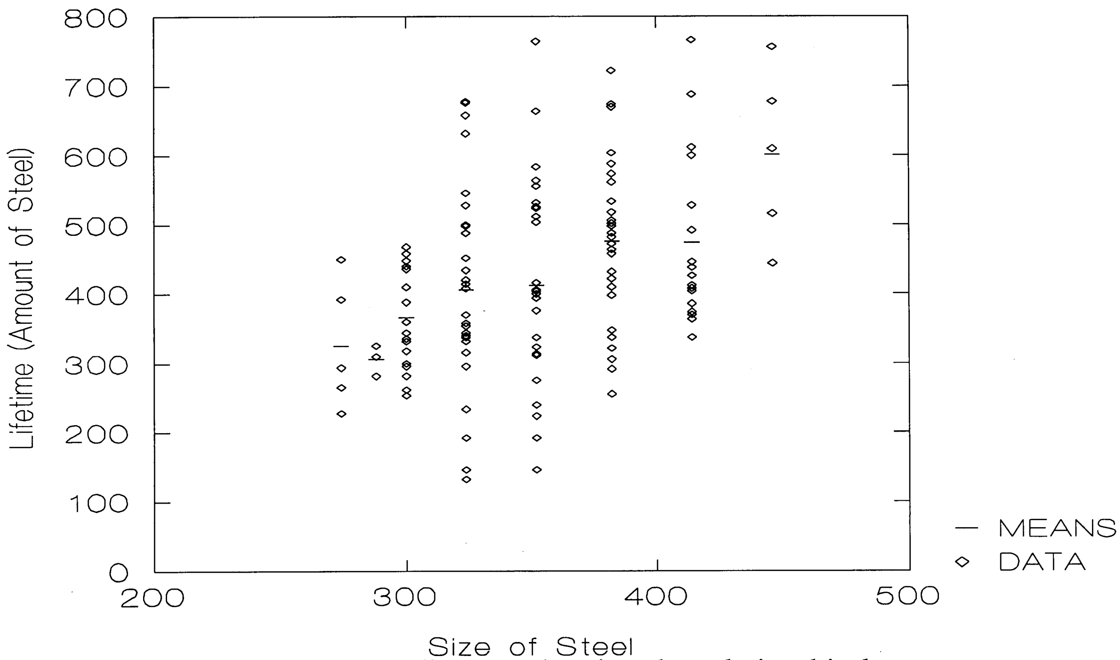


Figure 4b: Scatter diagram showing the relationship between amount and size for the *refined conditions*.

where

w_i is the i[th] observation of weight (amount)

t_i is the thickness (size) corresponding to w_i

M_i is an indicator variable which equals 1 for those observations collected under both modified and refined-modified conditions and 0 otherwise

R_i is an indicator variable which equals 1 for observations collected after the process was refined (refined-modified) and 0 otherwise

$M_i\ t_i$ is the product of M_i and t_i

$R_i\ t_i$ is the product of R_i and t_i

$[b_0 + b_1M_i + b_2R_i]$ is the effective intercept

$[b_3 + b_4M_i + b_5R_i]$ is the effective slope.

Since the refinement involved a relatively minor revision to the modified process, it was expected to be either similar to the modification or, hopefully, slightly better. In order to reflect this fact, the model was set up such that the indicator M_i treats data from *both* the modified and refined series in the same manner ($M_i = 1$ for both and 0 otherwise) while R_i differentiates between the two ($R_i = 1$ for the refinement and 0, otherwise). Thus the significance of terms involving M_i will indicate that, on average, *both the modified and refined conditions differ from the original;* the significance of terms involving R_i will indicate a *difference between the refinement and the modification.*

A more typical approach would be to have $M_i = 1$ for the modified process and zero otherwise and $R_i = 1$ for the refined process and zero otherwise. This approach would work, but the testing of the specific hypothesis of interest would be somewhat more difficult so the approach outlined above was used.

Further insight into the interpretation of the model and the corresponding significance tests may be gained by recognition of the fact that, in effect, the equation is that of a straight line as shown in equation [3a] with

$$intercept = b_0 + b_1 M_i + b_2 R_i \, . \qquad [4]$$

If neither b_1 nor b_2 are significant that would mean that the same intercept (b_0) adequately models all three process conditions; the significance of b_1 but not b_2 would mean that b_0, the intercept for the original conditions must be modified by an amount b_1 (*to* $b_0 + b_1$) for both the modified and refined conditions ($M_i = 1$). Finally, if b_1 and b_2 are both significant, then three separate and significantly different intercepts are needed (b_0 for original conditions, b_1 +

b_2 for modified, and $b_0 + b_1 + b_2$ for refined). Since the refinement involves a minor change to the modification, it would be very suprising to find b_2 significant and b_1 not significant.

