METHODOLOGY FOR PREDICTING DRILLING PERFORMANCE FROM ENVIRONMENTAL CONDITIONS

A Thesis

by

JOSE ALEJANDRO DE ALMEIDA

Submitted to the Office of Graduate Studies of Texas A&M University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2010

Major Subject: Petroleum Engineering
Methodology for Predicting Drilling Performance from Environmental Conditions

Copyright 2010 Jose Alejandro de Almeida
METHODOLOGY FOR PREDICTING DRILLING PERFORMANCE FROM ENVIRONMENTAL CONDITIONS

A Thesis

by

JOSE ALEJANDRO DE ALMEIDA

Submitted to the Office of Graduate Studies of Texas A&M University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee, Gene Beck
Committee Members, Jerome Schubert
                                 Michael Sherman
Head of Department, Steve Holditch

December 2010

Major Subject: Petroleum Engineering
ABSTRACT

Methodology for Predicting Drilling Performance from Environmental Conditions. (December 2010)

Jose Alejandro de Almeida, B.S., Colorado State University

Chair of Advisory Committee: Dr. Gene Beck

The use of statistics has been common practice within the petroleum industry for over a decade. With such a mature subject that includes specialized software and numerous articles, the challenge of this project was to introduce a duplicable method to perform deterministic regression while confirming the mathematical and actual validation of the resulting model. A five-step procedure was introduced using Statistical Analysis Software (SAS) for necessary computations to obtain a model that describes an event by analyzing the environmental variables. Since SAS may not be readily available, the code to perform the five-step methodology in R has been provided.

The deterministic five-step procedure methodology may be applied to new fields with a limited amount of data. As an example case, 17 wells drilled in north central Texas were used to illustrate how to apply the methodology to obtain a deterministic model. The objective was to predict the number of days required to drill a well using environmental conditions and technical variables. Ideally, the predicted number of days would be within +/- 10% of the observed time of the drilled wells. The database created contained 58 observations from 17 wells with the descriptive variables, technical limit (referred to as estimated days), depth, bottomhole temperature (BHT), inclination (inc), mud weight (MW), fracture pressure (FP), pore pressure (PP), and the average, maximum, and minimum difference between fracture pressure minus mud weight and mud weight minus pore pressure.
Step 1 created a database. Step 2 performed initial statistical regression on the original dataset. Step 3 ensured that the models were valid by performing univariate analysis. Step 4 history matched the models-response to actual observed data. Step 5 repeated the procedure until the best model had been found. Four main regression techniques were used: stepwise regression, forward selection, backward elimination, and least squares regression. Using these four regression techniques and best engineering judgment, a model was found that improved time prediction accuracy, but did not constantly result in values that were +/- 10% of the observed times.

The five-step methodology to determine a model using deterministic statistics has applications in many different areas within the petroleum field. Unlike examples found in literature, emphasis has been given to the validation of the model by analysis of the model error. By focusing on the five-step procedure, the methodology may be applied within different software programs, allowing for greater usage. These two key parameters allow companies to obtain their time prediction models without the need to outsource the work and test the certainty of any chosen model.
DEDICATION

To my father, mother, and sister, who always support me in my decisions and are present to guide me with their advice. To my fiancé and her love towards me.
ACKNOWLEDGEMENTS

I would like to give special thanks to Texas A&M University’s Statistics Department. Their Statistical Consulting Team and PhD graduate student Karl Gregory have taught me more about statistics then I ever thought I would learn.
# NOMENCLATURE

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Actual Days</td>
</tr>
<tr>
<td>BHT</td>
<td>Bottomhole Temperature</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>Regression Coefficient</td>
</tr>
<tr>
<td>Cp</td>
<td>Mallow Cp Value</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of Variance</td>
</tr>
<tr>
<td>D</td>
<td>Depth</td>
</tr>
<tr>
<td>F</td>
<td>F-number</td>
</tr>
<tr>
<td>FP</td>
<td>Fracture Pressure</td>
</tr>
<tr>
<td>inc</td>
<td>Inclination</td>
</tr>
<tr>
<td>$k$</td>
<td>Number of Regressor Coefficients</td>
</tr>
<tr>
<td>MS</td>
<td>Mean Squares</td>
</tr>
<tr>
<td>$MS_T$</td>
<td>Model Mean Square</td>
</tr>
<tr>
<td>MW</td>
<td>Mud Weight</td>
</tr>
<tr>
<td>n</td>
<td>Number of Observations</td>
</tr>
<tr>
<td>N</td>
<td>Number of Data Points</td>
</tr>
<tr>
<td>p</td>
<td>number of parameters</td>
</tr>
<tr>
<td>PP</td>
<td>Pore Pressure</td>
</tr>
<tr>
<td>$R^2$</td>
<td>Coefficient of Determination</td>
</tr>
<tr>
<td>adj$R^2$</td>
<td>Adjusted R-squared</td>
</tr>
<tr>
<td>RMS$_E$</td>
<td>Root Mean Squared</td>
</tr>
</tbody>
</table>
\( \hat{\sigma}^2 \)  
Estimator of Variance

SS\(_E\)  
Error Sum of Squares

SS\(_R\)  
Model Sum of Squares

SS\(_T\)  
Total Sum of Squares

Tbh  
Bottomhole Temperature

Te  
Estimated Days

\( x_i \)  
Regressor Variable

Y  
Dependent Variable

Y\(_i\)  
Calculated Values from Models

\( \hat{Y} \)  
Observed Experimental Values

\( \bar{Y} \)  
Overall Mean
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>DEDICATION</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENTS</td>
<td>vi</td>
</tr>
<tr>
<td></td>
<td>NOMENCLATURE</td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td>TABLE OF CONTENTS</td>
<td>ix</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>xvii</td>
</tr>
<tr>
<td>I</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>UNDERSTANDING SAS’S STEPWISE REGRESSION, FORWARD SELECTION, BACKWARD ELIMINATION, AND LEAST SQUARES REGRESSION</td>
<td>3</td>
</tr>
<tr>
<td>III</td>
<td>UNDERSTANDING UNIVARIATE OUTPUT</td>
<td>8</td>
</tr>
<tr>
<td>IV</td>
<td>PROCEDURE FOR STEPWISE REGRESSION, FORWARD SELECTION, AND BACKWARD ELIMINATION</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>4.1 Method for Stepwise Regression (Loucks 2003)</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>4.2 Method for Forward Selection (Loucks 2003)</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>4.3 Method for Backward Elimination (Loucks 2003)</td>
<td>13</td>
</tr>
<tr>
<td>V</td>
<td>REASON FOR A NEGATIVE INTERCEPT</td>
<td>14</td>
</tr>
<tr>
<td>VI</td>
<td>APPLYING THE METHODOLOGY FOR WELLS IN NORTH CENTRAL TEXAS</td>
<td>15</td>
</tr>
<tr>
<td>VII</td>
<td>SAS RESULTS FOR REGRESSION ANALYSIS OF NORTH CENTRAL TEXAS WELLS</td>
<td>17</td>
</tr>
<tr>
<td>Chapter/Appendix</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>VIII</td>
<td>HISTORY MATCHING OF SIGNIFICANT MODELS</td>
<td>36</td>
</tr>
<tr>
<td>IX</td>
<td>LITERATURE COMPARISON WITH METHODOLOGY AND RESULTS</td>
<td>38</td>
</tr>
<tr>
<td>X</td>
<td>CONCLUSION</td>
<td>43</td>
</tr>
<tr>
<td>REFERENCES</td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>A</td>
<td>PROCEDURE FOR PERFORMING ANALYSIS OF VARIANCE IN SAS</td>
<td>48</td>
</tr>
<tr>
<td>B</td>
<td>R SOFTWARE CODE</td>
<td>63</td>
</tr>
<tr>
<td>C</td>
<td>MODELS C, D, E, H, I, J, AND K’S MODEL EQUATION, RESIDUAL PLOT, QQ PLOT, AND THREE HISTORY MATCH CURVES</td>
<td>80</td>
</tr>
<tr>
<td>D</td>
<td>GRAPHICAL COMPARISON OF DEPENDENT VERSE INDEPENDENT VARIABLES</td>
<td>98</td>
</tr>
<tr>
<td>E</td>
<td>SAS OUTPUT FOR ALL REGRESSIONS AND UNIVARIATE CALCULATIONS</td>
<td>103</td>
</tr>
<tr>
<td>F</td>
<td>SAS CODE NECESSARY FOR REGRESSIONS AND UNIVARIATE CALCULATIONS</td>
<td>159</td>
</tr>
<tr>
<td>G</td>
<td>AN EXAMPLE OF R’S CODE AND OUTPUT FOR REGRESSIONS AND UNIVARIATE CALCULATIONS</td>
<td>166</td>
</tr>
<tr>
<td>H</td>
<td>A TEMPLATE OF R’S CODE FOR REGRESSION AND UNIVARIATE CALCULATIONS</td>
<td>175</td>
</tr>
<tr>
<td>VITA</td>
<td></td>
<td>177</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Normal Distribution Around Residual Point</td>
<td>8</td>
</tr>
<tr>
<td>Figure 7.1</td>
<td>Model A’s Residual Plot. Actual Days = f(Te,Tbh)</td>
<td>18</td>
</tr>
<tr>
<td>Figure 7.2</td>
<td>Model A’s Q-Q Plot. Actual Days = f(Te,Tbh)</td>
<td>19</td>
</tr>
<tr>
<td>Figure 7.3</td>
<td>Model B’s Residual Plot. Actual Days = f(Te,D)</td>
<td>20</td>
</tr>
<tr>
<td>Figure 7.4</td>
<td>Model C’s Residual Plot. Actual Days = f(Te,D)</td>
<td>21</td>
</tr>
<tr>
<td>Figure 7.5</td>
<td>Model C’s Q-Q Plot. Actual Days = f(Te,D)</td>
<td>21</td>
</tr>
<tr>
<td>Figure 7.6</td>
<td>Actual Days vs. Estimated Days</td>
<td>23</td>
</tr>
<tr>
<td>Figure 7.7</td>
<td>Actual Days vs. Depth</td>
<td>23</td>
</tr>
<tr>
<td>Figure 7.8</td>
<td>Actual Days vs. BHT</td>
<td>24</td>
</tr>
<tr>
<td>Figure 7.9</td>
<td>ln(Actual Days) vs ln(Estimated Days)</td>
<td>25</td>
</tr>
<tr>
<td>Figure 7.10</td>
<td>ln(Actual Days) vs ln(Depth)</td>
<td>25</td>
</tr>
<tr>
<td>Figure 7.11</td>
<td>ln(Actual Days) vs Depth</td>
<td>26</td>
</tr>
<tr>
<td>Figure 7.12</td>
<td>ln(Actual Days) vs BHT</td>
<td>26</td>
</tr>
<tr>
<td>Figure 7.13</td>
<td>ln(Actual Days) vs Inclination</td>
<td>27</td>
</tr>
<tr>
<td>Figure 7.14</td>
<td>ln(Actual Days) vs (Estimated Days)$^2$</td>
<td>27</td>
</tr>
<tr>
<td>Figure 7.15</td>
<td>Model E’s Residual Plot. ln(Actual Day) = f(ln(Te),D)</td>
<td>28</td>
</tr>
<tr>
<td>Figure 7.16</td>
<td>Model E’s Q-Q Plot. ln(Actual Day) = f(ln(Te),D)</td>
<td>29</td>
</tr>
<tr>
<td>Figure 7.17</td>
<td>Model G’s Residual Plot. ln(Actual Days) = f(ln(Te),Tbh)</td>
<td>30</td>
</tr>
<tr>
<td>Figure 7.18</td>
<td>Model G’s Q-Q Plot. ln(Actual Days) = f(ln(Te),Tbh)</td>
<td>31</td>
</tr>
<tr>
<td>Figure 7.19</td>
<td>Model H’s Residual Plot. ln(Actual Days) = f(Te,D)</td>
<td>32</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>7.20</td>
<td>Model H’s Q-Q Plot. $\ln(\text{Actual Days}) = f(Te,D)$</td>
<td>32</td>
</tr>
<tr>
<td>7.21</td>
<td>Model J’s Residual Plot. $\ln(\text{Actual Days}) = f(Te,\ln(D))$</td>
<td>34</td>
</tr>
<tr>
<td>7.22</td>
<td>Model J’s Q-Q Plot. $\ln(\text{Actual Days}) = f(Te,\ln(D))$</td>
<td>34</td>
</tr>
<tr>
<td>A.1</td>
<td>Example Dataset from North Central Texas Wells</td>
<td>48</td>
</tr>
<tr>
<td>A.2</td>
<td>Initial Steps to Import a Dataset into SAS</td>
<td>49</td>
</tr>
<tr>
<td>A.3</td>
<td>Middle Steps to Import a Dataset into SAS</td>
<td>50</td>
</tr>
<tr>
<td>A.4</td>
<td>Final Steps to Import a Dataset into SAS</td>
<td>51</td>
</tr>
<tr>
<td>A.5</td>
<td>Last Step to Import a Dataset into SAS</td>
<td>51</td>
</tr>
<tr>
<td>A.6</td>
<td>Variable Names Seen within the Log Window</td>
<td>52</td>
</tr>
<tr>
<td>A.7</td>
<td>Attach Dataset into SAS Workbook</td>
<td>53</td>
</tr>
<tr>
<td>A.8</td>
<td>Example SAS Input for Stepwise Regression</td>
<td>54</td>
</tr>
<tr>
<td>A.9</td>
<td>Example SAS Input for Forward Selection</td>
<td>54</td>
</tr>
<tr>
<td>A.10</td>
<td>Example SAS Input for Backward Elimination</td>
<td>55</td>
</tr>
<tr>
<td>A.11</td>
<td>Example SAS Input for Least Squares Regression</td>
<td>55</td>
</tr>
<tr>
<td>A.12</td>
<td>Example SAS Input for Univariate Calculations</td>
<td>56</td>
</tr>
<tr>
<td>A.13</td>
<td>Example SAS Input for Residual Plot</td>
<td>57</td>
</tr>
<tr>
<td>A.14</td>
<td>Example SAS Input for Stepwise Regression, Univariate Calculation, and Residual Plot</td>
<td>58</td>
</tr>
<tr>
<td>A.15</td>
<td>Example SAS Input for Least Sum Regression, Univariate Calculation, and Residual Plot</td>
<td>59</td>
</tr>
<tr>
<td>A.16</td>
<td>Example SAS Input for Removal of a Data Point</td>
<td>60</td>
</tr>
<tr>
<td>A.17</td>
<td>Example SAS Input for Creating a Plot</td>
<td>61</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>A.18</td>
<td>Example SAS Input for Variable Manipulation</td>
<td>62</td>
</tr>
<tr>
<td>B.1</td>
<td>Code to Open Window to Import Database</td>
<td>64</td>
</tr>
<tr>
<td>B.2</td>
<td>Code to See Database Values</td>
<td>64</td>
</tr>
<tr>
<td>B.3</td>
<td>Code to Attach Database</td>
<td>65</td>
</tr>
<tr>
<td>B.4</td>
<td>Code to Perform Least Squares Regression</td>
<td>66</td>
</tr>
<tr>
<td>B.5</td>
<td>Code to Create a Model with all Variables</td>
<td>66</td>
</tr>
<tr>
<td>B.6</td>
<td>Code to Perform Stepwise Regression</td>
<td>67</td>
</tr>
<tr>
<td>B.7</td>
<td>Code to Create a Model with 1 Variable</td>
<td>68</td>
</tr>
<tr>
<td>B.8</td>
<td>Code for Forward Selection R</td>
<td>68</td>
</tr>
<tr>
<td>B.9</td>
<td>Code for Backward Elimination</td>
<td>69</td>
</tr>
<tr>
<td>B.10</td>
<td>Code for Stepwise Regression with Mallow Cp</td>
<td>70</td>
</tr>
<tr>
<td>B.11</td>
<td>Code for Forward Selection with Mallow Cp</td>
<td>70</td>
</tr>
<tr>
<td>B.12</td>
<td>Code for Backward Elimination with Mallow Cp</td>
<td>71</td>
</tr>
<tr>
<td>B.13</td>
<td>Code to Create Four Plots</td>
<td>71</td>
</tr>
<tr>
<td>B.14</td>
<td>Results of plot.lm Code</td>
<td>72</td>
</tr>
<tr>
<td>B.15</td>
<td>Code for Residual Plot</td>
<td>73</td>
</tr>
<tr>
<td>B.16</td>
<td>Residual Plot</td>
<td>73</td>
</tr>
<tr>
<td>B.17</td>
<td>Code for QQ Plot in R</td>
<td>74</td>
</tr>
<tr>
<td>B.18</td>
<td>QQ Plot from Previous Code</td>
<td>75</td>
</tr>
<tr>
<td>B.19</td>
<td>Code for Shapiro Wilk Normality Test</td>
<td>76</td>
</tr>
<tr>
<td>B.20</td>
<td>Code for Pearson Normality Test</td>
<td>76</td>
</tr>
</tbody>
</table>
Figure B.21 Code for Summary Information ............................................................ 77
Figure B.22 Example Code of Basic Variable Manipulation .................................... 77
Figure B.23 Code to Bind a New Variable to an Existing Database ...................... 78
Figure B.24 Code to Remove Outliers .................................................................... 79
Figure C.1   Residual Plot of Model C ................................................................. 80
Figure C.2   QQ Plot of Model C ......................................................................... 81
Figure C.3   History Match between Actual and Predicted Days for Model C ........ 81
Figure C.4   History Match, Predicted vs Actual Days for Model C ...................... 82
Figure C.5   History Match, Predicted vs Actual Days for Model C Zoomed In ...... 82
Figure C.6   Residual Plot for Model D ................................................................. 83
Figure C.7   QQ Plot for Model D ......................................................................... 83
Figure C.8   History Match between Actual and Predicted Days for Model D ....... 84
Figure C.9   History Match, Predicted vs Actual Days for Model D ...................... 84
Figure C.10 History Match, Predicted vs Actual Days for Model D Zoomed In .... 85
Figure C.11 Residual Plot for Model E ................................................................. 85
Figure C.12 QQ Plot for Model E ......................................................................... 86
Figure C.13 History Match between Actual and Predicted Days for Model E ....... 86
Figure C.14 History Match, Predicted vs Actual Days for Model E ...................... 87
Figure C.15 History Match, Predicted vs Actual Days for Model E Zoomed In .... 87
Figure C.16 Residual Plot for Model H ................................................................. 88
Figure C.17 QQ Plot for Model H ......................................................................... 88
Figure C.18 History Match between Actual and Predicted Days for Model H........... 89
Figure C.19 History Match, Predicted vs Actual Days for Model H ...................... 89
Figure C.20 History Match, Predicted vs Actual Days for Model H Zoomed In ..... 90
Figure C.21 Residual Plot for Model I .................................................................... 90
Figure C.22 QQ Plot for Model I .............................................................................. 91
Figure C.23 History Match between Actual and Predicted Days for Model I ......... 91
Figure C.24 History Mach, Predicted vs Actual Days for Model I ......................... 92
Figure C.25 History Mach, Predicted vs Actual Days for Model I Zoomed In ........ 92
Figure C.26 Residual Plot for Model J .................................................................... 93
Figure C.27 QQ Plot for Model J .............................................................................. 93
Figure C.28 History Match between Actual and Prediced Days for Model J ........... 94
Figure C.29 History Match, Predicted vs Actual Days for Model J ....................... 94
Figure C.30 History Match, Predicted vs Actual Days for Model J Zoomed In ...... 95
Figure C.31 Residual Plot for Model K .................................................................... 95
Figure C.32 QQ Plot for Model K ............................................................................. 96
Figure C.33 History Match between Actual and Predicted Days for Model K....... 96
Figure C.34 History Match, Predicted Days vs Actual Days for Model K ............... 97
Figure C.35 History Match, Predicted Days vs Actual Days for Model K ............... 97
Figure D.1 Actual Days vs Estimated Days ............................................................ 98
Figure D.2 Actual Days vs Depth ............................................................................ 98
Figure D.3 Actual Days vs. BHT ............................................................................ 99
| Figure D.4   | ln(Actual Days) vs. ln(Estimated Days) | 99 |
| Figure D.5   | ln(Actual Days) vs. ln(Depth)          | 100|
| Figure D.6   | ln(Actual Days) vs. Depth              | 100|
| Figure D.7   | ln(Actual Days) vs. BHT                | 101|
| Figure D.8   | ln(Actual Days) vs. inc                | 101|
| Figure D.9   | ln(Actual Days) vs. (Estimated Days)^2 | 102|
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 8.1 Summation of Percent Error and Number of Times the +/-10% Objective was Obtained per Well Section</td>
<td>35</td>
</tr>
<tr>
<td>Table 8.2 Summation of Percent Error and Number of Times the +/-10% Objective was Obtained per Well</td>
<td>36</td>
</tr>
<tr>
<td>Table 9.1 Average Standard Deviation</td>
<td>39</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Statistical regression allows the influence and significance of variables to be chosen without bias. The method to produce a statistical deterministic model has well been established through the use of analytical techniques. The objective of this thesis was to introduce a methodology applicable to petroleum engineering to find a deterministic model from data through the use of statistics.

