Identification of Critical Key Performance Indicators for Anglo Coal Underground Mining Operations

by

Madinoge Jeanette Maserumule
s25310454

Submitted in partial fulfilment of the requirements for the degree of
BACHELOR OF INDUSTRIAL AND SYSTEMS ENGINEERING

in the

FACULTY OF ENGINEERING, BUILT ENVIRONMENT AND INFORMATION TECHNOLOGY

UNIVERSITY OF PRETORIA

October 2009
Executive summary

The most essential task of any production company is to be able to manage production with reliable methods developed using realistic data, thus creating a platform for continuous improvement using forecasting methods that can predict the future given past data. The objective of this paper is to study key performance indicators using statistical tools and develop a model that can be used to identify a combination of key performance indicators that are critical in terms of reaching production targets. Literature review was done to highlight relevant tools, techniques and methods that can be used to develop an analysis model that will reach the objectives of the project. This should aid Anglo Coal South Africa in accelerated decision making capabilities utilising proposed solution strategies to achieve stated business objectives.
Acknowledgements

I wish to express my gratitude to people who through their contribution made this project possible.

Thank you Mr. Derek Laing from Anglo Coal South Africa, Mr. Zaaids Van Zyl, Mr. Gary Smith and Mrs. Lauren Theron from Mining Consulting Services.

Additionally I would like to acknowledge the contribution of the following people:
Mr Olufemi Adentunji, for the reviews and provision of theoretical guidance. Andile Ndiweni and Takalani Mudau for guidance and support.
## Contents

1. Introduction ..................................................................................................................... 1  
   1.1 Mining Process ............................................................................................................. 2  
   1.2 Pick Six ....................................................................................................................... 3  
   1.3 Problem Statement ..................................................................................................... 6  
   1.4 Research design and methodology ............................................................................. 7  

2. Literature review ............................................................................................................. 8  
   2.1 Problem formulation .................................................................................................... 8  
   2.2 Time series analysis .................................................................................................... 8  
      2.2.1 Smoothing ............................................................................................................ 9  
   2.3 Multiple regression analysis ...................................................................................... 9  
      2.3.1 Least squares estimation .................................................................................... 10  
      2.3.2 Partial ordinary least squares analysis ........................................................... 10  
      2.3.3 Sequential ordinary least square analysis ......................................................... 11  
   2.4 Principal component analysis ................................................................................... 12  
   2.5 Linear programming .................................................................................................. 12  
   2.6 Excel model ............................................................................................................... 13  
   2.7 Conclusion ................................................................................................................ 14  

3. Multiple regression analysis ........................................................................................ 15  
   3.1 Terminology .............................................................................................................. 16  
   3.2 General structure ...................................................................................................... 16  
   3.3 Statistical significance ............................................................................................... 21
4. Computational Results ................................................................. 24
   4.1 Model outputs ................................................................. 24
   4.2 Solution quality ............................................................... 26

5. Recommendations ................................................................. 27
   5.1 Future work ................................................................. 27
List of figure

1.1: Shift breakdown...........................................................................................................................................2
1.2: Components of a shift time..................................................................................................................................2
1.3: Mining process, cutting cycle flow chart........................................................................................................3
1.4: KPI tree............................................................................................................................................................4
1.5: Production process..............................................................................................................................................5
2.1: Least squares estimation....................................................................................................................................10
3.1: Studentised residual graph of the final model: Vlaaklagte-Simunye Section.................................23
List of table

3.1: Regression terminology..................................................................................................16
3.2: ‘Pick six’ model symbols..........................................................................................17
4.1: Significant parameter estimates................................................................................25
Acronyms

ACSA  Anglo Coal South Africa
CM    Continuous miner
KPI   Key performance indicator
SAS   Statistical Analysis Software

Definitions

Continuous miner: remote-controlled machine with a large rotating steel drum equipped with tungsten carbide teeth that scrapes coal from the seam.

Production time: production unit of a mine, where coal is mined

Shuttle car: electric powered vehicle with a capacity of 20 tons used to transport coal from a mined seam to a section conveyor belt.

Sump: primary cutting motion of moving cutter head of CM into the coal face.

Tram: the movement process that a CM undergoes to cut coal
Part I

Introduction

Business organisations make use of key performance indicators (KPI’s) to measure the success of production management and to define a suitable set of performance indicators for monitoring ongoing performance. KPI’s are quantifiable measurements that reflect the critical success factors of an organisation thus Anglo Coal South Africa (ACSA) has identified and defined six key performance indicators referred to as the ‘Pick Six’ to manage underground coal mining operations.

ACSA is a division of Anglo American and one of South Africa’s largest coal producers. It operates four underground and five open cut mines in Mpumalanga, Witbank Coalfield. It is well known for producing thermal and metallurgical coal to customers in the inland domestic markets as well as the export market. ACSA’s coal demand has grown tremendously inducing a need to improve production managing methods that can help with managing coal operations, budgeting processes and forecasting tools to predict and lessen the risk of not meeting coal demand.

A standard model has been built to assist in managing coal production using the ‘Pick Six’ and fixed mine parameters. With the aid of a third-party data capturing and management company, Mining Consulting Services, data used to manage coal production is captured daily using data cards that are attached to each operating Continuous miner during a scheduled shift. A face shift report document is prepared daily using the mining data from the previous day illustrating actual key performance indicators relative to the benchmarked KPI’s, required production tonnes, production time etc.

The reports are used as

- a check tool to measure actual coal against budgeted coal targets and
- illustrate graphically the actual coal tonnes per day against the ‘Pick Six’ values
1.1 Mining Process

Every mine has a three shift daily mining processes that consists of nine hours per shift. An illustration of the total shift time is as shown in Figure 1.1. The travelling in and out time has been estimated to be an hour long thus it results in a section time of eight hours which is referred to as face time. The available time daily is 24 hours and for the three shifts 27 hours is worked due to the three hours shift change-over phases. Figure 1.2 shows a typical shift breakdown for most of the underground mines.

**Figure 1.1: Shift breakdown**

- **Day shift**
  - 06:00-15:00
  - 1-1:30 hour travel time
  - 30-40 minutes pre and post meeting

- **Afternoon shift**
  - 14:00-23:00
  - 1-1:30 hour travel time
  - 30-40 minutes pre and post meeting

- **Night shift**
  - 22:00-07:00
  - 1-1:30 hour travel time
  - 30-40 minutes pre and post meeting

**Figure 1.2: Components of a shift time**
During a mining operation, a continuous miner (CM) trams coal using the cutter head that is designed to mitigate methane gas level by absorbing in the fumes and using water sprays to put out fire sparks. The mining process is illustrated graphically in figure 1.3.

![Mining process diagram](image)

**Figure 1.3: Mining process, cutting cycle flow chart**

Mines are different in terms of
- shift patterns
- maintenance allowances
- pillar centers
- size of coal
- haulers
- daily production targets etc.

### 1.2 Pick Six

The ‘Pick Six’ can be categorised into productive and unproductive times. The unproductive times are used as a monitoring tool that has to be minimised and if possible eliminated as to enhance the production time. Productive times are those that have a direct proportional relationship to coal extracted in a shift.
Operational critical success factors for Anglo Coal for underground operations are to;

- achieve planned coal production targets in order to meet customer demand
- operate mining equipments efficiently by reducing the machine idle time

Critical factors are only achievable if the KPI's benchmarked values are reliable and their effect on the coal production tonnes is well understood. The three past projects defined and reviewed the KPI's to ensure that Anglo Coal has a coal managing standard that can be deployed in all the underground mines. [3][6][11] The KPI's are explained as follows;

**First sump late**

It is the time difference between the scheduled start of a shift and the first sump that the CM makes. It decreases the face time available during a shift by the same amount of minutes the shift started late.

*Measurement unit: Minutes*

**Load time**

The total time a CM takes to fill up a shuttle car or battery loader with coal. It makes up the production process together with the away time as depicted in figure 1.5. It is calculated by the following formula per shift;

\[
\text{Total loading time} = \text{sum of all loading time per shift/number of loads per shift}
\]
Away time
It is the time that the CM awaits a shuttle car or battery loader to load it with coal as shown in figure 1.5. The time begins when the CM stops loading coal onto a shuttle car and ends when it starts loading coal onto the next shuttle car.

*Measurement unit: seconds

Tram time per meter cut
Tram time per meter cut is the time taken to tram coal over a meter. It is calculated by summing the total tram time divided over the number of meters cut during a shift.

*Measurement unit: seconds

Downtime
Total non productive time recorded in minutes during a shift that minimises the overall production time. The breakdowns are categorised as follows;

- Engineering breakdowns
- Operational breakdowns
- Sectional conveyor breakdowns
- Trunk conveyor breakdowns
- Other

*Measurement unit: Minutes

**Last sump early**

The time difference between the planned end of a shift and the last sump the CM makes. The KPI time decreases the face time available during a shift.