In a similar manner the slope of the line is given by

$$slope = b_3 + b_4 M_i + b_5 R_i \ . \qquad [5]$$

The same argument applies to the slope adjustment terms b_4 and b_5. In effect the model [3] simultaneously fits the three separate lines,

$$w_i = b_0 + b_1 t_i \qquad \text{(original conditions)}$$

$$w_i = (b_0 + b_1) + (b_3 + b_4)t_i \qquad \text{(modified conditions)}$$

and

$$w_i = (b_0 + b_1 + b_2) + (b_3 + b_4 + b_5)t_i \qquad \text{(refined conditions)}$$

simultaneously. Tests for the significance of b_2, b_3, b_4 and b_5 enable the assessment of the significance of differences between the lines.

An analysis of variance table and other details of an analysis of the above model are given in Table 4. The format of the entire table is typical of the output from most standard multiple regression computer packages.

The results in the ANOVA table show overwhelming evidence ($p < .0001$) that the model is statistically significant. However, the results in the lower "Parameter Estimates" portion of the table show only the slope *(b_1)* of size as being strongly significant ($p < .0001$) with R marginally significant. As well, the intercept, M × Size and R × Size might also be regarded as marginally significant (p around .10). It is, however, important to take great care in interpreting these tests. In each of these t-tests the significance of the contribution of the parameter is tested when it is *added* to a model *containing all the other parameters*. When simple correlations, r, are calculated between the predictor (independent) variables for this model, it is found that R and R × Size (r=.97) and M and M × Size (r=.99) are very highly correlated. This means that once M (or R) is in the model M × Size (or R × Size) is unlikely to improve the model unless there is a dramatic change in the slope.

Table 4: ANOVA table for model [3] with tests for individual parameters

Source	**DF**	**SS**	**MS**	**F**	**P**
Model	5	2958369.1938	591673.83877	43.645	<<.0001
Error	337	4568558.293	13556.552798		
Corrected Total	342	7526927.4869			

Root MSE	116.43261	R-square	0.3930
Dep Mean	370.47230	Adj R-sq	0.3840
C.V.	31.42816		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H_0: Parameter=0	Prob > |T|
Intercept	1	-139.127	86.523	-1.608	0.1088
Size	1	1.185	0.249	4.757	0.0001
M	1	-72.052	113.357	-0.636	0.5255
R	1	203.319	110.510	1.840	0.0667
M × Size	1	0.527	0.325	1.623	0.1055
R × Size	1	-0.476	0.313	-1.522	0.1288

The marginal significance of the intercept is of little interest, especially as there is no data anywhere near the origin and there is no reason to assume linearity outside the range of the data. In any case, there is no reason to expect a zero intercept.

In view of the above results, a C_p analysis of all possible models was performed (see Draper and Smith, Chapter 6). The C_p statistic is calculated in such a way that, if no important variables have been left out of the model C_p should be approximately equal to p, where p is the total number of parameters in the model (including the constant term b_0). There is no statistical difference between models with $C_p \approx p$. The use of C_p enables an analyst to see if there are several equally good models. The output from a C_p analysis of these data is provided in Table 5. *Note that the basic variable size (T) was included in all models.*

A study of the information in this table shows that a model with T, R and MT has $C_p \approx p$ ($p = 4$, $C_p = 4.3$). Models with T, MT and RT or with T, M and R have slightly larger C_p values (5.4 and 5.5 with p=4) but are also worth considering. In particular, the model containing only T, M and R has the advantage of simplicity in that is corresponds to three parallel lines. The adequacy of this model can be tested by finding the difference in the residual sums of squares between it and the model with all the parameters in it. This is given by the additional sum of squares

$$SS(MT,\ RT/Constant,\ T,\ M,\ R) = 4615622.0 - 4568558.3 = 47063.7\ .$$

This sum of squares has two degrees of freedom.