The goal of this project was to create a model that explains how certain variables affect an outcome. By combining the variables and finding out how much influence each variable has, a predictive model can be found that emulates observed results. Shown below is the general formula for a multiple linear model (Montgomery and Runger 2007).

\[ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \]  

\[ \text{(1.1)} \]

\( \beta_i \) are regression coefficient and \( x_i \) regressor variables. First, significant regressor variables and regression coefficients are found. Using the significant variable and regression coefficients, \( Y_{\hat{Y}} \) are calculated from the model. Experimental values are \( \hat{Y} \). The smaller the difference between \( Y_i \) and \( \hat{Y} \), the better the model follows the observed data. Residuals, also known as error or deviation, are the individual values calculated from the difference between \( \hat{Y} \) and \( Y_i \) (Dallal 2008). An assumption for model fitting was that all residuals, \( \varepsilon \), are independent of one another. To test the adequacy of the model, univariate calculations and residual plots were used.

This thesis introduces a regression methodology for creating a prediction model that requires five steps. Step 1 creates a database with a dependent variable and
independent variables that are believed to affect the dependent variable. Step 2 performs the initial statistical regression on the original data. Software with statistical capabilities determines which variables are significant and calculates the regression coefficients. An alternative allows individually chosen independent variables to be tested through least squares regression. Step 3 ensures that the models are valid by performing univariate analysis. Step 4 history matches the model’s response to actual observed data. Step 5 repeats the procedure till the best model has been found. If no significant models are found after performing the calculations using the initial dataset, data manipulation may be required.

This thesis introduces a methodology to obtain a predictive model that explains the relationship between a dependent and multiple independent variables through the use of linear regression. The methodology was tested using actual data from north central Texas. The number of days required to drill a well was the dependent variable. The independent variables tested were technical limit (referred as estimated days), depth, bottomhole temperature (BHT), inclination (inc), mud weight (MW), fracture pressure (FP), pore pressure (PP), and the average, maximum, and minimum difference between fracture pressure minus mud weight and mud weight minus pore pressure.

Many available commercial software perform specific analysis for predicting the amount of time required to drill a well, and others have the ability to calculate the necessary mathematical equations. A common basic software package used is Microsoft Excel; add-ons for Microsoft Excel assist users with prediction calculations. Independent of the software, the primary method for prediction calculations involves statistical analysis of data. Statistical Analysis Software (SAS) was specifically chosen since the software tailors to all statistic calculations. The second software used to perform statistical calculations is simply called R. Statistical analysis typically does not produce a unique answer but requires interpretation of results. A certain amount of intuition and individualized decision making are needed, and if these are not used carefully, calculation that result may have no real-world descriptive capabilities.
CHAPTER II
UNDERSTANDING SAS’S STEPWISE REGRESSION, FORWARD SELECTION, BACKWARD ELIMINATION, AND LEAST SQUARES REGRESSION

SAS performs stepwise regression, forward selection, backward elimination, and least squares regression. All four methods are a particular type of analysis of variance, ANOVA. Stepwise regression, forward selection, and backward elimination calculate regressor coefficients and significant regressor variables, while least squares regression calculates the regressor coefficients. All basic equations required for regression are described below and written out in Appendix A.

The sum of squares, also known as the sum of squared deviations, measures the dispersion or variability in the model’s response. There are three different types of sum of squares in ANOVA calculations: total corrected sum of squares, error sum of squares, and model sum of squares. Model sum of squares is also known as regression sum of squares. Total corrected sum of squares, $SS_T$ (Montgomery and Runger 2007), is used to see the deviation from the simplest model, $Y = \beta_0$, where $\beta_0$ equals the average value of $\hat{Y}$ (Orlov 1996).

$$SS_T = SS_E + SS_R$$ \hfill (2.1)

$SS_E$ takes into account the randomness of the data set. It tests the model’s ability to replicate $\hat{Y}$ using a combination of settings (Montgomery and Runger 2007).

$$SS_E = \sum_{i=1}^{n} (Y_i - \hat{Y})^2$$ \hfill (2.2)

Smaller values of $SS_E$ correlate to more accurate regression coefficients. While $SS_E$ illustrates the randomness of the data, $SS_R$ describes the regressor variables in the model (Dallal 2008).
Type II Sum of Squares indicates the amount each variable reduces the total sum of squares. More significant variables have large Type II Sum of Squares values while insignificant variable have small values.

Mean squares are estimates of variance. They are determined by dividing a sum of squares by the degrees of freedom. Degrees of freedom are defined by the difference between the number of observations and number of parameters. Total mean square, MSₜ, equals SSₜ divided by the degrees of freedom of SSₜ. Since SSₜ has only one parameter, β₀, the equation for MSₜ equals SSₜ divided by n-1(Montgomery and Runger 2007).

\[ MSₜ = \frac{SSₜ}{n-1} \]  
\[ \text{(2.4)} \]

Error mean square, MSₑ, is SSₑ divided by the degrees of freedom of SSₑ. The parameters associated with SSₑ are all regressor variables, so the degree of freedom equals n-k where n equals the number of observations and k represents the number of regressor variable (Orlov 1996).

\[ MSₑ = \frac{SSₑ}{n-k} \]  
\[ \text{(2.5)} \]

RMSₑ is the square root of the MSₑ. Smaller values of RMSₑ, more accurate the model response.

\[ MSₜ = \frac{SSᵣ}{k-1} \]  
\[ \text{(2.6)} \]
k-1 represents the difference between the total mean square degrees of freedom and error mean square degrees of freedom (Orlov 1996).

The F-number found through the F-test indicates whether the model has statistically significant predictive capability (Dallal 2008). It tests the difference between two variances. Examples of variances are mean squares.

\[ F = \frac{MS_R}{MS_E} \] \hspace{1cm} (2.7)

The F-number determines if the null hypothesis, that all the regressor coefficients are equal to zero, is false (Montgomery and Runger 2007). SAS calculates the probability value, also known as the p-value, from the F distribution using the F-number and the degrees of freedom of the variables being tested. The probability that the variances are different equals 1 minus the p-value; this is known as confidence level (Orlov 1996). An F-test p-value has significance if it has a value less than \( \alpha \). \( \alpha \) may be independently chosen; for this thesis \( \alpha \) has a value of 0.05. If the F-number shows no significance, the p-value should be ignored (Dallal 2008).

The t-value tests the null hypothesis that the parameter estimator equals zero. Small values of standard errors of the mean represent sample means that are near true means. Larger t-values and smaller p-values occur for small standard error of the mean values (2007).

Both the F-test and t-test test their hypotheses by assuming that the samples came from a normal residual distribution. The chi-squared test checks whether the variances of the errors are constant throughout all observations. Constant variance errors, residuals with the same variance, are known as being homoscedastic. If the variances of the errors differ, then they are known as heteroscedastic. If the variances are constant, then the chi-squared p-value will be larger than 0.05 (2010).

Standard errors are the standard deviation of the estimator, the square root of the mean square error. R squared, also known as the coefficient of determination, describes
the proportion of variance and response that the model fits (Montgomery and Runger 2007).

\[ R^2 = \frac{SS_R}{SS_T} \] \hspace{5cm} (2.8)

Caution must be used when analyzing R-squared values because though an R squared value near one may indicate a good fit to the data; the model may actually be a bad fit. With the addition of each new variable, the R-squared will increase or stay the same, independent whether the added variable was significant. The adjusted R-squared, or adjusted coefficient of determination, takes into account the number of parameters within a model. As the number of parameters increases, the adjusted R-squared value decreases.

\[ R^2_{adj} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} \] \hspace{5cm} (2.9)

Multiple Rs are the square root of R-squared which estimates the influence of variables on the dependent variables (Orlov 1996).

The dependent mean represents the mean of the dependent variable (2007). The coefficient of variation represents a unitless measure of the variation in the data.

\[ CV = \frac{RMS_E}{\bar{y}} \] \hspace{5cm} (2.10)

Mallows Cp values are used to measure the accuracy of the model response.

\[ Cp = \frac{SS_E}{SS_T} - n + 2k \] \hspace{5cm} (2.11)

A model with the a Cp value nearest to 1+x indicates a good model (Beal 2007).
The variance inflation factor tests how variables are intercorrelated. Any value greater than 10 indicates that the variables are too intercorrelated. Pearson’s correlation coefficients also measure how intercorrelated variables are. Pearson’s correlation tests whether a linear relationship exists between two variables. If a positive correlation exists between two variables, then when one variable increase the other increases. A negative correlation indicates that as one variable increases, the other decreases. As the Pearson correlation value approaches one or negative one, the greater the intercorrelation is (Lawrence 2010).

If any changes are done to the model, such as the removal or addition of a regressor coefficient, the regression has to be redone. Many times one regressor coefficient may affect another one such that the removal of one variable may impact the model in ways unknown. Designers should always check the values of both hypothesis tests to best understand how the model mirrors the sample data.
CHAPTER III

UNDERSTANDING UNIVARIATE OUTPUT

After obtaining the regressor variables and regression coefficients, a univariate analysis tests the accuracy of the model by performing multiple tests on the model’s residuals. Residuals are the unexplained variations from the regression model. A key assumption for regression analysis requires that all residual points have a normal distribution variation; Figure 3.1 illustrates the concept.

![Fig. 3.1- Normal Distribution Around Residual Point](image)

SAS splits univariate output into six tables: Moments, Basic Statistical Measures, Tests for Location, Tests for Normality, and Quantiles.