*Measurement unit: Minutes

The ‘Pick Six’ is used to calculated figures such as the following that can be used to calculate the actual coal production per shift.

- production time
  \[\text{production time} = \text{face time} - \text{downtime} - \text{tramming time}\]
- Average production rate
  \[\text{Average production rate} = \frac{3600}{\text{load time} + \text{away time} + \text{tram time per metre cut}} \times \text{shuttle car capacity}\]

**1.3 Problem Statement**

Each KPI is evaluated individually without any comparison to the other KPI’s as to which one(s) has the highest impact on coal production target. This gives minute results as to how the KPI’s can be collectively utilised to the maximum in order to make better operational decisions. ACSA requires a combination of KPI’s that are significant for reaching coal production targets.
1.4 Research design and methodology

Past and present similar projects were studied as to adapt the methodology and obtain the solution strategy using such methods. Such projects have been performed mostly in the financial, biological, medical environments and all had a common solution strategy which is to predict the response variable using KPI's, in this instance the coal production tonnes per shift/day.

The first phase focused on identifying a solution strategy technique that can be used to identify a set of KPI's. Using KPI information obtained from the Mining Consulting Services, the best technique was selected for the identification process of critical KPI's.

The second phase focused on the solution strategy formulation based on the selected technique to determine the KPI combination. Data was extracted from excel and all the necessary computations were performed as to achieve the desired end goals. The model was inputted with all the different mines data.

For the third phase, computational results for each mine were studied and compared to the other mines. In the case of uncertainties, the model outputs were reviewed as to get an in-depth understanding of the uncertainties. This was mostly due to the fact that some of the mines characteristics differed.
Part II

Literature review

The problem should be defined in such a way that each KPI can be analysed both individually and collectively with the other KPI’s. As mentioned before, each KPI is used to calculate the production time and thus there exists a relationship between the KPI’s and the actual coal tonnes mined. These relationships will not be considered imperative for technique selection, however they will only be considered where applicable.

2.1 Problem formulation

Looking at the objectives of the study, the problem can be formulated in such a way that the KPI’s will be used or tested as to yield maximum coal production tonnages. The ‘Pick Six’ will be classified as independent variables where \( X = \{X_1, \ldots, X_6\} \) denotes the KPI set. The coal production tonnages will be classified as the dependent variable.

2.2 Time series analysis

Time series analysis entails analysis of a sequence of measurements made at specified time intervals. The two main goals of time series analysis [14]:

- Identifying the nature of the phenomenon represented by the sequence of observations,
- Forecasting i.e. predicting future values of the time series variable

Both of these goals require that the pattern of the observed time series data be identified and more or less formally described. Interpretation and integration can be made once a pattern has been identified with the other data i.e., use it in the theory of the investigated phenomenon, e.g., seasonal death rates. Extrapolation of the identified pattern can be made as to predict future events.
2.2.1 Smoothing

Smoothing method uses the averaging methods and exponential smoothing methods to reveal clearly the underlying trend, seasonal and cyclic components from data sets. The methodology behind smoothing methods is taking the averages of the data sets and analyzing the data using the following:

- the mean squared errors
- sum of the squared errors
- mean of the squared errors

The above mentioned can be derived from a built forecasting model that when analysed, the ‘Flaw of averages’ is sometimes checked for in the model. What the ‘Flaw of averages’ states is that plans based on the assumption that average conditions will occur are usually wrong.

2.3 Multiple regression analysis

The basic purpose of a statistical investigation is to make predictions and optimise by determining what values of the independent variable/s is the dependent value a maximum or minimum. [5] Multiple regression analysis is a flexible method of data analysis that is appropriate whenever a quantitative (response) variable is to be examined in relationship to any predictor variable and regression analysis is most widely used for such functions. The basic regression formula is as follows;

\[ y = \beta_0 + \beta_1 \, x_1 + \cdots + \beta_p \, x_p + \epsilon \]  \hspace{1cm} (2.1)

Where

- \( y \) = dependent variable
- \( x \) = independent variable
- \( E(y) \) = deterministic component
- \( \epsilon \) = random error component
- \( \beta_0 \) = \( y \)-intercept of the line, i.e. point at which the line cuts through the \( y \)-axis
- \( \beta_1 \) = slope of the line, i.e. amount of increase in the mean of \( y \) for every 1-unit increase in \( x \).
2.3.1 Least squares estimation

The regression coefficients, \( \ldots \), are estimated using the method of ordinary least squares of which the coefficients minimises the error sum of squares.

This means the distance between the fitted line which consists of the predicted values and the actual Y values denoted by asterisks in figure 2.1 is being minimised. This distance is referred to as the residual or prediction error.

For a test to be accurate, certain assumptions have to be satisfied [1]

- data is sampled randomly and independently from the population
- and the deviations of the predicted values from the actual Y values are normally distributed with equal variance for all predicted values of

![Figure 2.1: Least squares estimation](image)

2.3.2 Partial ordinary least squares analysis

Randall D. Tobias [12] defines partial least square as a method for constructing predictive models when the factors are many and highly collinear. Essentially what partial least square does is show how much a predictor improves over all the other predictors. If it is added over all the predictors, ‘Pick six’, it will be defined as the increase in the regression sum of squares. The regression sum of squares is a variation analysis of Y that shows the variation of Y predicted by the predictor variable/s.
The total variation is measured by the following formula;

\[ Total \ sum \ of \ squares = \sum_{i}^{n} (Y_i - \bar{Y})^2 \]  \hspace{1cm} (2.3)

And the variation that can be predicted is measured as follows;

\[ Regression \ sum \ of \ squares = \sum_{i}^{n} (\hat{Y}_i - \bar{Y})^2 \]  \hspace{1cm} (2.4)

The residual sum of squares is measured as follows;

\[ Residual \ error \ sum \ of \ squares = \sum_{i}^{n} (Y_i - \hat{Y}_i)^2 \]  \hspace{1cm} (2.5)

Emphasis is put on the fact that the predictive models are used for predicting the responses and also on trying to understand the underlying relationship between the explanatory variables.

2.3.3 Sequential ordinary least square analysis

Sequential least square analysis shows how parameter predictions are improved as more predictor variables are added in a defined order. The same concept of the increase in sum of squares as explained under partial least square analysis is the same for sequential least square analysis and equations 2.3, 2.4 and 2.5 were used. It is beneficial to use sequential least square analysis if there is a natural order of the predictor variables.

From figure 1.3, the only order that can be distinguished is that the load time has to occur before the away time. For the sumps KPI’s, the first sump late has to occur before the last sump early and all the other KPI’s occur in between the two. This will not have any effect on the integrity of the model if not followed as the KPI measurements are deemed independent. Downtime can occur without any of the KPI’s could occur such as the loading of coal onto a shuttle car.
2.4 Principal component analysis

Principal component analysis (PCA) is a widely used multivariate method that is seen and described as a “complementary statistical” method used to run rough preliminary investigations, to sort out ideas, to put a new light on problem, or to point out aspects which would not come out in a classical approach.”[9] The main objectives of PCA that are in line with the problem at hand are; dimensionality reduction, feature selection i.e. the selection of the most useful variables, identification of underlying variables. [9]

In terms of analysing large data subsets, PCA is a great tool due to its condensing of information. If all mines KPI’s are to be analysed for trends, it will aid in enhancing the chances of producing useful results.

2.5 Linear programming

A mathematical model can serve a number of purposes. The first function is that a mathematical model can serve in regards to production as a tool for prediction. The types of models that can be used are the descriptive mathematical models and simulation models. A descriptive model can be used to predict the consequences of increasing a KPI value, that is, the effect on coal production tonnes. With a simulation model, it can be used to predict the outcome of prescribed set of strategies so as to provide needed input-output information to solve a decision problem.

Secondly, it can be used as a tool for control or decision-making purposes. Decision problems regarding controllable aspects of a process influence the operation of the process by changing or controlling the value of some decision variable. The type of decision models are modeled as to find the values of the decision variables that

• satisfy all constraints simultaneously
• achieve the stated objective

The methodology followed in generating a mathematical model fits perfectly with the methodology planned for developing the solution strategy for the KPI identification process. However, problems arise when generating the algorithm as known KPI production tonnes relationships will be used thus the productive KPI’s that have a direct relationship to the production tonnes will be favored.
For an example, the first sump late, last sump early and downtime KPI’s are merely subtracted from the face time unlike the other KPI’s that are included in the production time calculations used to calculate the actual coal produced in a shift/day.

**2.6 Excel model**

The modern trend in programming applications is graphical interfaces. This means that the user interacts with a program through a graphical interface e.g. values are entered in text boxes, options are selected from graphical menus and answers are presented graphically. Excel macros are an option in terms of their ability to create options that can be selected from a button etc. The macros are stored as VBA code that uses an ActiveX interface to cause excel’s applications to perform actions such as change formulas, create charts etc. The type of action that the application supports are defined by what is called an object model.