The significance of the contribution of the two additional terms is given by

$$p = Prob\{F(2,337) > \frac{47063.7 \div 2}{13556.6}\}$$

$$= Prob\{F(2,327) > 1.74\}$$

$$\approx 0.18\ .$$

Thus, adding slope adjustments to a model containing the intercept adjustments does not significantly improve predictive ability. The resulting model is

$$w = -217.13 + 1.41t + 110.28M + 36.98R \ . \qquad [6]$$

Table 5: The results of a C_p analysis of the model.
All models studied contain the linear term in size (T).

Number in Model (p)	C_p	Residual Sum of Squares	Variables in Model
2	88.3	5793922.5	T
***Note**: The above variable (T) is included in all models to follow.*			
3	8.5	4685045.6	MT
3	9.8	4702539.6	M
3	49.4	5239059.8	R
3	51.8	5271278.7	RT
4	4.3	4599982.8	R MT
4	5.4	4614830.4	MT RT
4	5.5	4615622.0	M R
4	6.2	4626287.7	M RT
4	10.5	4684814.2	M MT
4	50.3	5224468.9	R RT
5	4.4	4574035.4	R MT RT
5	6.3	4599979.8	M R MT
5	6.6	4604263.2	M R RT
5	7.3	4614446.3	M MT RT
6	6.0	4568558.3	M R MT RT

This means that, relative to the original operating conditions, on average, one can expect that about 110 more units steel can be processed using the modified conditions and an additional

37 units using the refined conditions. The differences between the conditions are very significant from a statistical point of view. The results of an analysis of this model are given in Table 6. A test of lack of fit of the model is included. It is not remotely significant.

Realistically, it is very unlikely that the lines are parallel, or even perfectly linear, the table below does, however, model the data adequately from a statistical point of view and, most importantly, it is relatively easy to explain to non-statisticians. As well, it clearly shows the direction, magnitude and significance of the improvement due to the changes.

Although there is very strong evidence to support the conclusions, it will be noticed that $R^2 = .387$. This would be judged to be very small by many analysts. In fact, R^2 can be very misleading (see Weisburg, 1985, Example 3.4). A more meaningful statistic is the Root MSE. In this case, it is 116.7. This says (approximately) that individual data points (amounts of steel processed) are distributed around the predicted values with a standard deviation $s=116.7$. This value is very large considering the magnitude of the numbers. There is a need for further study of why the amounts of steel processed vary so much for a given size (from 146 to 764 units at $t = 352$ for the refined conditions). Note that the assumption of normality of the data around the line is supported by normal plots of the residuals.

It was also found that s was lower for the original conditions ($s = 78.4$) than for the modified ($s = 128.5$) and the refined ($s = 125.4$) conditions. The overall Root MSE is essentially a weighted average of these values. Further efforts should be made to find out why the variation in the amount of steel processed has increased. However, changeover costs are directly proportional to the average number of changeovers, and hence to the average amount of steel processed between changeovers. Thus the apparent increase in variability does not detract from the substantial average increases found when moving to the modified and then to the refined modification.

Table 6: ANOVA table for the final model with a test of lack of fit.

Source	**DF**	**SS**	**MS**	**F**	**P**
Model	3	2911305.4418	970435.14726	70.17	<< .0001
Lack of Fit	19	189864.37	9992.86	.72	> .50
Pure Error	320	4425757.68	13830.49		
Corrected Total	342	7526927.4869			

Root MSE	116.68506	R-square	0.3868
Dep Mean	370.47230	Adj R-sq	0.3814
C.V.	31.49630		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H_0: Parameter=0	Prob > |T|
Intercept	1	-217.130	47.146	-4.605	0.0001
Size	1	1.412	0.132	10.672	0.0001
M	1	110.279	16.297	6.767	0.0001
R	1	36.983	14.637	2.527	0.0120

Residual plots indicated that there might be a change in slope for the modified conditions. This was found to be due entirely to the data for size = 484. Removal of this set of data lead to a model

$$w = -176.00 + 1.29\, t + 103.03M + 45.21R \,. \qquad [7]$$

which differs slightly from the previous model [6]. There is, however, essentially no difference in the magnitude of the improvement of moving from original to modified conditions [+103 vs. +110] and from modified to refined [+45 vs. +37]. No reason for the slightly unusual behaviour of this set of runs could be found.

An indication of the magnitude of the savings can be seen on inspection of the Table 7. Although specific dollar values are confidential, the average increase in the region of 52.3% in the amount of steel processed between changeovers resulted in very substantial savings in both production time and the costs involved in changeovers and reconditioning the passes. When combined with the corresponding improvement in quality, the gains were especially impressive. The cost of using the new method was very small relative to the gains obtained.