Moments are standardized descriptive statistics. SAS provides details of the data distribution under the Moments table. N equals the number of data points. Weights are the value given to each data point. In this regression, all the data points are considered equal so each point has a value of one and the sum of weights equals the number of data points (2000). The mean of residual points equals the average of the difference between the predicted values and the observed data. Note that a residual mean equal to zero does not ensure that the model follows the observed data’s trend (Montgomery and Runger 2007).

\[
\sigma^2 = \frac{SS_g}{n-p} \tag{3.1}
\]

Standard deviation equals the square root of the variance.
Skewness measures the standardization of the mean and mode and illustrates the asymmetry of the distribution. A positive value for skewness represents the tail extending to the right while a negative number has the opposite trend. Normal data would have a skewness of zero while a large number would indicate that outliers play an important role in the data. Kurtosis describes the flatness of the top of a symmetric distribution compared to a normal distribution (Wuensch 2007). A normal distribution has a kurtosis value of zero while a positive value indicates a peaked distribution and a negative number a flat distribution.

The corrected sum of squares equals the squared difference between the response variable and the mean response variable. Uncorrected sum of squares should equal the square of the sum of the response variable. If the uncorrected sum of squares and the corrected sum of squares do not equal each other, then the mean of the residuals does not equal zero and there is a discrepancy within the model (2007).

The standard error of the mean equals the standard deviation divided by the square root of the number of data points. Confidence intervals are multipliers of standard errors. Calculate a 90% confidence interval requires multiplying the standard mean error by +/- 1.64; 95%, standard mean error multiply by +/-1.96; and 99%, standard mean error multiply by +/-2.58 (Montgomery and Runger 2007).

Basic statistical measures illustrate the central tendency and spread of the model. The mean equals the sum of all data points divided by the number of data points. The median represents where half the points are larger and the other half smaller, locating the exact middle. The mode represents the number that occurs with greatest frequency. The range gives an indication of the dispersion by subtracting the largest value by the smallest (Croarkin and Guthrie 2006). The range is based on the extremes. It does not indicate if majority of the data are in the middle or out in the tail. Interquartile range represents the value of the 75th percentile minus the value of the 25th percentile. The interquartile range attempts to measure the variability in the middle (Croarkin and Guthrie 2006).
There are three tests for location: student’s t-test, sign test, and Wilcoxon signed rank test. All three test the null hypothesis that the mean or median equals a certain number. This number may be assigned, but for these tests the mean equals zero (2007). The student’s t-test works best when the data has a normal distribution while the sign and signed rank can be used with nonnormal distributions. The student’s t-test calculates a t value by subtracting the sample mean from the actual mean and dividing by the standard error of the sample mean. The sign test illustrates whether the model over or underestimates (Montgomery and Runger 2007). The Wilcoxon signed test ranks the differences between observations and not the mean. A negative signed rank number means that the model has a tendency to underestimate, and a positive value indicates a tendency to overestimate (2007).

The test for normality assesses whether a normal distribution occurs for the residuals. Each of the following four tests has a different method to obtain this conclusion. The Shapiro-Wilk number calculates the ratio of the best estimator of variance to the corrected sum of squares. If the Shaprio-Wilk number equals one, that indicates a normally distributed sample (Park 2008). Small values of W for Shapiro-Wilk indicate departure from normal distribution. Kolmogorov-Smirnov does not depend on any specific distribution; instead it tests the hypothesis that the data follow a selected distribution. Anderson-Darling identifies the distribution the data come from. Modified from Kolmogorov-Smirnov, Anderson-Darling places more emphasis on the tails. Critical values of Anderson-Darling determine the distribution (Croarkin and Guthrie 2006). Cramer-von-Mises tests the hypothesis if the data came from a certain distribution by having tabulated data. If the value are larger than the tabulated, data then the hypothesis is rejected (2007).

Q-Q plots can graphically tell if the residuals follow a normal distribution. If the residuals follow a normal distribution, then a straight line is formed with the residual mean as the intercept and the residual standard deviation as the slope. The x-axis indicates which percentiles of residuals have values below the matching y-point (1999).
CHAPTER IV
PROCEDURE FOR STEPWISE REGRESSION, FORWARD SELECTION, AND BACKWARD ELIMINATION

The objective of this project was to find the significant regressor variables and regressor coefficients that describe the behavior of the dependent variable from a selected data set. Selection of regressor variables was done using different regression methods. All significant variables or combinations of variables have p-values less than $\alpha$. Within this thesis, $\alpha$ was assigned a value of 0.05.

Stepwise regression happens to be the mostly widely used variable selection technique (Montgomery and Runger 2007). Variables are added one at a time as long as the F-statistic p-values are less than or equal to $\alpha$. After the addition of a variable, all of the variable’s F-statistic p-values are evaluated with a larger $\alpha_2$, equal to 0.15. Any variable with the an $\alpha_2$ value greater than 0.15 is removed (Beal 2007). It is possible that no variables are removed; then the stepwise regression acts as a forward selection regression. Forward selection simplifies stepwise regression by removing backward elimination, the process of removing variables after selection. In forward selection, regressor variables are introduced one at a time until no significant p-values remain. A weakness of forward selection occurs because it does not take into account the effects of the added regressor variable on previous regressor variables. Backward elimination begins will all regressor variables in the model. The regressor variable with the largest F-test p-value gets removed, and the process repeats till all regressors have an F-test p-value less than $\alpha$ (Montgomery and Runger 2007).
4.1 Method for Stepwise Regression (Loucks 2003)

1. Compute F-number on all individual regressor variables not included in the model.
2. Select the variable with the lowest p-value.
   a. Value of variable has to be less than 0.05 to be selected.
3. Compute F-number on all regressor variables within the model.
4. Remove any regressor variable with p-value larger than $\alpha_2$.
   a. $\alpha_2$ has been set to 0.15 by SAS.
5. If no regressor variables are removed, restart from Step 1.

4.2 Method for Forward Selection (Loucks 2003)

1. Compute F-number on all individual variables.
2. Select the variable with the lowest p-value.
   a. Value of variable has to be less than 0.05 to be selected.
3. The first variable selected in kept and of the remaining variables, the one with the smallest p-value is added.
4. If the p-values of all remaining variables are larger than $\alpha$, then stop the forward selection.
5. If p-values are less than $\alpha$, continue the forward selection.
6. Continue this process till p-value calculations are greater than $\alpha$. 
4.3 Method for Backward Elimination (Loucks 2003)

1. Start with a model that includes all regressor variables.
2. Calculate the individual F-number for all variables in the model.
3. Remove the variable with the highest p-value.
4. Calculate the F-number on all remaining variables.
5. Remove the variable with the highest p-value.
6. Continue this procedure till all remaining variables have p-values less than \( \alpha \).
CHAPTER V

REASON FOR A NEGATIVE INTERCEPT

After obtaining a negative intercept for some of the preliminary models, I decided not to force the intercept to pass through the origin. The reason was that forcing the intercept hinders the model and gives a deceptively higher R-squared value.

The increase in R-square was an artificial increase. Simple linear regression tests whether the null hypothesis of $\beta_1=0$ is true for equation;

$$y = \beta_0 + \varepsilon = \mu + \varepsilon$$ \hspace{1cm} (5.1)

The values of $\hat{Y}_i$ are calculated using $\beta_1$ and compared to $Y_i$. The $SS_T$ equals Eq. 8.2:

$$SS_T = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$ \hspace{1cm} (5.2)

If there is no intercept, $\beta_1=0$ and the linear model equals Eq. 8.3:

$$y = \varepsilon = 0$$ \hspace{1cm} (5.3)

The $SS_T$ then equals Eq. 8.4:

$$SS_T = \sum_{i=1}^{n} (Y_i)^2 = \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2 + \sum_{i=1}^{n} (\bar{Y}_i)^2$$ \hspace{1cm} (5.4)

R-squared equals $SS_R/SS_T$. The goal behind regression is to minimize the sum of squares. With no intercept, there are larger $SS_R$ and $SS_T$ values are larger, and the R-squared value will be higher because of the difference between the mathematical operations on each side of the equals sign. High $SS_R$ and $SS_T$ value indicate a bad model. For a model with an intercept, the $SS_R$ and $SS_T$ values and R-squared value are lower. Lower $SS_R$ and $SS_T$ indicate a better model (Truong).
CHAPTER VI
APPLYING THE METHODOLOGY FOR WELLS IN
NORTH CENTRAL TEXAS

To test the proposed methodology, I constructed a model using SAS to find the number of days required to drill a well in northern central Texas from a total of 57 observations. Each observation was the number of days required to drill, test, case, and cement a particular section of a well. Initially 13 variables were tested, technical limit (referred as estimated days), depth, bottomhole temperature (BHT), inclination (inc), mud weight (MW), fracture pressure (FP), pore pressure (PP), and the average, maximum, and minimum difference between FP-MW and MW-PP.

All the independent variables were collected using daily drilling reports or daily mud reports. I ran two rounds of tests. The first round of tests used all 13 variables. An issue arose when testing the models that had FP or PP in the model: estimated values of FP and PP available prior to drilling were higher than actual FP or PP encountered within the wellbore. This caused all the models to overestimate the number of days required to drill. Since the only information available prior to drilling a well is analogous data (values derived from adjacent wells) I removed FP, PP, and MW from the analysis, reducing some of the uncertainty factor of the second test using technical limit (referred as estimated days), depth, BHT, and inc.

SAS’s output contains many results. Though all results of analysis of variance and univariate help describe how a model fits the data, some test results are more meaningful to a model’s validation than others. To validate a model, I analyzed seven basic criteria: residual plot, Shapiro-Wilk residual p-values value, chi-squared p-values, variance inflation factor, R-squared, Mallow Cp (if applicable), and F-test p-value. The following criteria for each test indicate whether any assumption needed for regression analysis has been violated.
- Residual plot should have no discernable pattern and the residual points should be evenly distributed above and below the origin x-axis.
- Both the Shapiro-Wilk p-value and chi-squared p-value need to be above an $\alpha$ value of 0.05 not to reject their null hypothesis that the residuals are normally distributed.
- Variance inflation factor values need to be below 10 to indicate no inter-correlation behavior between variables.
- R-square ideally equals 1. The closer the value to 1, the better the match between predicted and observed values. This comparison may only be done when models are similar to one another since models with more variables will have higher R values. The reasoning why some poor model may have a higher R-squared values has been discussed in Section 5.
- Mallow Cp values are only calculated for stepwise regression, forward selection, and backward elimination. The Mallow Cp value should equal one plus the number of regressor variables within the model.
- F-test p-value have to be less than $\alpha$ for every variable to be considered significant.

This order established a systematic approach to obtain the most significant model.
CHAPTER VII
SAS RESULTS FOR REGRESSION ANALYSIS OF
NORTH CENTRAL TEXAS WELLS

Eleven runs of the regression methodology were applied to the wells drilled in north central Texas. To better compare the results of each run, I show SAS results in Section 7, followed by the history match done in Microsoft Excel. SAS code, output, and Microsoft Excel charts and tables are found in the Appendix.

The first model (Eq. 7.1) found estimated days and bottomhole temperature (BHT) to be significant regressor variables by all three regression techniques: stepwise regression, forward selection, and backward elimination.

\[ y_A = -16.84055 + 0.87604 \times T_e + (0.09255 \times T_{bh}) \]  
……………………………………(7.1)

To test the significance of the model, I verified all parameters in the validation method. The residual plot for Model A fans out and clusters towards the middle (Fig 7.1). Residual plots should have no discernable pattern and be evenly scattered along the origin horizontal axis.
Model A’s Shapiro-Wilk p-value was less than 0.0001. If the Shapiro-Wilk p-value does not equal a value greater than 0.05, then the null hypothesis of the Shapiro-Wilk test—that the residuals are normal—gets rejected. A Q-Q plot helps visualize how the data deviates from the normal distribution. A normal residual distribution will have the data points on a linear line with a slope equal to the standard deviation and intercept equal to the mean. Model’s A Q-Q plot slope (Fig. 7.2) was less than the residual standard deviation slope line, confirming the Shapiro-Wilk test result that the residuals are not normal.
Since the Shapiro-Wilk test was rejected for Model A, all results of the regression are not significant (Nau 2005).

Though all three previous regressions obtained the same significant regressor variables for Model A, it appeared as though depth would be a more significant regressor variable than BHT. To test this conclusion I selected actual days, estimated days, and depth. Model B (Eq. 7.2) incorporates the desired variables, allowing SAS to calculate the regressor coefficients and the significance of each variable.

\[ y_B = -9.88984 + (0.90118 \times T_e) + (0.00150 \times D) \] ..............................................(7.2)

The residual plot for Model B (Fig. 7.3) appeared almost identical to that of Model A.
Fig. 7.3-Model B’s Residual Plot. Actual Days = f(Te,D)

All the key results were similar to Model A’s, leading to the same conclusion: Model B does not adequately model the data. When observing the residual plots of both Model A and B, a residual outlier with a value of 130 stands out. That point corresponds to a problem well section that took many days to resolve. Though a well may encounter serious issues that take a long time to remediate, they occur sporadically. To improve the model’s ability to predict the number of days required to drill a well, I removed that data.

Model C (Eq. 7.3), resulted from a new dataset created without the problem section. The regressor variables were inputted into SAS and the following regression coefficients calculated.

\[ y_C = -5.28257 + (0.87691 \times Te) + (0.00104 \times D) \] ...............................(7.3)

The residual plot of Model C (Fig. 7.4) has a discernable bow-tie pattern, but the residual plot appears better than Model A or B’s residual plot.
The Shapiro-Wilk p-value was greater than 0.05, so the null hypothesis—that the residuals are normal—was not rejected. The Q-Q plot for Model C (Fig. 7.5) had a normal trend but deviation occurred at lower percentiles.
The chi-squared p-value was barely above the rejection value of 0.05. Chosen regressor variables had low inflation variance factors, and the F-test p-value was less than 0.0001. The R-squared value was 0.71.

Using previous regressor variables selected by stepwise regression, forward selection, and backward elimination, I found new regressor coefficients by performing least squares regression with the new data set. Model D (Eq. 7.4) had many validation parameters similar to Model C.

\[ y_D = -9.98231 + (0.86113 \times T_e) + (0.06342 \times T_{bh}) \] (7.4)

Models C and D passed all the required validation parameters. A reason that Models C and D may lack significance was the residual plots scatter and low chi-square p-values. Residual plots of both models showed a slight double bow pattern while chi-squared values were near the rejection criteria of 0.05. To better understand why the models barely passed the validation process, I examined the dynamics between each individual variable and the dependent variable. The goal was to see if there was a linear relationship between the dependent variable and independent variables. Plots of actual days versus all independent functions were done. Shown below in Fig. 7.6, 7.7, and 7.8, are three key plots indicating the relationship between Actual Days and Estimated Days, Depth, and BHT. Figure 7.6 shows the linear relationship found between Actual Days verse Estimated Days. Both Depth and BHT appeared to have an exponential relationship when plotted against Actual Days as shown in Fig 7.7, and 7.8.
Fig 7.6–Actual Days vs. Estimated Days

Fig 7.7–Actual Days vs. Depth
To linearize the relationship, transformation of certain variables were done. Shown below are all the transformations done to the dataset. In statistics, the use of log is synonymous with natural log.