The model was built by excel designers to provide an interface so that the programming language can cause the application to do what a user normally would do interactively with a mouse and keyboard. It includes the following:

- list of application objects that can be managed
- properties of these objects that can be examined or altered; in the coal production model the value of the KPI’s can be altered.
- methods that can be performed on the object

An excel model can be embedded with calculations and measured values to output the desired end goals. Using this tool all the requirements can be met using defined formulations with the use of macros to show each KPI effect on the total coal tonnes. Each KPI factor can be calculated which can easily be viewed on the excel spreadsheet. Further formulation can be made using the KPI factors to select a combination of KPI that are critical in order to reach budgeted coal production tones.
2.7 Conclusion

From the different techniques studied, it was deemed multiple regression analysis is the most appropriate as it can yield the accurate required solution; however multiple regression has been extensively used in medical fields than engineering fields when dealing with variable selection studies. The other methods give rise to more problems as the solution formulation gets complicated due to extensive computations that defeats the purpose of the study. Due to the availability of software’s that can perform statistical computations, Statistical Analysis Software (SAS) was the prominent in terms of usage and simplicity understanding. The partial ordinary least square regression analysis methods will be used as it is considered to yield accurate solution outcomes.
Part III

Multiple regression analysis

Multiple regression analysis is widely used in the psychology, dietary, population and financial fields for prediction purposes. The relationship between the predictor variables and response variable will be analysed as to determine which predictor variable/s has the most impact on the response variable.

The following are the model assumptions as explained by Mendenhall for a regression model [7]

Assumption 1: The mean of the probability distribution of \( \epsilon \) is 0 i.e. the average of the errors over an infinitely long series of experiments is 0 for each setting of the independent variable \( x \).

Assumption 2: The variance of the probability distribution of \( \epsilon \) is constant for all settings of the independent variable \( x \).

Assumption 3: The probability of \( \epsilon \) is normal

Assumption 4: The errors associated with any two different observations are independent i.e. the error associated with one \( y \) has no effect on the errors associated with other \( y \) values.

A test for residual normality is crucial before any inferences can be made from the model. The p-value [table 3.1] can only be used if the residual normality test is satisfied, if not transformations are made and explained before using the p-value for parameter selection.

Test for normality

\[ H_0 : \text{Residual come from a normal distribution} \]
\[ H_1 : \text{Residual deviate from a normal distribution} \]

Shapiro-Wilk test

\( n \leq 50 \)

Reject \( H_0 \) if \( W \) is small

\[ p\text{-value} = p(w) < 0.05 : \text{deviate from normal} \]
\[ >0.05: H_0 \text{ is not rejected} \]
A Shapiro-Wilk test is used only when the number of observation is less or equal to 50. The rule of thumb is used for the confidence level i.e. $\alpha = 0.05$ unless stated otherwise due to failure to comply with normality tests. Data transformations are performed on the $y$ values as to make them nearly satisfy the assumptions, and for the latter reason, to achieve a model that provides a better approximation to $E(y)$.

### 3.1 Terminology

The basic terminology presented in part II; section 2.2 will be used with the additional terminology explained in table 3.1 throughout the rest of the document.

**Table 3.1: Regression terminology**

| **Least square:** | Squared distance between the data points and the model.  
The aim is to minimise the distance as to get a best fit linear graph. |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R-square:</strong></td>
<td>It measures how much of $y$, response variable is explained by $x$, predictor variable</td>
</tr>
<tr>
<td><strong>P–value:</strong></td>
<td>In statistical hypothesis testing, the p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. The fact that p-values are based on this assumption is crucial to their correct interpretation.</td>
</tr>
</tbody>
</table>

### 3.2 General structure

The steps applicable for developing a regression model are as listed below. Collection of data and variable classification was done in Step 1 and 2. From the code written for the model with the required output formulation for step 3, 4 and 5, the computations were done simultaneously. Step 6 calculations are for validation. The model was stopped when the t-test was not violated i.e. the p-value was less than 0.005.
Regression Analysis steps

**Step 1**: Collection of data for each experimental unit in the sample

**Step 2**: Variable classification

**Step 3**: Estimation of the unknown parameters, $\beta_0, \beta_1, \ldots, \beta_6$.

**Step 4**: Estimation of the probability distribution of the random error component $\epsilon$ and its variance

**Step 5**: Evaluation of the utility of the model

**Step 6**: Validation of assumptions on $\delta$ and make model modifications if necessary

**Step 7**: If the model is deemed adequate, estimate the mean value of $Y$, identify the critical KPI and make other inferences.

Regression algorithm

\[
Y = \beta_0 + \sum_{i=1}^{6} \beta_i X_i + \epsilon_i \quad i = \{1, \ldots, 6\} \quad (3.1)
\]

In this formulation, the dependent variable is denoted by $Y$, which is the production tonnes and the independent variables by $X_i$ which are the ‘Pick six’ as described in table 3.1.

Table 3.2 ‘Pick six’ model symbols

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>first sump late</td>
</tr>
<tr>
<td>$X_2$</td>
<td>load time</td>
</tr>
<tr>
<td>$X_3$</td>
<td>away time</td>
</tr>
<tr>
<td>$X_4$</td>
<td>tram time per meter cut</td>
</tr>
<tr>
<td>$X_5$</td>
<td>Downtime</td>
</tr>
<tr>
<td>$X_6$</td>
<td>last sump early</td>
</tr>
</tbody>
</table>
SAS algorithm

The model explained here is for one mine section, a couple of other models were run as to accomplish a general KPI combination that can be used for all Anglo Coal underground mining processes. (Addendum A and B)

Input data: Simunye Section

```sas
options nodate pageno=1;
title 'Vlaklaagte: Simunye Section';
data vlakkies;
labeled y = 'Actual Production tonnes'
   x1 = 'Load time sec'
   x2 = 'Away time sec'
   x3 = 'FS late min'
   x4 = 'Tram time min'
   x5 = 'Downtime min'
   x6 = 'LS early min';
input y x1 x2 x3 x4 x5 x6;
datalines;
  59457 63 72 22 3 304 13
  54847 67 70 26 3 187 13
  51350 67 74 25 4 155 16
  60042 67 76 24 4 142 20
  60898 66 71 20 4 143 8
  51956 70 72 25 3 148 7
  46085 67 70 25 3 203 5
  45962 63 84 28 4 199 8
  58854 67 71 27 3 167 6
  52279 88 68 22 3 187 12
  56704 66 66 24 3 201 19
  68864 63 70 20 3 215 11
;```
Data points for each mine section were extracted from an excel spreadsheet and defined as variables as in line 4(input) of the model algorithm. For this algorithm Simunye’s data was tested using 12 data points.

**Iteration & Output selection**

```plaintext
proc reg data=vlakkies;
model y = x1 x2 x3 x4 x5 x6;
output out = a1 p = pred1 r = res1 student = studres1;

model y = x1 x3 x4 x5 x6;
model y = x1 x3 x4 x5;
model y = x1 x3 x4;
model y = x1 x3;
run;
```

The full model was ran first and further parameter selection steps were made as to obtain parameter estimates that were significant. For this model, $x_2$ was eliminated first, followed by $x_6$, $x_5$ then $x_4$ using the outputted p-value. If the p-value for any predictor variable was significantly large, the variable was eliminated from the model. This was done till the remaining p-value’s for each predictor variable fell within the 95% confidence interval.

**Check if errors are normally distributed**

```plaintext
goptions reset=all;
symbol i=none value=dot cv=green height=0.7;

proc gplot data=a1;
plot studres1*pred1 / vref=0;
run;
```

In order to use the t-test for the stepwise elimination process, errors have to be normally distributed with their mean equal to zero as mentioned earlier.
Check for correlation between the x variables and y

```
proc corr data=vlakkies;
var x1 x2 x3 x4 x5 x6;
with y;
run;
```

This checks for how y and x’s are related i.e. for example if there is a negative relationship between y and one of the x values, take for instance the away time, the expected actual production tonnes will decrease as the away time increases.

Check for correlation between the x variables

```
proc corr data=vlakkies;
var x1 x2 x3 x4 x5 x6;
run;
```

This was used to check for relationships between the x’s themselves. If highly correlated predictor variables are present in the model, the results may be confusing. In some cases the t test on the individual parameter estimates may be non-significant even though the F-test for the overall model adequacy is significant and the parameter estimates may have signs opposite from what is expected. Variation inflation factors aid in determining whether a serious multicollinearity problem exists.

Data Normality Test

```
proc univariate data=vlakkies normal;
var y x1 x2 x3 x4 x5 x6;
histogram y x1 x2 x3 x4 x5 x6 /normal;
run;
```

The data normality test was to check if some KPI’s data were normally distributed. The sample size was too small, i.e. less than 50 datasets thus it was infrequent for normality to be observed in the datasets.
Fitted model for significant KPI’s