Table 7: Average expected increases in the amount of steel processed between changeovers using the modified and refined operating conditions at a steel size of 352 units.

Operating Condition	Expected Average Amount of Steel Processed (Model [6])	Expected Increase over Original Conditions
Original	279.19	---
Modified	389.47	39.5%
Refined-Modified	426.45	52.3%

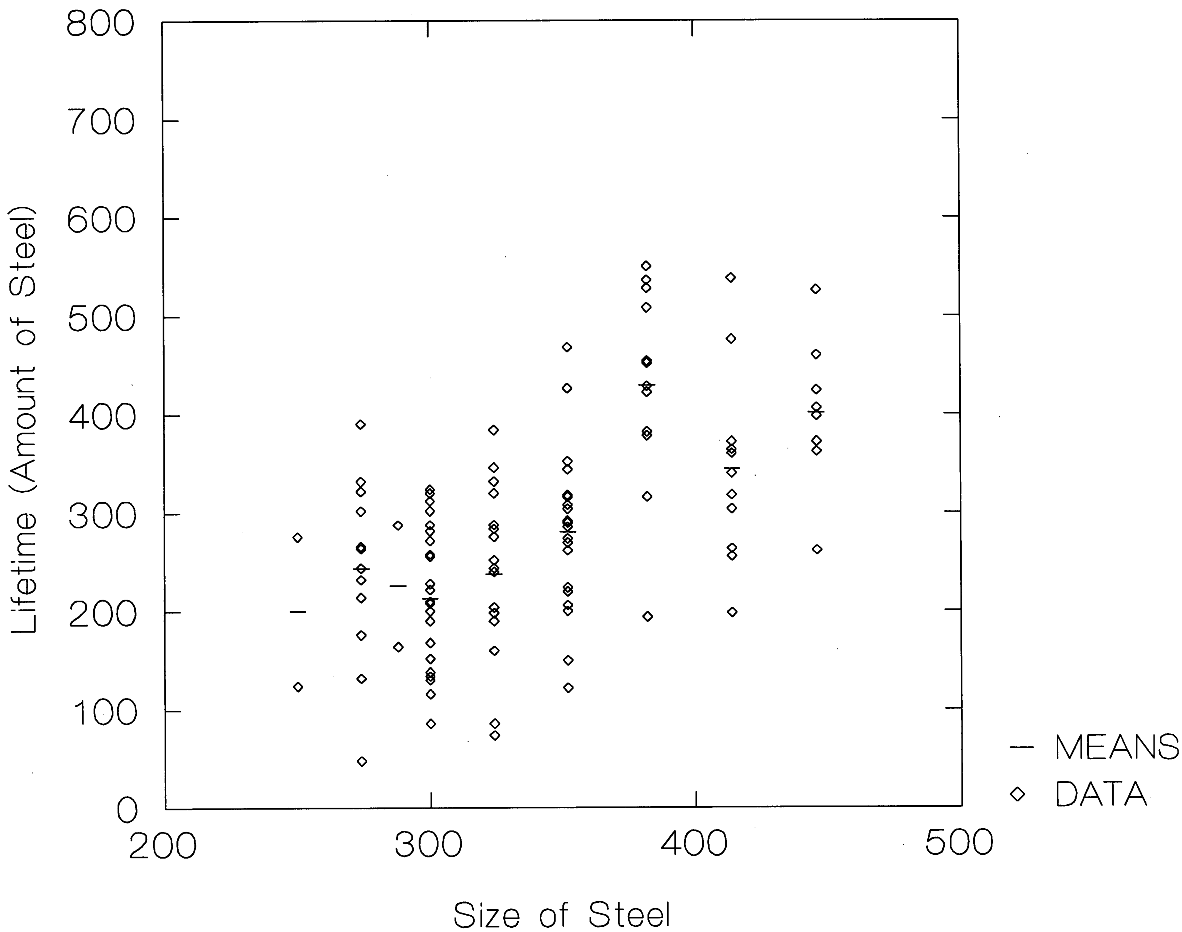
Conclusion

Very substantial savings resulted from moving to the refined changeover procedure and mill operating conditions. This resulted from finding the "assignable cause" for the outliers in the original data set. Although careful study of the appropriate scatter diagrams might lead to this result, the statistical analysis was able to estimate the magnitude of the improvement and the strength of the evidence supporting the conclusions. Although the initial goal was productivity improvement, the result was a substantial improvement in *both* quality and productivity.