- \( \log_{\text{actual days}} = \log(\text{actual days}) \)
  - The log command in SAS represents natural log
- \( \exp_{\text{estimated days}} = \exp(\text{estimated days}) \)
- \( \text{estimated days}^2 = \text{estimated days}^2 \)
- \( \log_{\text{estimated days}} = \log(\text{estimated days}) \)
- \( \log_{\text{Depth ft}} = \log(\text{Depth ft}) \)
- \( \log_{\text{TEMP BHT F}} = \log(\text{TEMP BHT F}) \)
- \( \text{Depth by Est days} = \text{Depth ft} \times \text{Estimated Days} \)
- \( \text{Est Days by BHT} = \text{Estimated Days} \times \text{Temp bht} \)
- \( \log_{\text{Temp BHT by Est}} = \log(\text{TEMP BHT F}) \times \text{Estimated Days} \)
- \( \log_{\text{Depth by Est Days}} = \log(\text{Depth ft}) \times \text{Estimated Days} \)

After transforming the dataset, plots were done to test the linearity of the variables.
Fig 7.9–ln(Actual Days) vs ln(Estimated Days)

Fig 7.10–ln(Actual Days) vs ln(Depth)
Fig 7.11–ln(Actual Days) vs Depth

Fig 7.12–ln(Actual Days) vs BHT
The ln(actual days) vs. ln(estimated days), shown in Fig. 7.9, had a linear relationship. Fig 7.10, ln(actual days) vs. ln(depth) did not have a linear relationship. ln(actual days) vs. depth and ln(actual days) vs. BHT, shown in Fig.7.11 and 7.12, also showed linear relationships. Figure 7.13 and 7.14 did not produce a linear relationship for the specific transformations between ln(actual days) vs. inclination and ln(actual days) vs. (estimated
The key variables for the new dataset were ln(actual days), ln(estimated days), depth, and BHT.

I performed stepwise regression, forward selection, and backward elimination to the new dataset. The three regressions did not return the same regressor variable combinations. Model E (Eq. 7.5), was found through stepwise wise regression.

\[
\ln(y_E) = 0.44017 + (0.60152 \times \ln(Te)) + (0.00006721 \times D) \quad \ldots (7.5)
\]

The residual plot for Model E (Fig. 7.15) had not improved from Models C and D.

The Shapiro-Wilk p-value of 0.0524 did not reject the null hypothesis. The Q-Q plot had noticeable deviation at the lower and higher percentiles with a staggered pattern from the 50th to 90th percentile.
Though the Shapiro-Wilk almost got rejected, the chi-squared p-value increased to 0.14. The variance inflation factor was low for both regressor variables, and the R-squared value equaled 0.7995. The F-number p-value was also very small.

Forward selection and backward elimination chose the same regressor variables. Model F (Eq. 7.6), included all variables being tested.

\[
\ln(y_F) = -0.72931 + (0.57459 \times \ln(Te)) - (0.0001936 \times D) + (0.01549 \times Tb) + (0.00743 \times Inc)
\] ..........................(7.6)

The residual plot of Model F still had the double bow appearance of models Eq. 7.3, Eq. 7.4, and Eq. 5.5. The Shapiro-Wilk p-value and chi-squared p-values were high. Mallow Cp with 4 regressor variables had a value of 5. The R-squared value of 0.82 indicated a very good fit. The problem with Model F was that the variance inflation factor for regressor variables depth and BHT was 130.5. Since all variables have to be independent for regressions to be useful, Model F was unacceptable.

Model F indicated the strong interrelationship between depth and BHT. All regressor variables must be independent of one another. Model E selected the regressor variable depth, while previous regressions selected BHT. To test the significance of
BHT, I analyzed the dependent variable actual days and regressor variables estimated
days and BHT. The returned regressor coefficients are seen in Model G (Eq. 7.7).

\[ \ln(y_G) = 0.18264 + (0.58106 \times \ln(Te)) + (0.00408 \times Tbh) \ldots \ldots \ldots \ldots \text{(Eq. 7.7)} \]

Model G’s residual plot (Fig. 7.17), had a better scatter pattern than the previous
models, with scatter evenly distributed along the horizontal axis.

![Residual Plot](image)

**Fig. 7.17**-Model G’s Residual Plot. \( \ln(\text{Actual Days}) = f(\ln(\text{Te}), \text{Tbh}) \)

Even with the better residual plot, the Shapiro-Wilk p-value rejected the null hypothesis.
The Q-Q plot for Model G (Fig. 7.18), does not follow the normal unit slope at the lower
percentiles and staggers stepwise at higher percentiles.
Though the chi-squared p-value, R-squared value, and F-number all showed good trends, Model G does not pass the Shapiro-Wilk test and was considered inadequate.

Looking at previous regressions, the three most significant regressor variables were estimated days, depth, and BHT. Models C, D, and E barely passed the validation process. The next step required further analysis of the transformed variables in an attempt to obtain a better model. The variables actual and estimated days and depth were selected and only actual days was kept transformed to ln actual days. The regression coefficients of Model H are shown in Eq. 7.8.

\[
\ln(y_H) = 1.30916 + (0.02003 \times Te) + (0.00010005 \times D) \tag{7.8}
\]

Model H’s residual plot (Fig. 7.19), had a similar scatter to Model G’s residual plot. It had better scatter distribution than Models C, D, and E.
Fig. 7.19-Model H’s Residual Plot. \( \ln(\text{Actual Days}) = f(Te,D) \)

The Q-Q plot appeared better than previous models, with a reduction in the stepwise stagger that was present for Model G. The central points fluctuate more above and below the unit slope, while at the top and lower percentile deviation still occurred.

Fig. 7.20-Model H’s Q-Q Plot. \( \ln(\text{Actual Days}) = f(Te,D) \)

Both the Shapiro-Wilk p-value of 0.25 and chi-squared p-value of 0.0807 do not reject the null hypothesis. Though the chi-squared p-value was small, it still passed the cutoff.
value of 0.05. Variance inflation factor of 1.35 confirmed that all variables were independent. R-squared was high with a value of 0.787. The F-test p-value was also below 0.05. Model H passed all validation criteria.

Even though Model H passed all validation criteria, an optimal model, if possible, would not have validation parameters that barely pass. I continued testing to try and find a model with higher validation parameters. BHT has often been used in previous models and was significant in Model D. Another test was done using ln Actual Days, Estimated Days, and BHT. Model I (Eq. 7.9), was the result of the analysis.

\[ \ln(y_I) = 0.88750 + (0.01893 \times T_e) + (0.00598 \times T_{bh}) \]  

Model I had a near-identical residual plot to that of Model H. Model I’s univariate analysis also calculated values similar to Model H. Model I did not reject the Shapiro-Wilk or chi-squared tests. Variance inflation factors were low. R-squared had a value of 0.787, the same as Model H. The only concern was that the chi-squared value was 0.073, a decrease from Model H’s chi-squared value of 0.0807.

Although models H and I are both valid models, it would be preferable to have a higher chi-squared p-value. To attempt to find a higher chi-squared, Model H and I were modified by transforming regressor variables Depth and BHT to ln Depth and ln BHT. The resulting regressor coefficients of Model J are shown in Eq. 7.10 and Model K in Eq. 7.11.

\[ \ln(y_J) = -4.1396 + (0.02007 \times T_e) + (0.72552 \times \ln(D)) \]  

\[ \ln(y_K) = -4.98991 + (0.01828 \times T_e) + (1.36169 \times \ln(T_{bh})) \]

Model J’s residual plot (Fig. 2.11), appeared slightly more homogenous than Models G, H, and I. The scatter pattern looked evenly distributed along the horizontal axis.
Model J’s Q-Q plot (Fig. 7.22) improved from Model’s I Q-Q plot; it does not stagger along the unit slope. The lower percentiles did not deviate as in previous models, while deviation still occurred at the upper percentiles.

The Shapiro-Wilk and chi-squared tests did not reject the null hypothesis. The Shapiro-Wilk p-value was 0.16 and the chi-squared p-value was 0.25. The variance
inflation factor was low, and the F-test p-value was very low. The R-squared value was 0.775. Model J had the desired validation criteria.

Both the residual plot and Q-Q plot of Model K appeared near identical to Model J’s. Model K’s Shapiro-Wilk and chi-squared tests were not rejected. The chi-squared value did reduce to 0.11 from Model J’s chi-squared value of 0.25. The variance inflation factor was low and the F-test p-value was very small. The R-squared value was 0.797. Model K has a lower chi-squared value than Model J but a higher R-squared value. Both Models J and K are valid models to use for predicting the amount of time required to drill a well.

A final test was done to see if combining the variables in each valid model would result in a significant positive change. New regressor variables were made by combining estimated days and ln depth into estimated days by ln depth. Estimated days and ln BHT were combined into estimated days by ln BHT. No positive significant results were obtained.
CHAPTER VIII
HISTORY MATCHING OF SIGNIFICANT MODELS

Seven models-Eq. 7.3, 7.4, 7.5, 7.8, 7.9, 7.10, and 7.11-passed all validation parameters. The next step checked the accuracy of each model by comparing predicted data to historical data.

\[ \text{Percent Error} = \frac{\text{Actual} - \text{Predicted}}{\text{Actual}} \]  
\[ \text{(Eq. 8.1)} \]

By looking at the difference between actual and predicted and by comparing the percent error per well section, the best model was chosen.

Shown below, Table 8.1 illustrates an overall tendency to overestimate or underestimate and how many times each model predicted a value that was +/- 10% of the actual number of days required to drill a well interval.

<table>
<thead>
<tr>
<th>Per Well Section</th>
<th>Technical Limit</th>
<th>Model C</th>
<th>Model D</th>
<th>Model E</th>
<th>Model H</th>
<th>Model I</th>
<th>Model J</th>
<th>Model K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over or Under Estimated</td>
<td>2.57</td>
<td>-16.06</td>
<td>-4.41</td>
<td>-4.11</td>
<td>-4.12</td>
<td>-4.07</td>
<td>-4.47</td>
<td>-4.02</td>
</tr>
<tr>
<td>+/- 10% out of 57 intervals</td>
<td>11</td>
<td>10</td>
<td>16</td>
<td>15</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>14</td>
</tr>
</tbody>
</table>

The parameter, Overestimated or Underestimated, was found by summing all percent errors for every well section. A smaller number indicates a smaller difference between actual and predicted values. Table 8.2 compares the same properties as in Table 8.1 but the individual well sections have been summed.
Both Table 8.1 and Table 8.2 showed that Model D best predicted the actual data. Overall Model D had a 47% success rate for comparing the total number of days to complete a well. The accuracy rate decreased to a 17.5% success rate when measuring individual well sections. Though the original estimate for technical limit had the lowest difference between the number of days per individual well interval, overall original estimate for the total number of days required to drill a well was greater than all models and never had an estimate within the +/- 10% time difference. The original estimates do not take into account the increase time required to drill the deeper hole sections while the predictive model do.

No model gave the optimum +/-10% time difference for every well section. Using regression techniques to find Model D, there was an improvement on the prediction of the total number of days to drill a well. To improve upon Model D, more observations would be required and the regressions redone.

Table 8.2
Summation of Percent Error and Number of Times the +/-10% Objective was Obtained per Well

<table>
<thead>
<tr>
<th>Per Well</th>
<th>Technical Limit</th>
<th>Model C</th>
<th>Model D</th>
<th>Model E</th>
<th>Model H</th>
<th>Model I</th>
<th>Model J</th>
<th>Model K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over or Under Estimated</td>
<td>4.69</td>
<td>-3.36</td>
<td>-0.33</td>
<td>0.86</td>
<td>0.23</td>
<td>0.44</td>
<td>0.45</td>
<td>0.58</td>
</tr>
<tr>
<td>+/- 10% out of 17 wells</td>
<td>0</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>


Kaiser and Pulsipher (2007) found the generalized function model through regression in four basic steps. Step 1 required the selection of independent descriptor variables. Step 2 defined the bounds of individual parameters for each well. Step 3 constructed a regression model from the well data and tested the model’s coefficients for significance. Step 4 maintained all significant variables and factors with p-values less than 0.05. As stated previously, the five step methodology introduced in this thesis was similar to Kaiser and Pulsipher’s: Step 1 created a database. Step 2 performed initial statistical regression on the original data. Step 3 ensured that the models are valid by performing univariate analysis. Step 4 history matched the model’s response to actual observed data. Step 5 repeated the procedure till the best model was found. If no significant models were found after performing the calculations using the initial dataset, data manipulation was required.

Step 1 creates a database. If a dataset of all wells with all parameters were to be done, it would cause significant variability to the output due to the heterogeneity of the data (Kaiser and Pulsipher 2007). To avoid heterogeneity problems, the first step requires defining a specific question. A dataset should represent a group of wells or well sections from similar geology, depth, and drilling method (Adams et al. 2009). In this thesis, data was used from wells that were drilled in the same fashion, using a top drive, and located in the same county in north central Texas.

The number of observations available determines the number of independent variables that may be analyzed. According Adams et al. (2009), 100 wells was the bare minimum for statistical accuracy and 200 or 300 wells were optimal. Jablonowski and MacEachern (2009), claim that 30 observations allow for enough significance. An observation may be the data obtained from a certain well segment or the total values for a single well. With a small dataset, 10 independent variables can be tested and for larger datasets, 20 independent variables (Kaiser and Pulsipher 2007). Noerager et al. (1987)
produced a model using 640 wells in the North Sea that included 489 platform wells and 51 subsea wells within a time span of 10 years, from 1976 to 1986. They tested 9 variables. Even with a large dataset with detailed measurements, the model did not obtain the +/- 10% time-deviation goal. They attribute the large amount of scatter to operational differences among the wells. If they had a more specific initial question, like “how long does it take to drill a subsea well in the North Sea using data from 1981 to 1986?”, a more accurate model could have been possible. The dataset form north central Texas created to test the regression methodology had 57 observations with 12 independent variables.

In Step 2, boundaries were set to normalize the data. Normalizing data has many advantages for flat time activities and the addition of new descriptive variables. An example of a new descriptive variable would be, assigning a value of 1 if bottomhole temperature got above 300°F, 0 if it did not. Though there are many advantages standardizing variables, normalizing rate dependent variables, such as drilling and tripping, can cause errors in the model (Adams et al. 2009). In this thesis the dataset had some variables manipulated to linearize key interactions, but none were normalized.

Many software programs perform the necessary calculations for Step 3. Some options are as accessible as Microsoft Excel, other programs are dedicated to statistics, and some are specific to drill-time prediction. SAS and R were chosen to perform the statistical calculations in this project. Both offer the flexibility to test many parameters that specific prediction software may not allow or does not calculate. Step 4 keeps all significant variables with an F-test p-value less than 0.05 within the final model.