```plaintext
proc reg data=vlakkies;
model y = x1 x3;
output out=a p=pred r=res student=studres;
run;

proc print data=a;
run;

goptions reset=all;
symbol i=none value=dot cv=red height=0.7;

proc gplot data=a;
plot studres*pred / vref=0;
title3 'Studentized residual of the final model';
run;
```

This ensures that the errors are normally distributed as to satisfy the error normality assumptions mentioned before. From the graph it can be seen that some of the errors do come from a normal distribution or are normally distributed. (Refer to figure 3.1)

### 3.3 Statistical significance

A number of the models output will answer and show how reliable the solution strategy is. The following are some of the important outputs that were studied;

**R²**: It provides a measure of how well Y can be predicted from the set of x scores i.e. the amount of the total variation in Y that can be explained by x.

**Adjusted R²**: Is the same as the R² however it takes into account the number of parameters being modeled, x. A high R² portrays the reliability of a model in terms of how much of the dependent variable is explained by the independent variables.

**Hypothesis testing**: This aid in testing the validity of the estimated value of a parameter and the utility of the model. By the rule of thumb, the significance level was defined to be 5% but for the Vlaaklagte, Simunye Section, 20% was used only for testing the
parameters. If parameter estimates fell within the significance level, the null hypothesis was rejected, stating that the parameters were significant from zero and if not, stepwise regression was used to eliminate the parameters that were not significant. The model was run until we had only significant parameters in the model. This can be viewed in the iteration and output step of the SAS algorithm. \( X_6 \) was eliminated first followed by \( x_5 \) and \( x_4 \) consecutively.

**F-test:** The test is done on the full model using the F test for all the predictor variables except for the intercept values as they have no significance on the solution strategy. The null hypothesis can be rejected if the p-value is less than the \( \alpha \) value.

Test for overall model significance

\[
\begin{align*}
H_0 & : \text{Model not significant} \\
H_1 & : \text{Model significant} \\
\alpha & = 0.05 \\
\text{Reject } H_0 & \text{ if p-value is less than 0.05}
\end{align*}
\]

**T-test:** Same test as the F-test but only done for parameter estimates.

Test for parameter significance

\[
\begin{align*}
H_0 & : \beta_i = 0 \text{(Not significantly different from zero)} \\
H_1 & : \beta_i \neq 0 \text{ (significantly different from zero)} \\
\alpha & = 0.05^* \\
\text{Reject } H_0 & \text{ if p-value is less than 0.05}
\end{align*}
\]

\* Only 20% significance level was used for one (Vlaaklagte: Simunye Section) model as to accommodate the actuality of not having sufficient data points.

The model has to be run in such a way that all the regression analysis assumptions are not violated using the normality tests for errors. In order to use the p and f-test, normality for errors must not be violated as the tests can only be used if the assumptions are met. If normality is violated, then the model did not pass to merit for further considerations.
**Studentised residual graph of the final model:** As explained before, residuals are values between the observed responses from the predicted responses. The studentised residual graph plots the scatter points within the 95% confidence level. In order for the model not to be rejected, the scatter points should not show any trend.

![Studentised residual graph of the final model](image)

Figure 3.1 Studentised residual graph of the final model: Vlaaklagte, Simunye Section
Part IV

Computational Results

4.1 Model outputs

Out of the five mine sections models, 5 x 6 parameter estimate outputs were studied as to identify the critical KPI's. As mentioned before the parameter estimate, $\beta_i$, is the amount of increase or decrease in the mean of $y$ for every 1-unit increase in $x$. Monthly data was used for all the models except for Greenside as it had weekly data. The five model outputs give explanation as per data set inputted in each mine section’s model and the output is demonstrated using the general regression equation (4.1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \epsilon$$  \hspace{1cm} (4.1)

Vlaaklagte

Simunye Section

$$y = 144556 - 508.425 x_1 + 666.005 x_2 - 2415.144 x_3 - 10749 x_4 - 75.699 x_5 + 485.44 x_6$$

Mangwapa Section

$$y = 288557 - 1173 x_1 - 595.603 x_2 - 87.78 x_3 - 224.161 x_4 + 7104.161 x_5 - 435.534 x_6$$

The parameter estimate for Simunye Section’s $x_1$ which is the load time can be interpreted as for every one unit increase in the load time, the coal production tonnes will decrease by 508.425 units and for $x_6$ which is the last sump early can be interpreted as for every one unit increase in the last sump early, the coal production tonnes will increase by 485.44 units.

The parameter estimates for Vlaaklagte KPI’s cannot be used to derive a general set of combinations as the estimates differ tremendously. The $R^2$ for the Simunye Section model was 73% and for Mangwapa Section model, 78% which is significantly high, however the significant parameter estimates for each mine differ.
Greenside

George Section

\[ y = 2367.638 - 1.526x_1 - 8.336x_2 + 14.837x_3 - 163.727x_4 - 2.462x_5 - 6.600x_6 \]

Vumangara Section

\[ y = 5948.492 + 5.610x_1 - 36.644x_2 - 15.223x_3 - 228.054x_4 - 0.454x_5 - 3.836x_6 \]

The parameter estimates for Greenside differ slightly as compared to the Vlaaklagte parameter estimates, however the \( R^2 \) for both models were significantly low, 45% and 48% for George and Vumagara respectively. If the model is not good enough, no KPI combination can be derived from such a model.

New Denmard Central

Simunye Section

\[ y = 74521 - 103.3x_1 - 9.837x_2 - 318.546x_3 - 5886.84x_4 + 41.528x_5 - 365.107x_6 \]

The \( R^2 \) for the New Denmark section was 55% and the adjusted \( R^2 \) has a negative value which shows that the model contains terms that do not help in predicting the response.

All the parameter estimates where tested for significance and the significant parameter estimates are as shown in the table 4.1.

Table 4.1: Significant parameter estimates

<table>
<thead>
<tr>
<th>Section Name</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \beta_5 )</th>
<th>( \beta_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Greenside: Vumagara</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2  NDC_Central :Simunye</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3  Vlaaklagte: Simunye</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4  Greenside: George</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5  Vlaaklgate: Mangwapa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
From the table, no parameter estimate(s) dominate the others. This can be due to the reality that all the mines differ due to the following:

- shift patterns
- maintenance allowances
- pillar centers
- size of coal
- haulers
- daily mine’s section production targets etc.

From the five models, only three KPI’s dominated and this cannot be deemed legitimate as the sample size of the models is minute. It should not be ignored that the five models did not yield favorable outputs thus impeding further investigation due to the sample size being too small &/other unknown factors.

4.2 Solution quality

Test for Model and parameter significance

As mentioned before, the t-test and F-test were used for checking model and parameter significance. The p-value had to be below 20% were measurable changes were made, and 5% for all the other models using the rule of thumb.
Part IV

Recommendations

The model outputs did not show any trend that could be used to select a general KPI combination. The data samples when tested using the normality test did not yield favorable results thus suggesting the small sample size might be a problem. The parameter estimates told a different story for each mine section even though some of the models were significant and had more or less the same data points as the others.

However the model outputs gave ample information on the effect they have on the actual production tonnes. As mentioned under future work, further regression analysis methodologies can be carried out as to improve the solution integrity with the addition of a bigger sample size.

5.1 Future work

One area that was not explored was the combination of some KPI’s by having a product of two KPI’s as to check if the KPI’s combined will have an impact on the actual production tonnes. What this would do is to increase the order of the regression model to two or three degrees but the model would still be categorised as a linear model. The same goes for having reciprocals or dividing some of the KPI’s.

It would be valuable to gather more data as to develop precise and bettered underground mining KPI regression models that can be used for KPI combination selection and coal production tonnes forecasting.

Data transformations can also aid in giving a better model fit and smaller prediction errors, but also give a more reasonable bias structure as to how the KPI combination output would fit a real life mining scenario. The type of transformation will depend on the theoretical relationships between the response variable and the predictor variables.

One important lesson learnt is that one should not trust global measures of prediction quality too much since they may be misleading. The values for all the models developed are very reasonable, but the general interpretation of the individual output seemed bias.
Bibliography


Addendum

Addendum A: Sections SAS models
A 1: Vlaaklagte: Mangwapa section
A 2: Vlaaklagte: Mangwapa Section
A 3: Greenside: George Section
A 4: Greenside: Vumagara Section
A 5: New Denmark Colliery: Simunye Section

Addendum B: SAS sections model outputs
B 1: Vlaaklagte: Simunye Section
B 2: Vlaaklagte: Mangwapa Section
B 3: Greenside: George Section
B 4: Greenside: Vumagara Section
B 5: New Denmark Colliery: Simunye Section
Addendum A: Sections SAS models
A 1: Vlaaklagte: Simunye Section