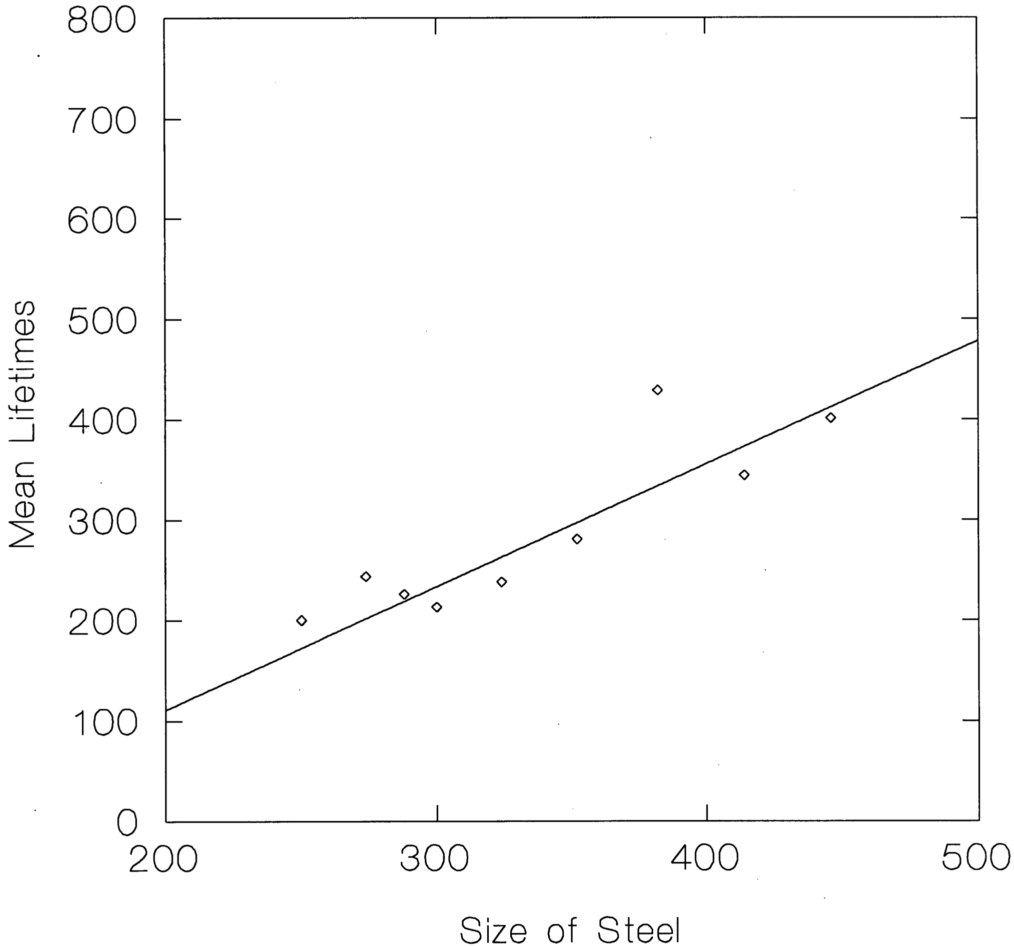
References

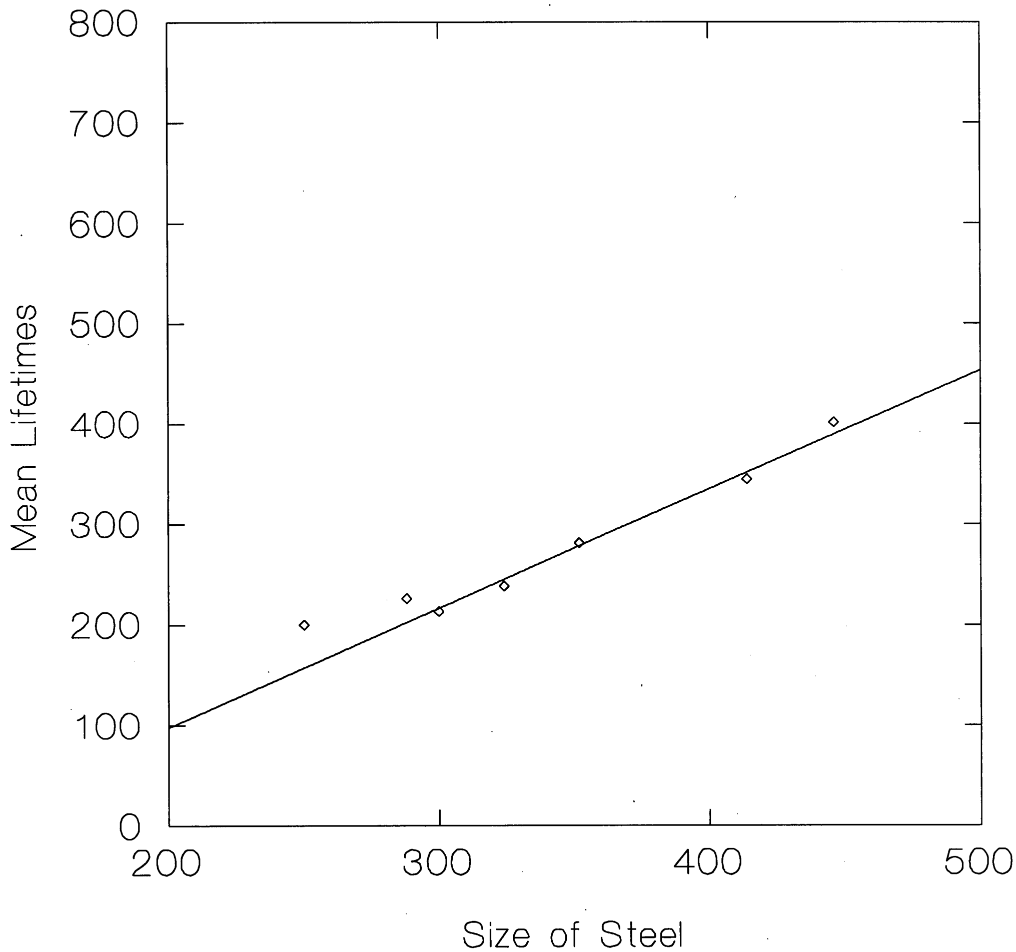
Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis.* John Wiley & Sons, New York

Weisburg, S. (1985). *Applied Linear Regression.* John Wiley & Sons, New York.