Noerager et al. (1987) produced two deterministic nonlinear models that contained dimensionless variables, penalty factor in learning curve, rate of learning, half-life of learning curve, and annual improvement factor (Noerager et al. 1987). The authors claimed that 21 days, or 35% of the mean, was the smallest deviation possible by any method of predicting the time required to drill and complete an individual well. Table 9.1 shows the standard deviation of actual and predicted values for the well drilled in north central Texas.
Table 9.1-Average Standard Deviation

<table>
<thead>
<tr>
<th>Days</th>
<th>Technical Limit</th>
<th>Model C</th>
<th>Model D</th>
<th>Model E</th>
<th>Model H</th>
<th>Model I</th>
<th>Model J</th>
<th>Model K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per Well</td>
<td>7</td>
<td>6.64</td>
<td>5.46</td>
<td>5.54</td>
<td>6.91</td>
<td>6.75</td>
<td>6.38</td>
<td>6.10</td>
</tr>
<tr>
<td>Section</td>
<td>28</td>
<td>29</td>
<td>32</td>
<td>34</td>
<td>37</td>
<td>36</td>
<td>34</td>
<td>34</td>
</tr>
</tbody>
</table>

That same year, Thorogood (1987) created a model using 85 wells that includes a linear formula for daily progress rate and additional values for flat-time unit operations. When the standard deviation of the drilling progress and unit operations were combined the total standard deviation equaled 10% of the total time of Noerager’s model, which tried to incorporate too many parameters that are not easily measured, such as learning curves. Working with a more targeted approach, as Thorogood had done, resulted in a more accurate, straightforward model that had better predictive capabilities. The models created using SAS and R emulated Thorogood’s strategic approach to deterministic estimator modeling.

Many operators use probabilistic instead of deterministic modeling. Some engineers claim that deterministic estimates tend to be optimistic and that many decisions are made ignoring the prediction errors (Loberg et al. 2008). In 1993, Peterson et al calculated drill time predictions using @Risk software to perform Monte Carlo simulations. In their model, total time equals total problem-free time plus total problem time, which required that the probability distribution functions of depth variation, drilling and evaluation problem days, and the problem-free drilling and evaluation days be chosen. Due to the large uncertainty present while drilling a well, deciding which probability distribution function to select may be difficult. Softwares are available to assist with the selection process, and some have preselected probability distribution functions if data is limited. To allow for a new field with a small datasets, my
deterministic methodology gave a more direct approach to obtain a time-predictive model.

Peterson et al. (1993) gave two examples, a 20,090 ft and 17,907 ft well. Using a probabilistic approach and the @ Risk software, they calculated problem-free days and problems day and validate their approach by showing that the difference between predicted and actual drill time was only 3 days. In the north central Texas example, Model D estimated three wells, 16,500 ft, 19,500 ft, and 12,300 ft, with less than 2 days’ error for the 17 wells tested. Peterson et al. does not describe the accuracy of the method when tested in a larger scale.

Kaiser and Pulsipher (2007) and Jablonowski and MacEachern (2009) calculated their predictor model using regression analysis. Jablonowski and MacEachern (2009) used the standard deviations of the model’s results to demonstrate the effectiveness of their model; Jablonowski and MacEachern (2009) showed R-squared values, standard error, and t-test values to indicate the adequacy of their model. The mean for the number of days required to drill a well in the Gulf of Mexico was 34.7 days with a standard deviation of 19.2 days using Louisiana Kaiser and Pulsipher (2007) model. Though both research teams gave statistical parameters to describe their models, they still lack the accuracy presented with the five-step methodology. To standardize the selection process and testing of model significance, I created a list of criteria for model selection and testing.

Jablonowski and MacEachern (2009) deterministic approach to allowed probabilistic range of values by changing the confidence interval. Using 66 observations and 30 independent variables, they developed a model with an R-squared value of 0.86 using a 95% confidence interval, whereas I used 57 observations with five independent variables. Using Model D as the optimal model with a 95% confidence interval, the R-squared value was 0.716. A possible improvement on predicting the time required to drill a well in north central Texas may occur with the introduction of new independent variables, such as hole size or indicating a variable for certain formations, similar to the 30 variables Jablonowski and MacEachern (2009) tested.
No article had proven the significance of their model. A model found using statistical regression needs to make sure certain key statistical test indicates whether the assumption required to create the model are accurate. Previous reported models included a calibration to fit historical data when using regression techniques. If certain key parameters such as Shapiro-Wilk residual normality test do not hold true, F-test p-values may indicate significant variables, but in actuality the model would be flawed and the final result may be a calibrated incorrectly. To avoid that problem, I developed a validation check list using results from univariate calculations. After the creation of a model by analysis of variance, every model had to pass the seven key parameters to decide if the model has mathematical significance. The seven models shown in Table 10.1 all passed the seven criteria. Historical matching of the data with predicted values of valid models then determined which model has the greatest predictive accuracy.
CHAPTER X
CONCLUSION

Deterministic models have a single value response, so a deterministic approach was chosen for this project because a linear model may be found with small amounts of data. The five step methodology introduced in this thesis includes validation parameters that ensure the required statistical assumptions for regression are followed. Step 1 created a database. Step 2 performed initial statistical regression on the original data. Step 3 ensured that the models are valid by performing univariate analysis. Step 4 history matches the model’s response to actual observed data. Step 5 repeated the procedure till the best model was found.

When the applied methodology was tested on the example case using SAS and R, the overall objective of having the predicted model within 10% of the actual time to drill a well was only achieved 50% of the time using Model D. When individual well sections were predicted, the 10% objective was achieved only 28% of the time. To improve the model, different options may be considered.

The first option would be to create a more homogenous database by redefining the example case. By changing the objective question, “How many days are required to drill a vertical well in north central Texas?”; the data could be more concise and may improve the accuracy of the model. A second option would be to include more data by including other analogous wells near the area of the well tested. The third option would be changing the independent variables. By removing the estimated days and keeping the more physical descriptive variables, a better model may be found.

Having a methodology that may be quickly implemented allows for greater flexibility. The most time-consuming labor was creating the database. After applying the methodology to the example case, greater accuracy was found than previously predicted values. The methodology helped improve prediction estimation but did not ensure the +/- 10% time deviation from actual observations.
REFERENCES


Venables, W.N., Smith, D.M., and R Development Core Team. 2002. An Introduction to
R: Notes on R: A Programming Environment for Data Analysis and Graphics,
0954161742.

Statistics Lessons: East Carolina University.
http://core.ecu.edu/psyc/wuenschk/docs30/Skew-Kurt.doc. Downloaded July,
2010.
APPENDIX A

PROCEDURE FOR PERFORMING ANALYSIS OF VARIANCE IN SAS

The chapter explains the basic code necessary to perform variable analysis within SAS. The process of transferring a dataset from Microsoft Excel to SAS 9.2, the basic format and commands required for stepwise regression, forward selection, and backward elimination will be shown and the code for least squares regression are given for testing of individually selected independent variables. To validate the response of SAS, the commands for univariate calculations are given.

A.1 Transferring Data from Microsoft Excel to SAS

The procedure starts with the initial dataset in Microsoft Excel. Figure A.1 shows how individual variables and values are arranged for a given dataset. Short, concise names are recommended when labeling individual variables. At the end of this section, the reason for short concise names has been given.

![Example Dataset from North Central Texas Wells](image)

Fig.A.1- Example Dataset from North Central Texas Wells
The next step is to save the worksheet as a comma-separated file, .csv file type. Though SAS can import workbooks from Microsoft Excel, certain issues arose when trying to import the data via a virtual access using the internet. There were no issues when the dataset was a .csv file, and it imported with no problems using a virtual access.

Described below is the procedure for importing a dataset into SAS 9.2. This procedure also works for older versions of SAS.

1. Click on the File tab
   a. Under the File tab, click on Import data.

![Fig. A.2-Initial Steps to Import a Dataset into SAS](image-url)
2. Click on the square marked, Standard Data Source.
3. Click on the pull-down bar marked Selected a data source from the list below.
4. Click on the fourth option, Comma-Separated Values (*.CSV).
5. Then hit Next on the bottom of the window.

6. Find your data file by clicking on the Browse tab.
7. Once the data file has been selected click Next.
8. Under library, make sure that the pull down bar has Work selected.
   a. Under Member, create a name for your dataset. This name will be used to access the dataset when writing the program necessary to perform calculations on SAS
After naming the dataset, click on Finish.

Fig. A.5 – Last Step to Import a Dataset into SAS
10. If SAS is readily available, you can construct personal user SAS libraries and store information.

After the dataset has been inputted, you can see the individual names of each variable under the, Log-(Untitled) window. Once the Log-(Untitled) window has been selected, scroll down to find the Input area. All variable names will be shown below. When working with SAS, specific variables have to be written as shown within the Log worksheet. For that reason, short, concise names are recommended when creating the initial dataset in Microsoft Excel. The symbol $ means that the variable has nonnumerical character values associated with it. When a variable has an underscore within the name, it represents a space that was given when the variable name was written within Microsoft Excel.

Fig. A.6-Variable Names Seen within the Log Window
A.2 Basic Code Necessary to Perform SAS Regression and Model Validation

The basic commands to perform the different calculations within SAS are explained below. All source code needs to be written within the Editor page. SAS requires an exact format when writing the open source code. Semicolons have to be placed at the end of every command. The floppy icon on the upper left-hand side of the toolbar saves the written code on the Editor page.

The first lines of code open up the dataset saved within SAS workspace:

```
proc contents data=filename;
run;
```

SAS recognizes individual variables by the space between them. Commands are differentiated by the use of a semicolon. A single command may take up multiple lines. Some backslashes appear below the model equation because there of a lack of space within the margins.

The command `spec` and `vif` are used for model validation. The code `spec` outputs analysis of variance results and `vif` outputs variance inflation factor. A full description of outputs can be found in CHAPTER III of this thesis.
The models presented have used the following basic code for stepwise regression, forward selection, backward elimination, and least square regression.

**Stepwise Regression**

```
proc reg data=filename;
model dependent_variable = variable_A variable_B variable_C
/selection=stepwise spec vif;
run;
quit;
```

![Fig. A.8-Example SAS Input for Stepwise Regression](image)

**Forward Selection**

```
proc reg data= filename;
model dependent_variable = variable_A variable_B variable_C
/selection=forward spec vif;
run;
quit;
```

![Fig. A.9-Example SAS Input for Forward Selection](image)
Backward Elimination

```sas
proc reg data= filename;
model dependent_variable = variable_A variable_B variable_C /selection=backward spec vif;
run;
quit;
```

Fig. A.10-Example SAS Input for Backward Elimination

Least Squares Regression

```sas
proc reg data= filename;
model dependent_variable = variable_A variable_B variable_C /spec vif;
output out=results r=resid p=fits;
run;
quit;
```

Fig. A.11-Example SAS Input for Least Squares Regression
The univariate code and residual plot test the normality of the residuals. The significance of the univariate output can be found in CHAPTER IV of this thesis. Residual plots are discussed in CHAPTER VII.

This model uses the following basic codes for univariate analysis and residual plots.

**Univariate**

\begin{verbatim}
proc univariate data=results normaltest ;
var resid;
probplot resid / normal(mu=est sigma=est);
run;
\end{verbatim}

**Residual Plot**

\begin{verbatim}
proc gplot data=results;
plot resid*fits;
run;
\end{verbatim}

Fig. A.12-Example SAS Input for Univariate Calculations
A recommended method to test the dataset would be to run the regression followed by the univariate calculation. This helps keep the output of the program organized with the results of the regression followed by the test of the residual normality. The title command places a header on the graphs. This helps differentiate if many graphs are made. The title command has to be written as shown in the example. The stepwise regression example below illustrates this concept.

```
proc reg data=filename;
model dependent_variable = variable_A variable_B variable_C
    /selection=stepwise spec vif;
run;
quit;

proc univariate data=results normaltest ;
var resid;
title 'Q-Q Plot: Step Regression of Dependent Variable and Variables A, B, C';
probplot resid / normal(mu=est sigma=est);
run;

proc gplot data=results;
title 'Residual Plot: Step Regression of Dependent Variable and Variables A, B, C';
plot resid*fits;
run;
```
Specific models can be entered and tested using least squares regression. If an individual wants to have a model with certain variables, the example below illustrates the method to perform that analysis.

```sas
proc reg data= filename;
model dependent_variable = variable_A variable_C variable_F /spec vif;
output out=results r=resid p=fits;
run;
quit;

proc univariate data=results normaltest;
var resid;
title 'Q-Q Plot: Least Squares of Dependent Variable and Variables A, C, F';
probplot resid / normal(mu=est sigma=est);
run;

proc gplot data=results;
title 'Residual Plot: Least Squares of Dependent Variable and Variables A, C, F';
plot resid*fits;
run;
```
The code above allows linear regression to be done on the dataset. To run the written commands, click on the icon of the running man on the tool bar. Any errors will be identified within the Log-(Untitled) window. To clear the Log or Output window, select the window and click Edit, clear text. The Edit icon can be found in the menu bar. As a shortcut to execute the same command, hold the control button and E at the same time.
A.3 Manipulation of Variables and Dataset

Outliers

If the resulting models do not prove to be significant, they may include a single or multiple outliers. Caution must be taken when deciding to remove outliers. Though a single outlier might be a unique event, multiple outliers may be a trend instead of an anomaly. There are two methods to remove outliers: the original dataset may be changed and uploaded again into SAS, or a new dataset may be created within SAS without the outlier/s. The following code demonstrates how to create a new dataset by removing unwanted data points.

```sas
data newfilename;
set filename;
where variable_name ne #;
run;
```

Fig. A.16- Example SAS Input for Removal of a Data Point

The new dataset removes any data point from a given variable that has a value greater than the numerical value given to #. Many different lines of code to perform basic manipulations of the dataset may be found at SAS’s homepage (2010) or its online user guide (1999).
A.4 Variable Trends

Plotting the dependent variable against the independent variable helps identify if the relationship follows a desirable linear trend. Shown below is the command to make a plot in SAS.

```sas
proc gplot data=filename;
title 'Independent Variable vs. Variable A';
plot independent_variable*variable_A;
run;
```

Fig. A.17- Example SAS Input for Creating a Plot

To create multiple plots, just repeat the previous code with the desired variables.

If a plot shows a nonlinear relationship, it may be possible to linearize the relationship by manipulating certain variables. First time users of SAS should use caution when using the `log` command. In SAS, the `log` command represents the natural log. Log10 calculates base 10 logs. New datasets are recommended whenever variables are manipulated. An example of variable manipulation appears below.

```sas
data newfilename;
set filename;
lnindependent_variable=log(independent_variable);
exp_variable_B=exp(variable_B);
variable_D_sq= variable_D**2;
log_variable_F=log10(variable_F);
run;
```
The new dataset, newfile, has four new variables that were not present in the original dataset newfile. Given names of the manipulated variables are left of the equal sign. The names were chosen to describe the action performed.

Appendix E contains the written code used in SAS for the example dataset of north central Texas. This may be used as a guide or template for future data regressions.
APPENDIX B

R SOFTWARE CODE

SAS’s cost can deter usage of the software. R software performs statistical operations and may be downloaded for free from www.r-project.org (2010). The same methodology and concepts used within SAS are applicable for R. A basic review of the code necessary to operate the software is explained below.

B.1 Transferring Data from Microsoft Excel into R

R does not import Microsoft Excel files; instead, the database has to be saved as a comma-separated file or text file. In Microsoft Excel, a database may be created and saved as a comma-separated file or text file (2010). The database has to be constructed similar to the example database of north central Texas wells. Unlike SAS, R does not allow empty cells within the database (Ripley et al. 2010). After working with R, I recommend that the first column be the dependent variable values and the following columns have the independent variable values. In R, having text within the first column and not the dependent variable caused problems.