```plaintext
options nodate pageno=1;
title' Vlaaklagte: Simunye Section';

data vlakkies;
label y = 'Actual Production tonnes'
x1 = 'Load time sec'
x2 = 'Away time sec'
x3 = 'FS late min'
x4 = 'Tram time min'
x5 = 'Downtime min'
x6 = 'LS early min';

input y x1 x2 x3 x4 x5 x6;
datalines;
59457 63 72 22 3 304 13
54847 67 70 26 3 187 13
51350 67 74 25 4 155 16
60042 67 76 24 4 142 20
60988 66 71 20 4 143 8
51956 70 72 25 3 148 7
46085 67 70 25 3 203 5
45962 67 72 28 4 199 8
58854 67 71 27 3 167 6
52279 88 68 22 3 187 12
56704 66 66 24 3 201 19
68864 63 70 20 3 215 11
;

proc reg data=vlakkies;
model y = x1 x2 x3 x4 x5 x6;
output out = a1 p = pred1 r = res1 student = studres1;
model y = x1 x3 x4 x5 x6;
model y = x1 x3 x4 x5;
model y = x1 x3 x4;
model y = x1 x3;
run;

options reset=all;
symbol i=none value=dot cv=green height=0.7;

proc gplot data=a1;
plot studres1*pred1 / vref=0;
run;

proc corr data=vlakkies;
var x1 x2 x3 x4 x5 x6;
with y;
run;

proc corr data=vlakkies;
```
var x1 x2 x3 x4 x5 x6;
run;

proc univariate data=vlakkies normal;
var y x1 x2 x3 x4 x5 x6;
histogram y x1 x2 x3 x4 x5 x6 /normal;
run;

**********Checking errors for normality for the fitted model**********;

proc reg data=vlakkies;
model y = x1 x3 ;
output out=a p=pred r=res student=studres;
run;

proc print data=a;
run;

goptions reset=all;
symbol i=none value=dot cv=red height=0.7;

proc gplot data=a;
plot student*pred / vref=0;
title3 'Studentized residual of the final model';
run;
A 2: Vlaaklagte- Mangwapa section

```plaintext
options nodate pageno=1;
title 'Vlaaklagte: Mangwapa Section';

data vlakkies1;
label y = 'Actual Production tonnes'
x1 = 'Load time sec'
x2 = 'Away time sec'
x3 = 'FS late min'
x4 = 'Tram time min'
x5 = 'Downtime min'
x6 = 'LS early min';

input y x1 x2 x3 x4 x5 x6;
datalines;
90257 67 37 517 1 2 155 3
83561 70 44 474 8 2 147 12
63425 69 47 467 18 3 224 16
77906 64 49 477 22 3 174 17
95465 58 53 485 11 2 139 21
10045956 52 500 9 2 165 9
97995 54 44 552 10 2 157 14
64849 55 48 524 14 2 208 16
93748 57 47 523 13 2 134 16
89527 57 48 543 12 2 141 10
47986 58 45 562 4 2 229 17
69023 58 43 570 12 2 201 25
95342 59 45 553 12 2 175 16
;
proc reg data=vlakkies1;
model y = x1 x2 x3 x4 x5 x6;
output out = a1 p = pred1 r = res1 student = studres1;
model y = x1 x2 x3 x5 x6;
model y = x1 x2 x3 x6;
model y = x1 x3 x6;
model y = x1 x6;
model y = x6;
run;

goptions reset=all;
symbol i=none value=dot cv=green height=0.7;

proc gplot data=a1;
plot studres1*pred1 / vref=0;
title3 'Studentized residual of the full model';
run;

proc corr data=vlakkies1
var x1 x2 x3 x4 x5 x6;
with y;
run;

proc corr data=vlakkies1;
var x1 x2 x3 x4 x5 x6;
run;
```
**proc univariate data=vlakkies1 normal;**
var y x1 x2 x3 x4 x5 x6;
histogram y x1 x2 x3 x4 x5 x6 / normal;
run;

**********Checking errors for normality for the fitted model**********;

**proc reg data=vlakkies1;**
model y = x6 ;
output out=a p=pred r=res student=studres;
run;

**proc print data=a;**
run;

goptions reset=all;
symbol i=none value=dot cv=red height=0.7;

**proc gplot data=a;**
plot studres*pred / vref=0;
title3 'Studentized residual of the final model';
run;
options nodate pageno=1;
title 'Greenside: George Section';

data greenside1;
label y = 'Actual Production tonnes'
   x1 = 'FS late min'
   x2 = 'Load time sec'
   x3 = 'Away time sec'
   x4 = 'Tram time min'
   x5 = 'Downtime min'
   x6 = 'LS early min';

input x1 x2 x3 x4 x5 x6 y;
datalines;
15 50 63 3 175 30 2104
17 66 57 2 163 16 1857
4 60 65 2 125 45 1903
13 69 64 2 138 32 1886
10 64 62 3 208 26 1382
25 56 62 3 69 21 1850
15 60 65 2 95 10 2563
10 60 61 2 106 26 2082
18 59 64 3 50 13 2299
27 60 56 3 59 11 2206
13 61 58 2 52 11 1894
19 62 62 2 199 11 1782
23 61 68 2 105 10 2003
13 68 65 2 98 22 2249
9 65 58 3 154 19 1815
5 64 64 3 142 24 1355
1 67 61 3 126 27 1722
15 69 64 3 159 21 1704
17 42 59 3 226 24 1500
9 67 58 3 50 23 2276
22 66 65 3 118 16 2138
14 64 64 3 138 20 1944
30 61 57 3 20 19 1376
2 64 64 3 81 17 1886
5 74 64 3 203 17 1368
2 71 60 3 330 22 1314
10 71 62 3 195 17 1703
3 70 62 2 138 19 1643
15 72 56 3 115 20 1761
15 71 61 3 178 21 1565
12 65 68 3 155 22 1593
;

proc reg data=greenside1;
model y = x1 x2 x3 x4 x5 x6 ;
output out = a r = res p = pred student = studres;

model y = x2 x3 x4 x5 x6 ;
model y = x2 x3 x5 x6 ;
model y = x2 x3 x5 ;
model y = x3 x5 ;
model y = x5 ;
run;
goptions reset=all;
symbol i=none value=dot cv=blue height=0.7;

proc gplot data = a;
plot studres*pred / vref = 0;
run;

proc corr data=greenside1;
var x1 x2 x3 x4 x5 x6;
with y;
run;

proc corr data=greenside1;
var x1 x2 x3 x4 x5 x6;
run;

proc univariate data=greenside1 normal;
var y x1 x2 x3 x4 x5 x6;
histogram y x1 x2 x3 x4 x5 x6 /normal;
run;

**********Checking errors for normality for the fitted model*************;

proc reg data = greenside1;
model y = x5;
output out = a1 r = res1 p = pred1 student = studres1;
title1 'Fitted regression model';
run;

goptions reset=all;
symbol i=none value=dot cv=red height=0.7;

proc gplot data = a1;
plot studres1*pred1 / vref=0;
title3 'Studentized residual of the final model';
run;
### A 4: Greenside: Vumagara Section

```plaintext
options nodate pageno=1;
title 'Greenside: Vumagara Section';

data greenside;
label y = 'Actual Production tonnes'
   x1 = 'FS late min'
   x2 = 'Load time sec'
   x3 = 'Away time sec'
   x4 = 'Tram time min'
   x5 = 'Downtime min'
   x6 = 'LS early min';

input x1 x2 x3 x4 x5 x6 y;
datalines;
16  61  88  2.5  221  15  1539
  5  61  99  2.5  202  16  1652
  9  63  98  2.5  169  18  1605
  9  66  70  2.5  183  14  1377
  9  70  68  2.5  172  21  1895
 15  71  68  2.5  228  19  1544
 10  70  70  2.5  158  22  1887
 16  70  59  2.5  174  7  1997
 13  68  67  2.5  273  23  1674
  9  69  67  2.5  141  23  1650
  5  71  67  2.5  176  19  1686
  9  70  64  2.5  42   16  1405
 23  69  68  2.5  159  50  1494
 11  64  65  2.5  163  28  1741
  6  68  65  2.5  165  18  1635
  4  68  60  2.5  105  4  2170
  2  66  62  2.5  169  10  1977
  4  62  57  2.5  101  25  2079
 37  60  58  2.5  166  28  1908
 32  65  60  2  143  14  1955
  3  57  62  2  144  45  2295
  5  55  65  2  126  71  2087
 10  64  57  2  188  19  1902
 15  56  65  3  149  47  2116
 26  69  61  3  222  14  1680
 19  67  58  3  76   22  2434
 16  63  58  3  188  23  2117
 27  62  63  2  103  30  3212
 20  59  63  2  183  22  2421
 13  59  68  2  135  26  2078
 16  59  61  2  129  31  2455
 11  61  62  2  145  26  2112
 10  63  59  3  240  25  2137
  0  64  67  3  235  27  1600
  8  64  68  3  130  20  1915
  3  67  63  3  215  24  1770
  2  68  55  3  220  22  1688
  7  65  60  3  260  30  1827
  6  61  63  3  40   25  1566
 12  60  61  2  223  18  2271
  4  62  63  2  141  18  2229
 16  64  67  2  99   19  1931
  2  66  63  2  134  26  1811
```
```
```plaintext
6  69  69  2  138  18  1844  
6  66  64  2  183  23  2171  
26  67  60  3  153  11  1898  
3  64  57  2  173  19  2169  
9  61  58  3  134  19  2012  
15  64  60  2  214  20  2290  
14  64  60  3  131  20  2068  

proc reg data=greenside;
model y = x1 x2 x3 x4 x5 x6 ;
output out = a r = res p = pred student = studres;
model y = x1 x2 x3 x4 x6 ;
model y = x1 x2 x3 x4 ;
model y = x2 x3 x4 ;
run;

goptions reset=all;
symbol i=none value=dot cv=green height=0.7;
proc gplot data = a;
plot studres*pred / vref = 0;
run;

proc corr data=greenside;
var x1 x2 x3 x4 x5 x6;
with y;
run;

proc corr data=greenside;
var x1 x2 x3 x4 x5 x6;
run;

proc univariate data=greenside normal;
var y x1 x2 x3 x4 x5 x6;
histogram y x1 x2 x3 x4 x5 x6 /normal;
run;

*******Checking errors for normality for the fitted model************;
proc reg data = greenside;
model y = x2 x3 x4 ;
output out = a1 r = res1 p = pred1 student = studres1;
title1 'Fitted regression model';
run;

goptions reset=all;
symbol i=none value=dot cv=red height=0.7;
proc gplot data = a1;
plot studres1*pred1 / vref=0;
title3 'Studentized residual of the final model';
run;
```
A 5: New Denmark Colliery - Simunye Section