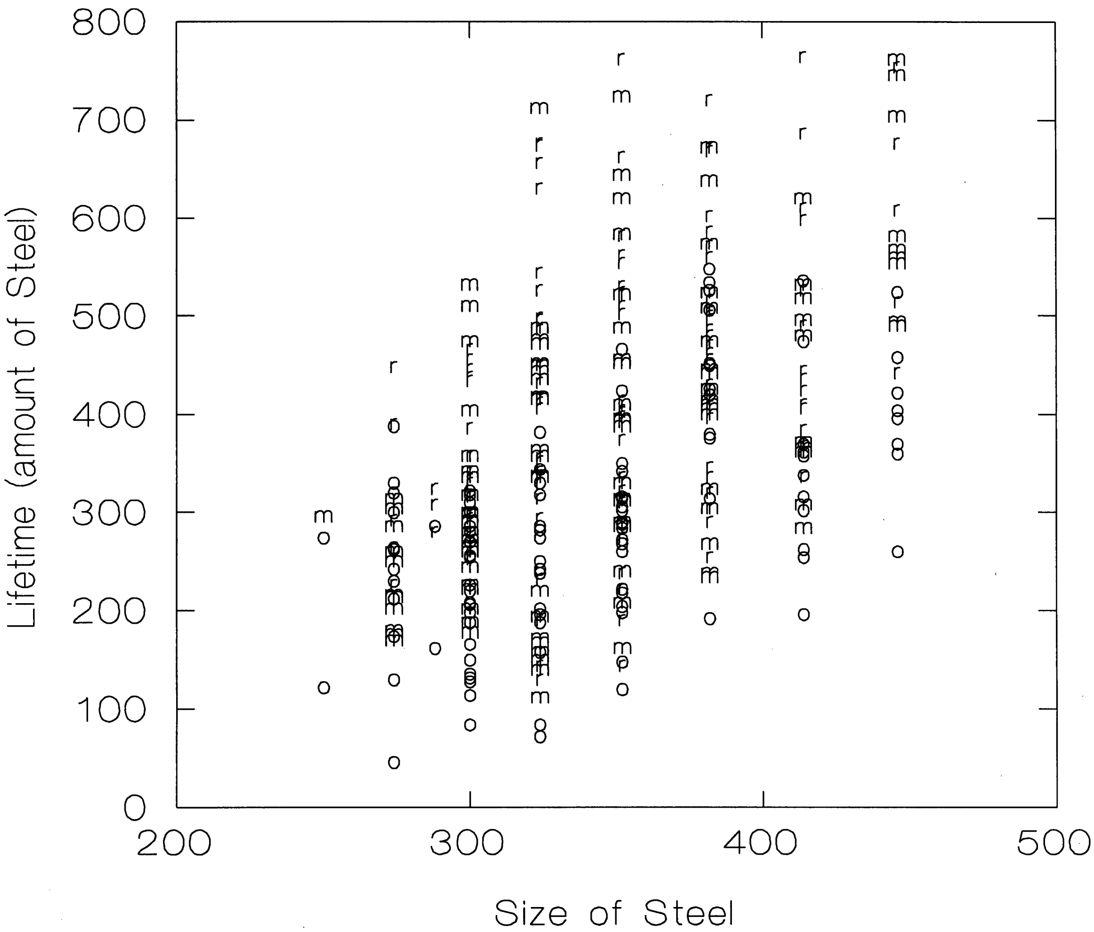


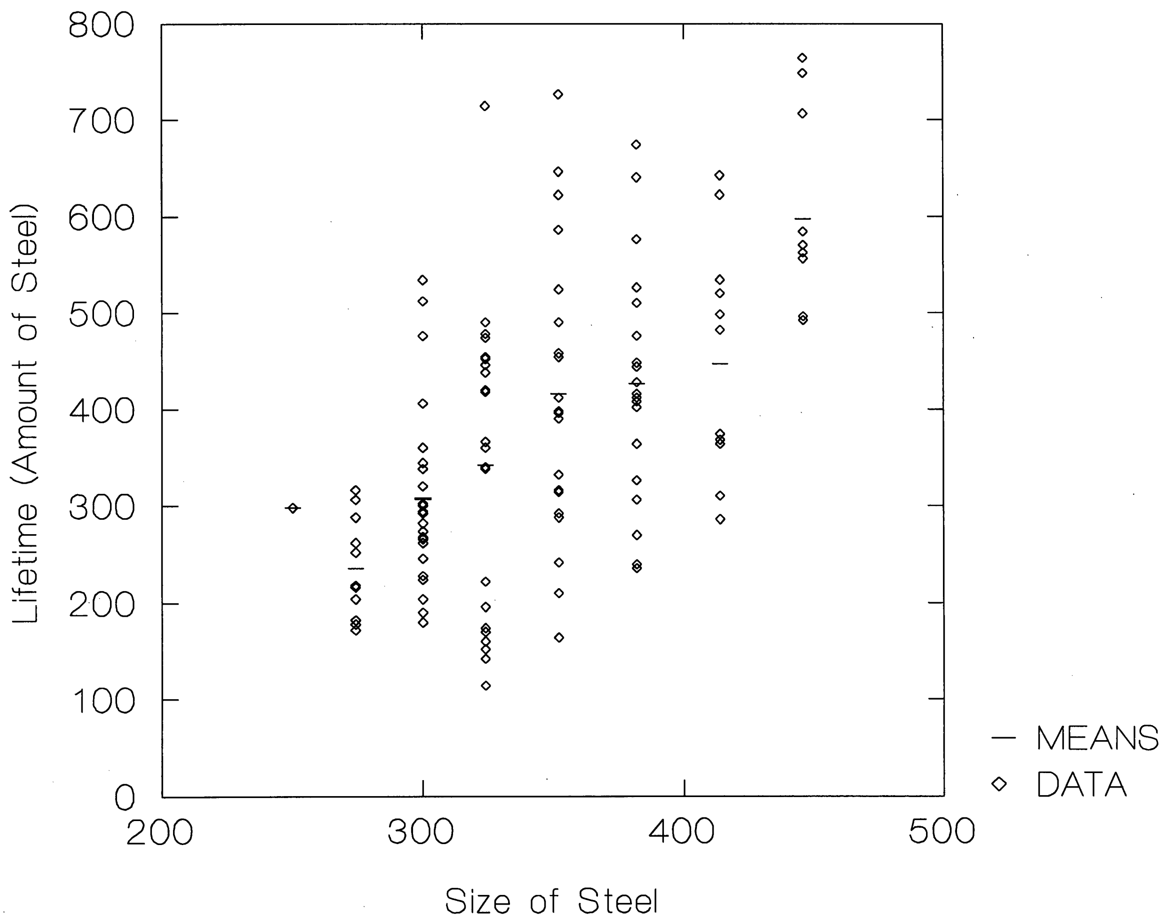
800
700
600
500
400
300
200
100
0
200
300
400
500
Lifetime (Amount of Steel)
Size of Steel
MEANS
DATA