Shown below are the steps to import a saved comma-separated file into R followed by an example using north central Texas database (Gardener) The following code will insert the saved .csv file as the work database in R. The following code has to be typed after the prompt " >". Labeled names of the columns are transferred from the .csv file. To clear the console screen in R at any time, hold down control L.

- > databasename= read.csv(file.choose(), header=T)
To see the inserted database and to know how to call upon individual variables, after the prompt, type the name of the dataset (Gardener).

- `>drilling`

To make the data accessible to R, insert the command after the prompt, `attach(name of the file)`. In the previous example the file name was “drilling”. To allow R to work with the variables within the example database, `>attach(drilling)`. 

Fig. B.1-Code to Open Window to Import Database

```
> drilling <- read.csv(file.choose(), header=T)
```

Fig. B.2-Code to See Database Values

```
> drilling
  ActualDays Estimated.Days  Depth.ft Temp.BHT.F     inc
  1    15        23     10260    249.420  24.600000
  2    32        26     14496    321.432  2.150000
  3    52        17     16025    347.425  2.000000
  4    54        23     19250    402.250  1.500000
  5    64        36     14620    323.540  33.850000
  6    22        21     16350    552.850  2.500000
  7    30        21     18700    392.900  2.990000
  8    67        42     13470    308.800  21.870000
```
B.2 Basic Code Necessary to Perform Regressions and Model Validation in R

The next step shows how to perform four types of regressions, least squares regression, stepwise regression, forward selection, and backward elimination. The example above is used to keep variable names constant; it is based on the example dataset for north central Texas. Least squares regression introduces a basic understanding of how to work with R before performing stepwise regressions, forward selection, and backward elimination. The code shown below represents the generic model fitting for assigned independent variables followed by an example (Venables et al. 2002).

- `LSR <- lm(y~x1+x2+x3,data)`
This code returned the same regressor coefficients as SAS for Model B.

Stepwise regression uses a similar form of the previous code to perform regression (Venables et al. 2002). Instead of indicating which independent variable to use, all independent variables are included. First a "full" model has to be created. >full <- lm(y~.,data)

The use of a period indicates to R that all independent variables are to be included. After creating the full model, type "step" to perform a stepwise regression. R will perform the regression without requiring an initial variable name. A variable name has been added to
the example to differentiate between stepwise regression, forward selection, and backward elimination (Venables et al. 2002).

- >step(full)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth.ft</td>
<td>1</td>
<td>208.1</td>
<td>22054</td>
</tr>
<tr>
<td>inc</td>
<td>1</td>
<td>311.5</td>
<td>22067</td>
</tr>
<tr>
<td>Temp.BHT.F</td>
<td>1</td>
<td>501.5</td>
<td>22257</td>
</tr>
<tr>
<td>&lt;none&gt;</td>
<td></td>
<td>21755</td>
<td>353.79</td>
</tr>
</tbody>
</table>

Fig. B.6-Code to Perform Stepwise Regression

After running the above code, I found the same regressor coefficients of Model A in SAS with R.

R has the basic code, >step(model, data, direction, scale, k, trace), to perform regressions (Phoa 2007). In, R forward selection comes from defining the step command, not to eliminate variables. Similar to stepwise regression’s full model, forward selection requires a "null" model with only one variable to start.

- >null<-lm(y~1,data)
To chose between stepwise regression, forward selection, or backward elimination, the "direction" part of the command may be written as, direction="both", direction="forward", and direction="back". As shown above, step has "both" as the default direction. Below, the necessary inputs for the step command are shown for forward selection (Phoa 2007).

- >step(null, scope=list(lower=null, upper=full, direction="forward")

The code for backward elimination matches closer to stepwise regression. Shown below, the necessary code to perform backward elimination in R (Phoa 2007).
The key parameters for validation were residual plots, Shapiro-Wilk p-value, Q-Q plot, chi-squared p-value, variance inflation factor, R-squared value, Mallow Cp, and F-test p-value. Stepwise regression, forward selection, and backward elimination within R use Akaike Information Criterion (AIC) to determine the optimal model. The step command may also be changed to perform regressions using Bayes Information Criterion (BIC) or Mallow Cp (Faraway 2002). To find the Mallow Cp of the regression as shown in SAS, the scale code needs to be adjusted. For Mallow Cp, \( \text{scale} = (\text{summary(full)}$sigma)^2 \) (Phoa 2007). Mallow Cp values in SAS equal the values found in R next to the variable named, none. Shown below, the code written using the examples are for Mallow Cp stepwise regression, forward selection, and backward elimination.

- \( > \text{Reg5}=\text{step(}\text{full, scale}=(\text{summary(}\text{full}$sigma)^2)\)
Fig. B.10-Code for Stepwise Regression with Mallow Cp

- > Reg6 = step(null, scope = list(lower = null, upper = full, direction = "forward"), scale = (summary(full)$sigma)^2)

Fig. B.11-Code for Forward Selection with Mallow Cp

- > Reg7 = step(full, direction = "backward", scale = (summary(full)$sigma)^2)
Residuals of the regression models are seen by using "$\text{res}$" after the model name. The code "plot.lm()" provides users with a method to see residual and QQ plots. After the prompt type, plot.lm(model name) to obtain Residual vs Fitted, Normal QQ, Scale Location, and Residual vs Leverage plots (Ricci 2005). Using `par(mfrow=c(2,2))` will allow the four plots to be seen within a single window.

- `par(mfrow=c(2,2))`
- `plot.lm(model)`
Fig. B.14-Results of plot.lm Code

Residual plot have the same attributes as seen with SAS’s residual plots. R’s QQ plot use standardized residuals instead of the raw data residuals as seen within SAS. Standardized residuals are residuals divided by their standard error (Dallal 2001). Scale location helps to assess nonconstant variance. Similar to a residual plot, a random scatter indicates a constant variance. The Loess line indicates the local trends of the residuals within the scale location plot. If the residuals have constant variance, then the Loess line will appear horizontal (Maathuis 2008). Within the residual vs leverage plot, the red lines correspond to Cook’s distance. Cook’s distance measures how much each point influences the estimated regression function and how much each point pull the function to itself (Maathuis 2008). Leverage indicates the difference between an observation from the given predictor value. It looks at the x’s and compares the difference between the individual groups of x’s with other groups of x’s. The majority of the points in the
residual vs. leverage plot should be bunched. Outliers indicate xs that are either higher or lower than the normal values of majority of the data (Maathuis 2008).

Residual plots in R can also be done using the plot command where the y-axis, x-axis, and title can be labeled within the plot code. The generic code for creating a residual plot and an example are shown below.

- `>plot(model$res, ylab="y-axis", xlab="x-axis", main="title")`

![Fig. B.15-Code for Residual Plot](image1)

![Fig. B.16-Residual Plot](image2)
Using the code shown below creates a QQ plot in R that resembles SAS’s QQ plot.

- `a <- modelname$res`
- `u <- (1:length(a) -.5)/length(a)`
- `Q <- quantile(a, u, type = 5)`
- `plot(qnorm(u), Q)`
- `abline(lm(Q ~ qnorm(u)))`

Fig. B.17-Code for QQ Plot in R
The Shapiro-Wilk residual normality p-value and Pearson chi-squared residual normality test require a single line of code. These tests are looking at the normality of the residuals and not the predicted values of the model. The code shown below tests the residuals normality using Shapiro-Wilk p-value and Pearson chi-squared test.

- `> shapiro.test(x)`
Fig. B.19-Code for Shapiro Wilk Normality Test

> shapiro.test(Reg$Sres)

Shapiro-Wilk normality test

data:  Reg$Sres
W = 0.6315, p-value = 8.31e-11

Fig. B.20-Code for Pearson Normality Test

> Pearson.test(x)

Pearson chi-square normality test

data:  Reg$Sres
\[ P = 34.1724, \ p-value = 3.78e^{-05} \]

R-squared and F-test p-values are found by looking at the summary of the model. After the prompt type “summary(model)” to see R-squared and F-test p-value.
B.3 Manipulation of Variables and Dataset

To look at the relationship between two variables, the plot command, shown above, may be used. When using the plot command, the first variable represents the x-axis and the second variable separated by a comma represents the y-axis. If manipulation of variables is needed to linearize the data, shown below are examples of basic variable manipulation in R.

- \( \text{lndepth} \leftarrow \log(\text{Depth.ft}) \)
- \( \text{logdepth} \leftarrow \log10(\text{Depth.ft}) \)
- \( \text{depthsqr} \leftarrow (\text{Depth.ft})^2 \)

![Fig. B.22-Example Code of Basic Variable Manipulation](image)
After creating a new variable, the “bind” command will add the new variable to the dataset being used. Every time a dataset changes, to minimize problems, a new dataset name should be made:

- \( \text{drilling2} \leftarrow \text{cbind(drilling,lndepth)} \)

If an outlier or data point needs to be removed, it may be done in R or by changing the dataset outside of R and importing it again. The code below removes any points within the variable ActualDays that has a value of 174. All other independent variables also are removed. When the new dataset gets called, \( \text{drilling3} \), the row with Actual days equal to 174 and all other independent variables associated with that row will be removed.

- \( \text{drilling3} \leftarrow \text{drilling2[which(ActualDays!=174),]} \)
Fig. B.24-Code to Remove Outliers

If any command shows the error message, “could not find function” the package for R with that command may not have been installed. To install them just click on the Packages tab and select Install Packages. After choosing a server a list of packages will appear. Select the package of interest and click on the OK button. After installation has been completed, upload the package by returning to the Packages tab and select Load Packages. After loading the package, the command should work. To avoid redundancy and the need to upload the same packages when R initiates, within R’s application file, make sure that wanted packages are found in the library folder. R creates a separate file within the user’s document file for new packages. Transfer the packages from the user R folder to the R’s library folder that may be found in R’s main application folder.
APPENDIX C

MODELS C, D, E, H, I, J, AND K’S MODEL EQUATION, RESIDUAL PLOT, QQ PLOT, AND THREE HISTORY MATCH CURVES

\[ y_C = -5.28257 + (0.87691 \times T_e) + (0.00104 \times D) \]  

(7.3)

Fig. C.1–Residual Plot of Model C
Fig. C.2—QQ Plot of Model C

Fig. C.3—History Match between Actual and Predicted Days for Model C
Fig. C.4–History Match, Predicted vs Actual Days for Model C

Fig. C.5–History Match, Predicted vs Actual Days for Model C Zoomed In
\[ y_D = -9.98231 + (0.86113 \times Te) + (0.06342 \times Tb) \] .................................(7.4)
Fig. D.8–History Match between Actual and Predicted Days for Model D

Fig. D.9–History Match, Predicted vs Actual Days for Model D
\[
\ln(E) = 0.44017 + (0.60152 \times \ln(Te)) + (0.00006721 \times D)
\] (7.5)
Fig. C.12–QQ Plot for Model E

Fig. C.13–History Match between Actual and Predicted Days for Model E
Fig. C.14–History Match, Predicted vs Actual Days for Model E

Fig C.15–History Match, Predicted vs Actual Days for Model E Zoomed In
\[ \ln(y_H) = 1.30916 + (0.02003 \times Te) + (0.00010005 \times D) \] (7.8)

**Fig. C.16**–Residual Plot for Model H

**Fig. C.17**–QQ Plot for Model H
Fig. C.18–History Match between Actual and Predicted Days for Model H

Fig. C.19–History Match, Predicted vs Actual Days for Model H
\[ \ln(y_I) = 0.88750 + (0.01893 \times Te) + (0.00598 \times Tbh) \] \hspace{1cm} (7.9)
Fig. C.22–QQ Plot for Model I

Fig. C.23–History Match between Actual and Predicted Days for Model I
Fig. C.24–History Mach, Predicted vs Actual Days for Model I

Fig. C.25–History Mach, Predicted vs Actual Days for Model I Zoomed In
\[ \ln(y_j) = -4.1396 + (0.02007 \times T_e) + (0.72552 \times \ln(D)) \] \hspace{1cm} \text{(7.10)}

Fig. C.26–Residual Plot for Model J

Fig. C.27–QQ Plot for Model J
Fig. C.28–History Match between Actual and Predicted Days for Model J

Fig. C29–History Match, Predicted vs Actual Days for Model J
\[ \ln(y_K) = -4.98991 + (0.01828 \times T_e) + (1.36169 \times \ln(T_{bh})) \] ..........................(7.11)
Fig C.32–QQ Plot for Model K

Fig C.33–History Match between Actual and Predicted Days for Model K
Fig C.34–History Match, Predicted Days vs Actual Days for Model K

Fig. C.35–History Match, Predicted Days vs Actual Days for Model K
APPENDIX D

GRAPHICAL COMPARISON OF DEPENDENT VERSE INDEPENDENT VARIABLES

Fig. D.1—Actual Days vs Estimated Days

Fig. D.2—Actual Days vs Depth
Fig. D.3–Actual Days vs. BHT

Fig. D.4–ln(Actual Days) vs. ln(Estimated Days)
Fig. D.5–ln(Actual Days) vs. ln(Depth)

Fig. D.6–ln(Actual Days) vs. Depth
Fig. D.7 – ln(Actual Days) vs. BHT

Fig. D.8 – ln(Actual Days) vs. inc
Fig. D.9–ln(Actual Days) vs. (Estimated Days)^2
APPENDIX E

SAS OUTPUT FOR ALL REGRESSIONS AND UNIVARIATE CALCULATIONS

The SAS System 10:08 Thursday, August 12, 2010

The CONTENTS Procedure

Data Set Name          WORK.DRILLING            Observations
                         58                          
Member Type            DATA                          
Engine                 V9                           
Created                Thursday, August 12, 2010 10:13:10 AM  
Last Modified          Thursday, August 12, 2010 10:13:10 AM  
Protection             NO                           
Data Set Type          NO                           
Label                  Data Representation  WINDOWS_64  
Encoding               wlatin1 Western (Windows)  

Engine/Host Dependent Information

Data Set Page Size          8192
Number of Data Set Pages    1
First Data Page             1
Max Obs per Page            145
Obs in First Data Page      58
Number of Data Set Repairs  0
Filename                    C:\Users\JDEALM-1\AppData\Local\Temp\SAS Temporary Files\TD3960\drilling.SAS7bdat
Release Created             9.0202M3
Host Created                X64_VSPRO

Alphabetic List of Variables and Attributes

<table>
<thead>
<tr>
<th>#</th>
<th>Variable</th>
<th>Type</th>
<th>Len</th>
<th>Format</th>
<th>Informat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Well</td>
<td>Char</td>
<td>9</td>
<td>$9.</td>
<td>$9.</td>
</tr>
<tr>
<td>2</td>
<td>ActualDays</td>
<td>Num</td>
<td>8</td>
<td>BEST12.</td>
<td>BEST32.</td>
</tr>
<tr>
<td>3</td>
<td>Depth_ft</td>
<td>Num</td>
<td>8</td>
<td>BEST12.</td>
<td>BEST32.</td>
</tr>
<tr>
<td>4</td>
<td>Estimated_Days</td>
<td>Num</td>
<td>8</td>
<td>BEST12.</td>
<td>BEST32.</td>
</tr>
<tr>
<td>5</td>
<td>Temp_BHT_F</td>
<td>Num</td>
<td>8</td>
<td>BEST12.</td>
<td>BEST32.</td>
</tr>
</tbody>
</table>
**Model A**

The SAS System 10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read 58
Number of Observations Used 58