options nodate pageno=1;
title 'NDC_Central: Simunye Section';

data ndc;
label y = 'Actual Production tonnes'
x1 = 'Load time sec'
x2 = 'Away time sec'
x3 = 'FS late min'
x4 = 'Tram time min'
x5 = 'Downtime min'
x6 = 'LS early min';
input y x1 x2 x3 x4 x5 x6;
datalines;
25628  96   90  26   5  267   26
25873  93   89  24   6  222   21
24785  88   95  13   7  236   8
24747 111   75   8   7  215   8
26208 114   76  21   5  232  24
28506 113   81  26   4  214  21
27930 115   79  21   6  260   9
22736 138   93  15   7  243   4
20597 116   85  14   6  159   5
36238 113  84  15   5  200   6
35034 113  84  18   5  173  13
;
proc reg data=ndc;
model y = x1 x2 x3 x4 x5 x6;
output out = a r = res p = pred student = studres;
model y = x1 x3 x4 x5 x6;
model y = x1 x4 x5 x6 ;
model y = x1 x4 x6;
model y = x4 x6 ;
run;

goptions reset=all;
symbol i=none value=dot cv=green height=0.7;
proc gplot data = a;
plot studres*pred / vref = 0;
run;
proc corr data=ndc;
var x1 x2 x3 x4 x5 x6;
with y;
run;
proc corr data=ndc;
var x1 x2 x3 x4 x5 x6;
run;
proc univariate data=ndc normal;
  var y x1 x2 x3 x4 x5 x6;
  histogram y x1 x2 x3 x4 x5 x6 /normal;
run;

*******Checking errors for normality for the fitted model***********;

proc reg data=ndc;
  model y = x4 x6;
  output out = a1 p = pred1 r = res1 student = studres1;
run;

proc print data = a1;
run;

goptions reset=all;
symbol i=none value=dot cv=red height=0.7;

proc gplot data = a1;
  plot studres1*pred1 / vref=0;
  title3 'Studentized residual of the final model';
run;
Addendum B: SAS sections model outputs

*Only one model output is included under this section, the rest of the output models are saved in the attached disc*
B 1: Vlaaklagte-Simunye Section

Vlaaklagte: Simunye Section

The REG Procedure
Model: MODEL1
Dependent Variable: y Actual Production tonnes

Number of Observations Read 12
Number of Observations Used 12

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>346946314</td>
<td>57824386</td>
<td>2.23</td>
<td>0.1987</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>129834085</td>
<td>25966817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11</td>
<td>476780400</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 5095.76462 R-Square 0.7277
Dependent Mean 55608 Adj R-Sq 0.4009
Coeff Var 9.16370

Parameter Estimates

| Variable     | Label          | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
|--------------|----------------|----|--------------------|----------------|---------|------|---|
| Intercept    | Intercept      | 1  | 144556             | 39939          | 3.62    | 0.0152 |
| x1           | Load time sec  | 1  | -508.42486         | 262.32092      | -1.94   | 0.1103 |
| x2           | Away time sec  | 1  | 666.00469          | 779.10867      | 0.85    | 0.4317 |
| x3           | FS late min    | 1  | -215.14412         | 891.85202      | -0.271  | 0.2024 |
| x4           | Tram time min  | 1  | -10749             | 7642.21477     | -1.41   | 0.2186 |
| x5           | Downtime min   | 1  | -75.69894          | 54.63028       | -1.39   | 0.2245 |
| x6           | LS early min   | 1  | 485.43998          | 368.77244      | 1.32    | 0.2452 |
Vlaklaagte: Simunye Section

The REG Procedure
Model: MODEL5
Dependent Variable: y Actual Production tonnes

Number of Observations Read 12
Number of Observations Used 12

Analysis of Variance

Source DF Sum of Squares Mean Square F Value Pr > F
Model 2 265536399 132768199 5.66 0.0257
Error 9 211244001 23471556
Corrected Total 11 476780400

Root MSE 4844.74516 R-Square 0.5569
Dependent Mean 55608 Adj R-Sq 0.4585
Coeff Var 8.71229

Parameter Estimates

| Variable | Label          | DF | Parameter Estimate | Error | t Value | Pr > |t| |
|----------|----------------|----|--------------------|-------|---------|------|
| Intercept| Intercept      | 1  | 122660             | 21851 | 5.61    | 0.0003 |
| x1       | Load time sec  | 1  | -338.95933         | 220.77709 | -1.54 | 0.1591 |
| x3       | FS late min    | 1  | -1835.79031        | 576.93002 | -3.18 | 0.0111 |

The CORR Procedure

1 With Variables: y
6 Variables: x1 x2 x3 x4 x5 x6

Simple Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Sum</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>12</td>
<td>55608</td>
<td>6584</td>
<td>667298</td>
<td>45962</td>
<td>68864</td>
</tr>
<tr>
<td>x1</td>
<td>12</td>
<td>67.83333</td>
<td>6.68558</td>
<td>814.00000</td>
<td>63.00000</td>
<td>88.00000</td>
</tr>
<tr>
<td>x2</td>
<td>12</td>
<td>72.00000</td>
<td>4.57265</td>
<td>864.00000</td>
<td>66.00000</td>
<td>84.00000</td>
</tr>
<tr>
<td>x3</td>
<td>12</td>
<td>24.00000</td>
<td>2.55841</td>
<td>288.00000</td>
<td>20.00000</td>
<td>28.00000</td>
</tr>
<tr>
<td>x4</td>
<td>12</td>
<td>3.33333</td>
<td>0.49237</td>
<td>40.00000</td>
<td>3.00000</td>
<td>4.00000</td>
</tr>
<tr>
<td>x5</td>
<td>12</td>
<td>187.58333</td>
<td>44.69179</td>
<td>2251</td>
<td>142.00000</td>
<td>304.00000</td>
</tr>
<tr>
<td>x6</td>
<td>12</td>
<td>11.50000</td>
<td>4.96360</td>
<td>138.00000</td>
<td>5.00000</td>
<td>20.00000</td>
</tr>
</tbody>
</table>

The UNIVARIATE Procedure

Variable: y (Actual Production tonnes)

Moments

13
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Sum Weights</th>
<th>Sum Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>55608.1667</td>
<td>55608.1667</td>
<td>667298</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>6583.59117</td>
<td>Variance 43343672.7</td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>0.28003329</td>
<td>Kurtosis 0.15128948</td>
<td></td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>3.7584E10</td>
<td>Corrected SS 476780400</td>
<td></td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>11.8392523</td>
<td>Std Error Mean 1900.51907</td>
<td></td>
</tr>
</tbody>
</table>
Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>55608.17</td>
</tr>
<tr>
<td>Median</td>
<td>55775.50</td>
</tr>
<tr>
<td>Mode</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests for Location: $\mu = 0$

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's $t$</td>
<td>$t = 29.25946$</td>
<td>$Pr &gt;</td>
</tr>
<tr>
<td>Sign</td>
<td>$M = 6$</td>
<td>$Pr \geq</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>$S = 39$</td>
<td>$Pr \geq</td>
</tr>
</tbody>
</table>