800
700
600
500
400
300
200
100
0
Mean Lifetimes
200
300
400
500
Size of Steel

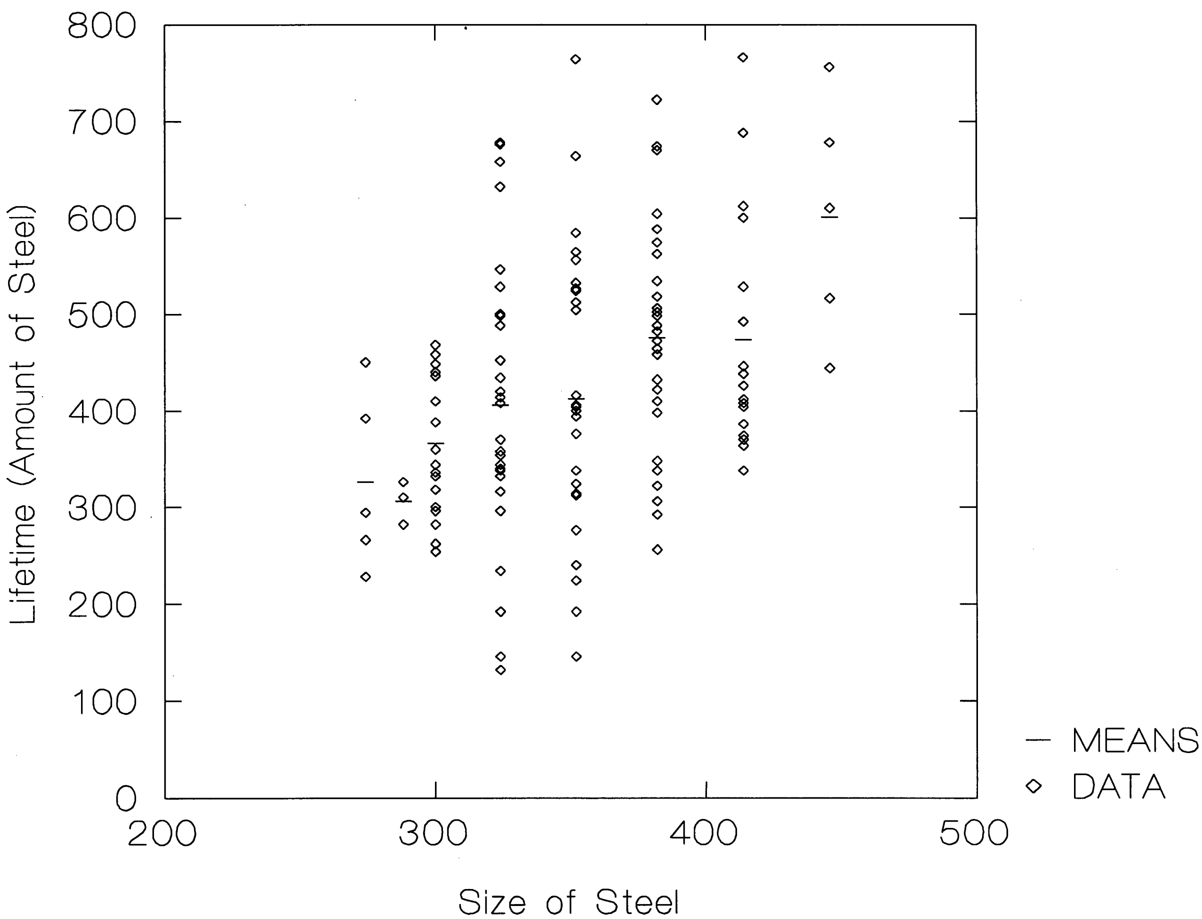


800
700
600
500
400
300
200
100
0
Mean Lifetimes
200
300
400
500
Size of Steel


800
700
600
500
400
300
200
100
0
Lifetime (amount of Steel)
200
300
400
500
Size of Steel



800
700
600
500
400
300
200
100
0
200
300
400
500
Lifetime (Amount of Steel)
Size of Steel
MEANS
DATA


800
700
600
500
400
300
200
100
0
200
300
400
500
Lifetime (Amount of Steel)
Size of Steel
MEANS
DATA