Stepwise Selection: Step 1

Variable Estimated_Days Entered: R-Square = 0.4218 and C(p) = 6.1922

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>1</td>
<td>18023</td>
<td>18023</td>
<td>40.85</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>56</td>
<td>24708</td>
<td>441.20679</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Standard Error</th>
<th>Type II SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.17256</td>
<td>265.78388</td>
<td>0.60</td>
<td>0.4409</td>
</tr>
<tr>
<td>Estimated_Days</td>
<td>0.18129</td>
<td>18023</td>
<td>40.85</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable Temp_BHT_F Entered: R-Square = 0.4829 and C(p) = 1.8317

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>2</td>
<td>20634</td>
<td>10317</td>
<td>25.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>22097</td>
<td>401.75946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Standard Error</th>
<th>Type II SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.55474</td>
<td>1248.07502</td>
<td>3.11</td>
<td>0.0835</td>
</tr>
<tr>
<td>Estimated_Days</td>
<td>0.20547</td>
<td>7303.17586</td>
<td>18.18</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Temp_BHT_F</td>
<td>0.03630</td>
<td>2610.80963</td>
<td>6.50</td>
<td>0.0136</td>
</tr>
</tbody>
</table>

The SAS System 10:08 Thursday, August 12, 2010
All variables left in the model are significant at the 0.1500 level.
No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Selection

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable Entered</th>
<th>Variable Removed</th>
<th>Number Vars In</th>
<th>Partial R-Square</th>
<th>Model R-Square</th>
<th>C(p)</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Estimated_Days</td>
<td></td>
<td>1</td>
<td>0.4218</td>
<td>0.4218</td>
<td>6.1922</td>
<td>40.85</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Temp_BHT_F</td>
<td></td>
<td>2</td>
<td>0.0611</td>
<td>0.4829</td>
<td>1.8317</td>
<td>6.50</td>
<td>0.0136</td>
</tr>
<tr>
<td>0.0136</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The SAS System 10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read 58
Number of Observations Used 58

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>2</td>
<td>20634</td>
<td>10317</td>
<td>25.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>22097</td>
<td>401.75946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 20.04394
R-Square 0.4829
Dependent Mean 31.98276
Adj R-Sq 0.4641
Coeff Var 62.67107

Parameter Estimates

| Parameter   | DF | Estimate | Error  | t Value | Pr > |t| |
|-------------|----|----------|--------|---------|-------|---|
| Inflation   |    |          |        |         |       |   |
| Intercept   | 1  | -16.84055| 9.55474| -1.76   | 0.0835|   |
| Estimated_Days | 1 | 0.87604  | 0.20547| 4.26    | <.0001|   |
| Temp_BHT_F  | 1  | 0.09255  | 0.03630| 2.55    | 0.0136|   |

The SAS System 10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.48</td>
<td>0.6266</td>
</tr>
</tbody>
</table>

The SAS System  10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read          58
Number of Observations Used          58

Forward Selection: Step 1

Variable Estimated_Days Entered: R-Square = 0.4218 and C(p) = 6.1922

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>1</td>
<td>18023</td>
<td>18023</td>
<td>40.85</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>56</td>
<td>24708</td>
<td>441.20679</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variable Temp_BHT_F Entered: R-Square = 0.4829 and C(p) = 1.8317

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>2</td>
<td>20634</td>
<td>10317</td>
<td>25.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>22097</td>
<td>401.75946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Forward Selection: Step 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Standard Error</th>
<th>Type II SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-16.84055</td>
<td>9.55474</td>
<td>1248.07502</td>
<td>3.11</td>
<td>0.0835</td>
</tr>
<tr>
<td>Estimated_Days</td>
<td>0.87604</td>
<td>0.20547</td>
<td>7303.17586</td>
<td>18.18</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Temp_BHT_F</td>
<td>0.09255</td>
<td>0.03630</td>
<td>2610.80963</td>
<td>6.50</td>
<td>0.0136</td>
</tr>
</tbody>
</table>

Bounds on condition number: 1.4108, 5.643

No other variable met the 0.5000 significance level for entry into the model.

Summary of Forward Selection

<table>
<thead>
<tr>
<th>Step Entered</th>
<th>Variable Entered</th>
<th>Number Vars In</th>
<th>Partial R-Square</th>
<th>Model R-Square</th>
<th>C(p)</th>
<th>F Value</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &lt; .0001</td>
<td>Estimated_Days</td>
<td>1</td>
<td>0.4218</td>
<td>0.4218</td>
<td>6.1922</td>
<td>40.85</td>
<td></td>
</tr>
<tr>
<td>2 0.0136</td>
<td>Temp_BHT_F</td>
<td>2</td>
<td>0.0611</td>
<td>0.4829</td>
<td>1.8317</td>
<td>6.50</td>
<td></td>
</tr>
</tbody>
</table>
The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read 58
Number of Observations Used 58

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>2</td>
<td>20634</td>
<td>10317</td>
<td>25.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>22097</td>
<td>401.75946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 20.04394  R-Square 0.4829
Dependent Mean 31.98276  Adj R-Sq 0.4641
Coeff Var 62.67107

Parameter Estimates

| Parameter            | DF | Estimate  | Standard Error | t Value | Pr > |t| |
|----------------------|----|-----------|----------------|---------|------|
| Intercept            | 1  | -16.84055 | 9.55474        | -1.76   | 0.0835|
| Estimated_Days       | 1  | 0.87604   | 0.20547        | 4.26    | <.0001|
| Temp_BHT_F           | 1  | 0.09255   | 0.03630        | 2.55    | 0.0136|

The SAS System 10:08 Thursday, August 12, 2010 9

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.48</td>
<td>0.6266</td>
</tr>
</tbody>
</table>

The SAS System 10:08 Thursday, August 12, 2010 10

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read 58
Number of Observations Used 58

Backward Elimination: Step 0
All Variables Entered: R-Square = 0.4909 and C(p) = 5.0000

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>4</td>
<td>20976</td>
<td>5243.90663</td>
<td>12.78</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>53</td>
<td>21755</td>
<td>410.47842</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Standard

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Type II SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-40.20190</td>
<td>28.46162</td>
<td>818.96294</td>
<td>2.00</td>
<td>0.1636</td>
</tr>
<tr>
<td>Depth_ft</td>
<td>-0.00517</td>
<td>0.00607</td>
<td>298.12846</td>
<td>0.73</td>
<td>0.3979</td>
</tr>
<tr>
<td>Estimated_Days</td>
<td>0.87268</td>
<td>0.21013</td>
<td>7079.57568</td>
<td>17.25</td>
<td>0.0001</td>
</tr>
<tr>
<td>Inc</td>
<td>0.13626</td>
<td>0.15641</td>
<td>311.54542</td>
<td>0.76</td>
<td>0.3876</td>
</tr>
<tr>
<td>Temp_BHT_F</td>
<td>0.39575</td>
<td>0.35804</td>
<td>501.49284</td>
<td>1.22</td>
<td>0.2740</td>
</tr>
</tbody>
</table>

Bounds on condition number: 134.3, 1085.4

Backward Elimination: Step 1

Variable Depth_ft Removed: R-Square = 0.4839 and C(p) = 3.7263

The REG Procedure

Model: MODEL1
Dependent Variable: ActualDays

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>3</td>
<td>20677</td>
<td>6892.49935</td>
<td>16.88</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>22053</td>
<td>408.39786</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Standard

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Type II SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-17.43942</td>
<td>9.80741</td>
<td>1291.33397</td>
<td>3.16</td>
<td>0.0810</td>
</tr>
<tr>
<td>Estimated_Days</td>
<td>0.88525</td>
<td>0.20908</td>
<td>7321.09539</td>
<td>17.93</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Inc</td>
<td>0.03131</td>
<td>0.09619</td>
<td>43.28578</td>
<td>0.11</td>
<td>0.7460</td>
</tr>
<tr>
<td>Temp_BHT_F</td>
<td>0.09222</td>
<td>0.03662</td>
<td>2590.80623</td>
<td>6.34</td>
<td>0.0148</td>
</tr>
</tbody>
</table>

Bounds on condition number: 1.437, 11.612

Backward Elimination: Step 2
Variable inc Removed: R-Square = 0.4829 and C(p) = 1.8317

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>2</td>
<td>20634</td>
<td>10317</td>
<td>25.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>22097</td>
<td>401.75946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The SAS System 10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Backward Elimination: Step 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Type II SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-16.84055</td>
<td>9.55474</td>
<td>1248.07502</td>
<td>3.11</td>
<td>0.0835</td>
</tr>
<tr>
<td>Estimated_Days</td>
<td>0.87604</td>
<td>0.20547</td>
<td>7303.17586</td>
<td>18.18</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Temp_BHT_F</td>
<td>0.09255</td>
<td>0.03630</td>
<td>2610.80963</td>
<td>6.50</td>
<td>0.0136</td>
</tr>
</tbody>
</table>

Bounds on condition number: 1.4108, 5.643

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination

<table>
<thead>
<tr>
<th>Step &gt; F</th>
<th>Variable Removed</th>
<th>Number Vars In</th>
<th>Partial R-Square</th>
<th>Model R-Square</th>
<th>C(p)</th>
<th>F Value</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Depth_ft</td>
<td>3</td>
<td>0.0070</td>
<td>0.4839</td>
<td>3.7263</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>0.3979</td>
<td>inc</td>
<td>2</td>
<td>0.0010</td>
<td>0.4829</td>
<td>1.8317</td>
<td>0.11</td>
<td></td>
</tr>
</tbody>
</table>

The SAS System 10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read  58
Number of Observations Used   58

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>2</td>
<td>20634</td>
<td>10317</td>
<td>25.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>22097</td>
<td>401.75946</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Corrected Total          57          42731

Root MSE                     20.04394   R-Square      0.4829
Dependent Mean                 31.98276   Adj R-Sq       0.4641
Coef Var                       62.67107

Parameter Estimates

| Parameter       | DF  | Estimate | Error  | t Value | Pr > |t| |
|-----------------|-----|----------|--------|---------|-------|
| Inflation       |     |          |        |         |       |
| Intercept       | 1   | -16.84055| 9.55474| -1.76   | 0.0835|
| Estimated_Days  | 1   | 0.87604  | 0.20547| 4.26    | <.0001|
| Temp_BHT_F      | 1   | 0.09255  | 0.03630| 2.55    | 0.0136|
The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Test of First and Second
Moment Specification

DF    Chi-Square    Pr > ChiSq
      5          3.48        0.6266

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read          58
Number of Observations Used          58

Analysis of Variance

Source       DF    Sum of Squares    Mean Square    F Value    Pr > F
Model         2    20634          10317          25.68    <.0001
Error        55    22097        401.75946
Corrected Total        57         42731

Root MSE             20.04394    R-Square     0.4829
Dependent Mean       31.98276    Adj R-Sq     0.4641
Coef Var            62.67107

Parameter Estimates

Variance Variable       DF    Parameter  Standard
           Inflation
Intercept          1      -16.84055        9.55474      -1.76      0.0835
0
Temp_BHT_F         1        0.09255        0.03630       2.55      0.0136
1.41076
Estimated_Days     1        0.87604        0.20547       4.26      <.0001
1.41076

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Test of First and Second
Moment Specification

DF    Chi-Square    Pr > ChiSq
      5          3.48        0.6266
Q-Q Plot: Actualdays = TEMP_BHT_F  Estimated_Days

10:08 Thursday, August 12, 2010

The UNIVARIATE Procedure
Variable: resid (Residual)

Moments

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>58</td>
<td>Sum Weights</td>
</tr>
<tr>
<td>Mean</td>
<td>0</td>
<td>Sum Observations</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>19.6891503</td>
<td>Variance</td>
</tr>
<tr>
<td>Skewness</td>
<td>4.27757171</td>
<td>Kurtosis</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>22096.7705</td>
<td>Corrected SS</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>.</td>
<td>Std Error Mean</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.00000</td>
</tr>
<tr>
<td>Median</td>
<td>-2.72978</td>
</tr>
<tr>
<td>Mode</td>
<td>.</td>
</tr>
<tr>
<td>Range</td>
<td>155.99533</td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>14.01188</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</td>
<td>0</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>-6</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>-133.5</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</td>
<td>0.631467</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.174238</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.619197</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>3.965366</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>123.03080</td>
</tr>
<tr>
<td>99%</td>
<td>123.03080</td>
</tr>
<tr>
<td>95%</td>
<td>18.32326</td>
</tr>
<tr>
<td>90%</td>
<td>13.46468</td>
</tr>
</tbody>
</table>
Q-Q Plot: Actualdays = TEMP_BHT_F  Estimated_Days

10:08 Thursday, August 12, 2010

The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

Quantile     Estimate
75% Q3        6.11412
50% Median    -2.72978
25% Q1        -7.89776
10%           -14.16112
5%            -24.61165
1%            -32.96453
0% Min        -32.96453

Extreme Observations

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-32.9645</td>
<td>29</td>
<td>13.9134</td>
<td>8</td>
</tr>
<tr>
<td>-26.5824</td>
<td>26</td>
<td>15.4590</td>
<td>56</td>
</tr>
<tr>
<td>-24.6116</td>
<td>57</td>
<td>18.3233</td>
<td>35</td>
</tr>
<tr>
<td>-14.9525</td>
<td>46</td>
<td>19.3604</td>
<td>5</td>
</tr>
<tr>
<td>-14.9371</td>
<td>52</td>
<td>123.0308</td>
<td>39</td>
</tr>
</tbody>
</table>
Model B

Residual Plot: Actualdays = TEMP_BHT_F  Estimated_Days

19

12, 2010

10:08 Thursday, August

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read          58
Number of Observations Used          58

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>2</td>
<td>20474</td>
<td>10237</td>
<td>25.30</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>22257</td>
<td>404.67104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>57</td>
<td>42731</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE             20.11644  R-Square     0.4791
Dependent Mean       31.98276  Adj R-Sq     0.4602
Coeff Var            62.89775

Parameter Estimates

| Parameter          | DF | Estimate | Standard Error | t Value | Pr > |t| |
|--------------------|----|----------|----------------|---------|------|---|
| Intercept          | 1  | -9.88984 | 7.51429        | -1.32   | 0.1936|
| Depth_ft           | 1  | 0.00150  | 0.00061002     | 2.46    | 0.0170|
| Estimated_Days     | 1  | 0.90118  | 0.20271        | 4.45    | <.0001|

Residual Plot: Actualdays = TEMP_BHT_F  Estimated_Days

20

12, 2010

10:08 Thursday, August

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Test of First and Second
Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.45</td>
<td>0.6307</td>
</tr>
</tbody>
</table>

Q-Q Plot: Actualdays = Depth_ft Estimated_Days

21

12, 2010

10:08 Thursday, August

The UNIVARIATE Procedure
Variable: resid (Residual)
Moments

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>58</td>
</tr>
<tr>
<td>Sum Weights</td>
<td>58</td>
</tr>
<tr>
<td>Mean</td>
<td>0</td>
</tr>
<tr>
<td>Sum Observations</td>
<td>0</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>19.7603657</td>
</tr>
<tr>
<td>Variance</td>
<td>390.472053</td>
</tr>
<tr>
<td>Skewness</td>
<td>4.29124048</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>26.9643936</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>22256.907</td>
</tr>
<tr>
<td>Corrected SS</td>
<td>22256.907</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>.</td>
</tr>
<tr>
<td>Std Error Mean</td>
<td>2.59466313</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.00000</td>
</tr>
<tr>
<td>Median</td>
<td>-2.85754</td>
</tr>
<tr>
<td>Mode</td>
<td>.</td>
</tr>
<tr>
<td>Range</td>
<td>156.77331</td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>14.34040</td>
</tr>
</tbody>
</table>