Tests for Normality

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk</td>
<td>$W = 0.956764$</td>
<td>$Pr &lt; W = 0.7368$</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>$D = 0.127513$</td>
<td>$Pr &gt; D &gt; 0.1500$</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>$W-Sq = 0.032057$</td>
<td>$Pr &gt; W-Sq &gt; 0.2500$</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>$A-Sq = 0.24477749$</td>
<td>$Pr &gt; A-Sq &gt; 0.2500$</td>
</tr>
</tbody>
</table>

The UNIVARIATE Procedure

Fitted Normal Distribution for $y$

Parameters for Normal Distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$\mu$</td>
<td>55608.17</td>
</tr>
<tr>
<td>Std Dev</td>
<td>$\sigma$</td>
<td>6583.591</td>
</tr>
</tbody>
</table>

Goodness-of-Fit Tests for Normal Distribution

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>$D = 0.12751317$</td>
<td>$Pr &gt; D &gt; 0.150$</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>$W-Sq = 0.03205699$</td>
<td>$Pr &gt; W-Sq &gt; 0.250$</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>$A-Sq = 0.24477749$</td>
<td>$Pr &gt; A-Sq &gt; 0.250$</td>
</tr>
</tbody>
</table>

The UNIVARIATE Procedure

Variable: $x1$ (Load time sec)

Moments

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>12</td>
</tr>
<tr>
<td>Mean</td>
<td>67.83333333</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>6.68557923</td>
</tr>
<tr>
<td>Skewness</td>
<td>2.86864841</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>55708</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>9.85589076</td>
</tr>
<tr>
<td>Sum Weights</td>
<td>12</td>
</tr>
<tr>
<td>Sum Observations</td>
<td>814</td>
</tr>
<tr>
<td>Variance</td>
<td>44.696697</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>9.14528156</td>
</tr>
<tr>
<td>Corrected SS</td>
<td>491.666667</td>
</tr>
<tr>
<td>Std Error Mean</td>
<td>1.92996049</td>
</tr>
</tbody>
</table>
### Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>67.83333</td>
</tr>
<tr>
<td>Median</td>
<td>67.00000</td>
</tr>
<tr>
<td>Mode</td>
<td>67.00000</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tests for Location: \( \mu_0 = 0 \)

<table>
<thead>
<tr>
<th>Test</th>
<th>-Statistic-</th>
<th>( \text{---p Value---} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's t</td>
<td>( t ) = 35.14752</td>
<td>( \text{Pr} &gt;</td>
</tr>
<tr>
<td>Sign</td>
<td>( M ) = 6</td>
<td>( \text{Pr} \geq</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>( S ) = 39</td>
<td>( \text{Pr} \geq</td>
</tr>
</tbody>
</table>

Tests for Normality

<table>
<thead>
<tr>
<th>Test</th>
<th>-Statistic---</th>
<th>( \text{---p Value---} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk</td>
<td>( W ) = 0.601942</td>
<td>( \text{Pr} &lt; W ) = 0.0001</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>( D ) = 0.382932</td>
<td>( \text{Pr} &gt; D ) = &lt;0.0100</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>( W-Sq ) = 0.352356</td>
<td>( \text{Pr} &gt; W-Sq ) = &lt;0.0050</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>( A-Sq ) = 1.91812</td>
<td>( \text{Pr} &gt; A-Sq ) = &lt;0.0050</td>
</tr>
</tbody>
</table>

The UNIVARIATE Procedure
Fitted Normal Distribution for \( x_1 \)

Parameters for Normal Distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>( \mu )</td>
<td>67.83333</td>
</tr>
<tr>
<td>Std Dev</td>
<td>( \sigma )</td>
<td>6.685579</td>
</tr>
</tbody>
</table>

Goodness-of-Fit Tests for Normal Distribution

<table>
<thead>
<tr>
<th>Test</th>
<th>-Statistic-----</th>
<th>( \text{---p Value---} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>( D ) = 0.38293158</td>
<td>( \text{Pr} &gt; D ) = &lt;0.010</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>( W-Sq ) = 0.35235564</td>
<td>( \text{Pr} &gt; W-Sq ) = &lt;0.005</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>( A-Sq ) = 1.91812039</td>
<td>( \text{Pr} &gt; A-Sq ) = &lt;0.005</td>
</tr>
</tbody>
</table>

The UNIVARIATE Procedure
Variable: \( x_2 \) (Away time sec)

Moments

<table>
<thead>
<tr>
<th>N</th>
<th>Sum Weights</th>
<th>Sum Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td>864</td>
</tr>
<tr>
<td>Mean</td>
<td>72</td>
<td></td>
</tr>
</tbody>
</table>

- Std Deviation: 4.57264594
- Variance: 20.909090
- Skewness: 1.70465475
- Kurtosis: 4.11595463
- Uncorrected SS: 62438
- Corrected SS: 230
- Coeff Variation: 6.35089714
- Std Error Mean: 1.32000918

Basic Statistical Measures

Location Variability

17
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Std Deviation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>72.00000</td>
<td>4.57265</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>71.00000</td>
<td>20.90909</td>
<td></td>
</tr>
<tr>
<td>Mode</td>
<td>70.00000</td>
<td>18.00000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interquartile Range</td>
<td>3.00000</td>
</tr>
</tbody>
</table>
Tests for Location: Mu0=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student’s t</td>
<td>t 54.54508</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>Sign</td>
<td>M 6</td>
<td>Pr &gt;=</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S 39</td>
<td>Pr &gt;=</td>
</tr>
</tbody>
</table>

Tests for Normality

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk</td>
<td>W 0.846722</td>
<td>Pr &lt; W 0.0335</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D 0.25</td>
<td>Pr &gt; D 0.0376</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq 0.130847</td>
<td>Pr &gt; W-Sq 0.0383</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq 0.764345</td>
<td>Pr &gt; A-Sq 0.0351</td>
</tr>
</tbody>
</table>

The UNIVARIATE Procedure
Fitted Normal Distribution for x2

Parameters for Normal Distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Mu</td>
<td>72</td>
</tr>
<tr>
<td>Std Dev</td>
<td>Sigma</td>
<td>4.572646</td>
</tr>
</tbody>
</table>

Goodness-of-Fit Tests for Normal Distribution

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D 0.25000000</td>
<td>Pr &gt; D 0.038</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq 0.13084710</td>
<td>Pr &gt; W-Sq 0.038</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq 0.76434475</td>
<td>Pr &gt; A-Sq 0.035</td>
</tr>
</tbody>
</table>

The UNIVARIATE Procedure
Variable: x3 (FS late min)

Moments

<table>
<thead>
<tr>
<th>N</th>
<th>12</th>
<th>Sum Weights</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>24</td>
<td>Sum Observations</td>
<td>288</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>2.5584086</td>
<td>Variance</td>
<td>6.54545455</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.2736076</td>
<td>Kurtosis</td>
<td>-0.7231481</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>6984</td>
<td>Corrected SS</td>
<td>72</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>10.6600358</td>
<td>Std Error Mean</td>
<td>0.73854895</td>
</tr>
</tbody>
</table>

Basic Statistical Measures
<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>24.00000</td>
</tr>
<tr>
<td>Median</td>
<td>24.50000</td>
</tr>
<tr>
<td>Mode</td>
<td>25.00000</td>
</tr>
<tr>
<td></td>
<td>Std Deviation</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
</tr>
<tr>
<td></td>
<td>Range</td>
</tr>
<tr>
<td></td>
<td>Interquartile Range</td>
</tr>
</tbody>
</table>
Tests for Location: Mu0=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>-p Value-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's t</td>
<td>t 32.49615</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>Sign</td>
<td>M 6</td>
<td>Pr &gt;=</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S 39</td>
<td>Pr &gt;=</td>
</tr>
</tbody>
</table>

Tests for Normality

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>-p Value-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk</td>
<td>W 0.944175</td>
<td>Pr &lt; W 0.5540</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D 0.166667</td>
<td>Pr &gt; D &gt;0.1500</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq 0.051073</td>
<td>Pr &gt; W-Sq &gt;0.2500</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq 0.30962</td>
<td>Pr &gt; A-Sq &gt;0.2500</td>
</tr>
</tbody>
</table>

The UNIVARIATE Procedure
Fitted Normal Distribution for x3

Parameters for Normal Distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Mu</td>
<td>24</td>
</tr>
<tr>
<td>Std Dev</td>
<td>Sigma</td>
<td>2.558409</td>
</tr>
</tbody>
</table>

Goodness-of-Fit Tests for Normal Distribution

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>-p Value-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D 0.1666667</td>
<td>Pr &gt; D &gt;0.1500</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq 0.051073</td>
<td>Pr &gt; W-Sq &gt;0.2500</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq 0.30961958</td>
<td>Pr &gt; A-Sq &gt;0.2500</td>
</tr>
</tbody>
</table>
The UNIVARIATE Procedure  
Variable: x4  (Tram time min)

Moments

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Sum Weights</th>
<th>12</th>
<th>Sum Observations</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.333333333</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std Deviation</td>
<td>0.49236596</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>0.81240384</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>136</td>
<td>Corrected SS</td>
<td>2.66666667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>14.7709789</td>
<td>Std Error Mean</td>
<td>0.14213381</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.333333333</td>
</tr>
<tr>
<td>Median</td>
<td>3.0000000</td>
</tr>
<tr>
<td>Mode</td>
<td>3.0000000</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests for Location: Mu0=0

<table>
<thead>
<tr>
<th>Test</th>
<th>-Statistic-</th>
<th>------p Value------</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's t</td>
<td>t 23.45208</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>Sign</td>
<td>M 6</td>
<td>Pr &gt;=</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S 39</td>
<td>Pr &gt;=</td>
</tr>
</tbody>
</table>

Tests for Normality

<table>
<thead>
<tr>
<th>Test</th>
<th>-Statistic-</th>
<th>------p Value------</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk</td>
<td>W 0.608091</td>
<td>Pr &lt; W 0.0001</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D 0.417485</td>
<td>Pr &gt; D &lt;0.0100</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq 0.414795</td>
<td>Pr &gt; W-Sq &lt;0.0050</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq 2.32359</td>
<td>Pr &gt; A-Sq &lt;0.0050</td>
</tr>
</tbody>
</table>
The UNIVARIATE Procedure
Fitted Normal Distribution for x4

Parameters for Normal Distribution

Parameter  Symbol  Estimate
Mean       Mu        3.333333
Std Dev    Sigma     0.492366

Goodness-of-Fit Tests for Normal Distribution

Test                 ----Statistic-----  -----p Value------
Kolmogorov-Smirnov   D            0.41746470  Pr > D     <0.010
Cramer-von Mises     W-Sq         0.41479548  Pr > W-Sq  <0.005
Anderson-Darling     A-Sq         2.32358998  Pr > A-Sq  <0.005

The UNIVARIATE Procedure
Variable: x5 (Downtime min)

Moments

N                                     12  Sum Weights  12  Sum Observations  2251
Mean                                   187.583333       Variance   1997.35606
Std Deviation                          44.6917896       Kurtosis   3.7233541
Skewness                               1.60984936       Corrected SS 21970.9167
Uncorrected SS                        444221         Corrected SS 21970.9167
Coeff Variation                       23.8250322       Std Error Mean 12.9014084

Basic Statistical Measures

Location         Variability
Mean             187.5833       Std Deviation  44.69179
Median           187.0000       Variance      1997
Mode             187.0000       Range        162.00000
                  Interquartile Range 50.50000

Tests for Location: Mu0=0

Test                 -Statistic-  -----p Value------
Student's t          t            14.53976     Pr > |t|     <.0001
Sign                 M            6           Pr >= |M|    0.0005
Signed Rank          S            39          Pr >= |S|    0.0005

Tests for Normality

Test                 --Statistic---  -----p Value------
Shapiro-Wilk         W            0.839297    Pr < W    0.0271
Kolmogorov-Smirnov  D            0.198397    Pr > D    >0.1500
Cramer-von Mises     W-Sq         0.086496    Pr > W-Sq  0.1553
Anderson-Darling    A-Sq         0.641968    Pr > A-Sq  0.0747
The UNIVARIATE Procedure
Fitted Normal Distribution for x5

Parameters for Normal Distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Mu</td>
<td>187.5833</td>
</tr>
<tr>
<td>Std Dev</td>
<td>Sigma</td>
<td>44.69179</td>
</tr>
</tbody>
</table>

Goodness-of-Fit Tests for Normal Distribution

<table>
<thead>
<tr>
<th>Test</th>
<th>------</th>
<th>Statistic------</th>
<th>-------</th>
<th>p Value------</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.19839735</td>
<td>Pr &gt; D</td>
<td>&gt;0.150</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.08649558</td>
<td>Pr &gt; W-Sq</td>
<td>0.155</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.64196819</td>
<td>Pr &gt; A-Sq</td>
<td>0.075</td>
</tr>
</tbody>
</table>

The UNIVARIATE Procedure
Variable: x6 (LS early min)

Moments

<table>
<thead>
<tr>
<th>N</th>
<th>12</th>
<th>Sum Weights</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>11.5</td>
<td>Sum Observations</td>
<td>138</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>4.96350336</td>
<td>Variance</td>
<td>24.6363636</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.46033432</td>
<td>Kurtosis</td>
<td>-0.8621542</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>1858</td>
<td>Corrected SS</td>
<td>271</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>43.1608971</td>
<td>Std Error Mean</td>
<td>1.43283994</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>11.500000</td>
</tr>
<tr>
<td>Median</td>
<td>11.500000</td>
</tr>
<tr>
<td>Mode</td>
<td>8.000000</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>4.96350</td>
</tr>
<tr>
<td>Variance</td>
<td>24.63636</td>
</tr>
<tr>
<td>Range</td>
<td>15.00000</td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>7.00000</td>
</tr>
</tbody>
</table>

NOTE: The mode displayed is the smallest of 2 modes with a count of 2.

Tests for Location: Mu=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic-</th>
<th>------</th>
<th>p Value------</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's t</td>
<td>t</td>
<td>8.026019</td>
<td>Pr &gt;</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>6</td>
<td>Pr &gt;=</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>39</td>
<td>Pr &gt;=</td>
</tr>
</tbody>
</table>
Tests for Normality

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk</td>
<td>W 0.936711</td>
<td>Pr &lt; W 0.4567</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D 0.176307</td>
<td>Pr &gt; D &gt;0.1500</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq 0.043627</td>
<td>Pr &gt; W-Sq &gt;0.2500</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq 0.302842</td>
<td>Pr &gt; A-Sq &gt;0.2500</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>y</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
<th>x6</th>
<th>pred</th>
<th>res</th>
<th>studres</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>59457</td>
<td>63</td>
<td>72</td>
<td>22</td>
<td>3</td>
<td>13</td>
<td></td>
<td>60918.05</td>
<td>-1461.05</td>
<td>-0.33794</td>
</tr>
<tr>
<td>2</td>
<td>54847</td>
<td>67</td>
<td>70</td>
<td>26</td>
<td>3</td>
<td>187</td>
<td>13</td>
<td>52219.05</td>
<td>2627.95</td>
<td>0.58455</td>
</tr>
<tr>
<td>3</td>
<td>51350</td>
<td>67</td>
<td>74</td>
<td>25</td>
<td>4</td>
<td>155</td>
<td>16</td>
<td>54054.84</td>
<td>-2704.84</td>
<td>-0.58774</td>
</tr>
<tr>
<td>4</td>
<td>60042</td>
<td>67</td>
<td>76</td>
<td>24</td>
<td>4</td>
<td>142</td>
<td>20</td>
<td>55890.63</td>
<td>4151.37</td>
<td>0.89569</td>
</tr>
<tr>
<td>5</td>
<td>60898</td>
<td>66</td>
<td>71</td>
<td>20</td>
<td>4</td>
<td>143</td>
<td>8</td>
<td>63572.75</td>
<td>-2674.75</td>
<td>-0.67380</td>
</tr>
<tr>
<td>6</td>
<td>51956</td>
<td>70</td>
<td>72</td>
<td>25</td>
<td>3</td>
<td>148</td>
<td>7</td>
<td>53037.96</td>
<td>-1081.96</td>
<td>-0.23681</td>
</tr>
<tr>
<td>7</td>
<td>46085</td>
<td>67</td>
<td>70</td>
<td>25</td>
<td>3</td>
<td>203</td>
<td>5</td>
<td>54054.84</td>
<td>-7969.84</td>
<td>-1.73178</td>
</tr>
<tr>
<td>8</td>
<td>45962</td>
<td>63</td>
<td>84</td>
<td>28</td>
<td>4</td>
<td>199</td>
<td>8</td>
<td>49903.31</td>
<td>-3941.31</td>
<td>-0.99286</td>
</tr>
<tr>
<td>9</td>
<td>58854</td>
<td>67</td>
<td>71</td>
<td>27</td>
<td>3</td>
<td>167</td>
<td>6</td>
<td>50383.26</td>
<td>8470.74</td>
<td>1.96529</td>
</tr>
<tr>
<td>10</td>
<td>52279</td>
<td>88</td>
<td>68</td>
<td>22</td>
<td>3</td>
<td>187</td>
<td>12</td>
<td>52444.07</td>
<td>-165.07</td>
<td>-0.12184</td>
</tr>
<tr>
<td>11</td>
<td>56704</td>
<td>66</td>
<td>66</td>
<td>24</td>
<td>3</td>
<td>201</td>
<td>19</td>
<td>56229.59</td>
<td>474.41</td>
<td>0.10267</td>
</tr>
<tr>
<td>12</td>
<td>68864</td>
<td>63</td>
<td>70</td>
<td>20</td>
<td>3</td>
<td>215</td>
<td>11</td>
<td>64589.63</td>
<td>4274.37</td>
<td>1.12857</td>
</tr>
</tbody>
</table>

Studentized residual of the final model