Tests for Location: \( \mu_0 = 0 \)

<table>
<thead>
<tr>
<th>Test</th>
<th>-Statistic-</th>
<th>-----p Value------</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's t</td>
<td>( t )</td>
<td>0 Pr &gt;</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>-6 Pr &gt;=</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>-136.5 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 0.630373</td>
<td>Pr &lt; W &lt;0.0001</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D 0.178141</td>
<td>Pr &gt; D &lt;0.0100</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq 0.617013</td>
<td>Pr &gt; W-Sq &lt;0.0050</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq 3.974491</td>
<td>Pr &gt; A-Sq &lt;0.0050</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>123.52585</td>
</tr>
<tr>
<td>99%</td>
<td>123.52585</td>
</tr>
<tr>
<td>95%</td>
<td>18.96744</td>
</tr>
<tr>
<td>90%</td>
<td>13.81943</td>
</tr>
</tbody>
</table>

Q-Q Plot: Actualdays = Depth_ft Estimated_Days

22

10:08 Thursday, August 12, 2010

The UNIVARIATE Procedure

Variable: resid (Residual)

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% Q3</td>
<td>6.33997</td>
</tr>
<tr>
<td>50% Median</td>
<td>-2.85754</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-8.00043</td>
</tr>
<tr>
<td>10%</td>
<td>-13.78917</td>
</tr>
<tr>
<td>5%</td>
<td>-24.16070</td>
</tr>
<tr>
<td>1%</td>
<td>-33.24746</td>
</tr>
<tr>
<td>Value</td>
<td>Obs</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
</tr>
<tr>
<td>-33.2475</td>
<td>29</td>
</tr>
<tr>
<td>-26.1717</td>
<td>26</td>
</tr>
<tr>
<td>-24.1607</td>
<td>57</td>
</tr>
<tr>
<td>-14.6087</td>
<td>52</td>
</tr>
<tr>
<td>-14.1998</td>
<td>46</td>
</tr>
</tbody>
</table>
### Model C

Residual Plot: ActualDays = Depth_ft Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

| Number of Observations Read | 57 |
| Number of Observations Used  | 57 |

#### 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>2</td>
<td>15850</td>
<td>7925.04348</td>
<td>67.31</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>6358.15865</td>
<td>117.74368</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>22208</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE: 10.85098
R-Square: 0.7137
Dependent Mean: 29.49123
Adj R-Sq: 0.7031
Coeff Var: 36.79391

#### Parameter Estimates

| Parameter       | DF | Estimate | Standard Error | t Value | Pr > |t|   |
|-----------------|----|----------|----------------|---------|------|----|
| Intercept       | 1  | -5.28257 | 4.07262        | -1.30   | 0.2001|
| Depth_ft        | 1  | 0.00104  | 0.00033148     | 3.12    | 0.0029|
| 1.35154         |    |          |                |         |      |
| Estimated_Days  | 1  | 0.87691  | 0.10936        | 8.02    | <.0001|
| 1.35154         |    |          |                |         |      |

Residual Plot: ActualDays = Depth_ft Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10.41</td>
<td>0.0643</td>
</tr>
</tbody>
</table>

Q-Q Plot: ActualDays = Depth_ft Estimated_Days

10:08 Thursday, August 12, 2010

The UNIVARIATE Procedure
Variable: resid (Residual)
Moments

<table>
<thead>
<tr>
<th>Moment</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>57</td>
</tr>
<tr>
<td>Sum Weights</td>
<td>57</td>
</tr>
<tr>
<td>Mean</td>
<td>0</td>
</tr>
<tr>
<td>Sum Observations</td>
<td>0</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>10.655458</td>
</tr>
<tr>
<td>Variance</td>
<td>113.53855</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.15585</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.943013</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>6358.15865</td>
</tr>
<tr>
<td>Corrected SS</td>
<td>6358.15865</td>
</tr>
<tr>
<td>Coef Variation</td>
<td></td>
</tr>
<tr>
<td>Std Error Mean</td>
<td>1.411348</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Variability</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0000000</td>
</tr>
<tr>
<td>Median</td>
<td>0.529895</td>
</tr>
<tr>
<td>Mode</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>55.77660</td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>9.78507</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</td>
<td>0</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>2.5</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>10.5</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</td>
<td>0.972431</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.093778</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.099368</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.594442</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>23.993124</td>
</tr>
<tr>
<td>99%</td>
<td>23.993124</td>
</tr>
<tr>
<td>95%</td>
<td>19.178971</td>
</tr>
<tr>
<td>90%</td>
<td>15.385281</td>
</tr>
</tbody>
</table>

Q-Q Plot: Actualdays - Depth_ft Estimated_Days

The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% Q3</td>
<td>4.412667</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.529895</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-5.372400</td>
</tr>
<tr>
<td>10%</td>
<td>-10.711766</td>
</tr>
<tr>
<td>5%</td>
<td>-20.759939</td>
</tr>
<tr>
<td>1%</td>
<td>-31.783474</td>
</tr>
</tbody>
</table>

10:08 Thursday, August 12, 2010
0% Min \(-31.783474\)

**Extreme Observations**

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-31.7835</td>
<td>29</td>
<td>16.5033</td>
<td>8</td>
</tr>
<tr>
<td>-22.6537</td>
<td>26</td>
<td>18.5142</td>
<td>55</td>
</tr>
<tr>
<td>-20.7559</td>
<td>56</td>
<td>19.1790</td>
<td>4</td>
</tr>
<tr>
<td>-12.5988</td>
<td>11</td>
<td>22.5739</td>
<td>5</td>
</tr>
<tr>
<td>-11.9955</td>
<td>51</td>
<td>23.9931</td>
<td>35</td>
</tr>
</tbody>
</table>
Model D

Residual Plot: Actualdays = Depth_ft Estimated_Days

27
12, 2010

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read 57
Number of Observations Used 57

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>2</td>
<td>15907</td>
<td>7953.54216</td>
<td>68.16</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>6301.16130</td>
<td>116.68817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>22208</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 10.80223 R-Square 0.7163
Dependent Mean 29.49123 Adj R-Sq 0.7058
Coeff Var 36.62862

Parameter Estimates

| Variance          | DF | Parameter | Standard Error | t Value | Pr > |t| |
|-------------------|----|-----------|----------------|---------|------|---|
| Variable          |    | Inflation |                |         |      |   |
| Interception      | 1  | -9.98231  | 5.18294        | -1.93   | 0.0594 |
| Temp_BHT_F        | 1  | 0.06342   | 0.01972        | 3.22    | 0.0022 |
| Estimated_Days    | 1  | 0.86113   | 0.11074        | 7.78    | <.0001 |

Residual Plot: Actualdays = Depth_ft Estimated_Days

28
12, 2010

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10.60</td>
<td>0.0600</td>
</tr>
</tbody>
</table>
The UNIVARIATE Procedure
Variable:  resid (Residual)

Moments

N                           57    Sum Weights                 57
Mean                        0    Sum Observations             0
Std Deviation      10.6075792    Variance            112.520738
Skewness            -0.190869    Kurtosis            0.97192839
Uncorrected SS   6301.1613    Corrected SS         6301.1613
Coeff Variation             .    Std Error Mean      1.40500819

Basic Statistical Measures

Location                    Variability
Mean     0.000000     Std Deviation           10.60758
Median   0.693429     Variance               112.52074
Mode      .           Range                   55.21750
Interquartile Range      9.66674

Tests for Location: Mu0=0

Test           -Statistic-    -----p Value------
Student's t    t         0    Pr > |t|    1.0000
Sign           M       2.5    Pr >= |M|   0.5966
Signed Rank    S      11.5    Pr >= |S|   0.9282

Tests for Normality

Test          --Statistic---    -----p Value------
Shapiro-Wilk          W     0.969891    Pr < W      0.1659
Kolmogorov-Smirnov    D     0.095845    Pr > D     >0.1500
Cramer-von Mises      W-Sq  0.113858    Pr > W-Sq  0.0753
Anderson-Darling      A-Sq  0.662165    Pr > A-Sq   0.0833

Quantiles (Definition 5)

Quantile        Estimate
100% Max       23.579001
99%            23.579001
95%            18.667073
90%            15.385396
Q-Q Plot: Actualdays = TEMP_BHT_F  Estimated_Days

The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% Q3</td>
<td>4.333042</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.693429</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-5.333699</td>
</tr>
<tr>
<td>10%</td>
<td>-11.201434</td>
</tr>
<tr>
<td>5%</td>
<td>-21.058065</td>
</tr>
<tr>
<td>1%</td>
<td>-31.638498</td>
</tr>
<tr>
<td>0% Min</td>
<td>-31.638498</td>
</tr>
</tbody>
</table>

Extreme Observations

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-31.6385</td>
<td>29</td>
<td>16.5369</td>
<td>8</td>
</tr>
<tr>
<td>-22.9302</td>
<td>26</td>
<td>18.3696</td>
<td>55</td>
</tr>
<tr>
<td>-21.0581</td>
<td>56</td>
<td>18.6671</td>
<td>4</td>
</tr>
<tr>
<td>-12.6892</td>
<td>11</td>
<td>22.4639</td>
<td>5</td>
</tr>
<tr>
<td>-12.2253</td>
<td>51</td>
<td>23.5790</td>
<td>35</td>
</tr>
</tbody>
</table>

10:08 Thursday, August 12, 2010
**Model E**

```
logactualdays*logDepth_ft

10:08 Thursday, August 12, 2010
```

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read          57
Number of Observations Used          57

Stepwise Selection: Step 1

Variable log_estimated_days Entered: R-Square = 0.7232 and C(p) = 29.5573

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>1</td>
<td>26.45166</td>
<td>26.45166</td>
<td>143.70</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>10.12436</td>
<td>0.18408</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Type II SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.25572</td>
<td>0.24557</td>
<td>0.19961</td>
<td>1.08</td>
<td>0.3023</td>
</tr>
<tr>
<td>log_estimated_days</td>
<td>0.96484</td>
<td>0.08049</td>
<td>26.45166</td>
<td>143.70</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable Depth_ft Entered: R-Square = 0.7995 and C(p) = 8.7999

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>2</td>
<td>29.24250</td>
<td>14.62125</td>
<td>107.66</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>7.33352</td>
<td>0.13581</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

logactualdays*logDepth_ft

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Stepwise Selection: Step 2

Parameter Estimate Standard Error
### Parameter Estimates

<p>| Parameter               | DF | Standard Error | t Value | Pr &gt; |t| |
|-------------------------|----|----------------|---------|------|---|
| Intercept               | 1  | 0.21481        | 2.05    | 0.0453 |
| Depth_ft                | 1  | 0.00001483     | 4.53    | &lt;.0001|</p>
<table>
<thead>
<tr>
<th>log_estimated_days</th>
<th>1</th>
<th>0.60152</th>
<th>0.10584</th>
<th>5.68</th>
<th>&lt;.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>logactualdays*logDepth_ft</td>
<td>2.34394</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>logactualdays</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:08 Thursday, August 12, 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>8.26</td>
<td>0.1423</td>
</tr>
</tbody>
</table>
10:08 Thursday, August 12, 2010

The UNIVARIATE Procedure
Variable: resid (Residual)

Moments

<table>
<thead>
<tr>
<th>Moment</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>57</td>
</tr>
<tr>
<td>Sum Weights</td>
<td>57</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0000000</td>
</tr>
<tr>
<td>Sum Observations</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>10.60758</td>
</tr>
<tr>
<td>Variance</td>
<td>112.52074</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.190870</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.9719283</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>6301.1613</td>
</tr>
<tr>
<td>Corrected SS</td>
<td>6301.1613</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>.000000</td>
</tr>
<tr>
<td>Std Error Mean</td>
<td>1.405008</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.000000</td>
</tr>
<tr>
<td>Median</td>
<td>0.693429</td>
</tr>
<tr>
<td>Mode</td>
<td>.</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>10.60758</td>
</tr>
<tr>
<td>Variance</td>
<td>112.52074</td>
</tr>
<tr>
<td>Range</td>
<td>55.21750</td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>9.66674</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>0</td>
<td>1.0000</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>2.5</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>11.5</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</td>
<td>0.969891</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.095845</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.113858</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.662165</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>23.579001</td>
</tr>
<tr>
<td>99%</td>
<td>23.579001</td>
</tr>
<tr>
<td>95%</td>
<td>18.667073</td>
</tr>
<tr>
<td>90%</td>
<td>15.385396</td>
</tr>
</tbody>
</table>
Q-Q Plot: log Actual Days = Depth_ft log_Estimated_Days

10:08 Thursday, August 12, 2010

The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% Q3</td>
<td>4.333042</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.693429</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-5.333699</td>
</tr>
<tr>
<td>10%</td>
<td>-11.201434</td>
</tr>
<tr>
<td>5%</td>
<td>-21.058065</td>
</tr>
<tr>
<td>1%</td>
<td>-31.638498</td>
</tr>
<tr>
<td>0% Min</td>
<td>-31.638498</td>
</tr>
</tbody>
</table>

Extreme Observations

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-31.6385</td>
<td>29</td>
<td>16.5369</td>
<td>8</td>
</tr>
<tr>
<td>-22.9302</td>
<td>26</td>
<td>18.3696</td>
<td>55</td>
</tr>
<tr>
<td>-21.0581</td>
<td>56</td>
<td>18.6671</td>
<td>4</td>
</tr>
<tr>
<td>-12.6892</td>
<td>11</td>
<td>22.4639</td>
<td>5</td>
</tr>
<tr>
<td>-12.2253</td>
<td>51</td>
<td>23.5790</td>
<td>35</td>
</tr>
</tbody>
</table>
Model F

Residual Plot: log Actual Days = Depth_ft log_Estimated_Days

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

Forward Selection: Step 1

Variable log_estimated_days Entered: R-Square = 0.7232 and C(p) = 29.5573

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>1</td>
<td>26.45166</td>
<td>26.45166</td>
<td>143.70</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>10.12436</td>
<td>0.18408</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter     Standard

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Type II SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.25572</td>
<td>0.24557</td>
<td>0.19961</td>
<td>1.08</td>
<td>0.3023</td>
</tr>
<tr>
<td>log_estimated_days</td>
<td>0.96484</td>
<td>0.08049</td>
<td>26.45166</td>
<td>143.70</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Bounds on condition number: 1, 1

------

Forward Selection: Step 2

Variable Depth_ft Entered: R-Square = 0.7995 and C(p) = 8.7999

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>2</td>
<td>29.24250</td>
<td>14.62125</td>
<td>107.66</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>7.33352</td>
<td>0.13581</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Residual Plot: log Actual Days = Depth_ft log_Estimated_Days

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Forward Selection: Step 2

Parameter     Standard
Variable               Estimate        Error   Type II SS  F Value  Pr > F
Intercept               0.44017      0.21481      0.57021     4.20  0.0453
Depth_ft             0.00006721   0.00001483      2.79084    20.55  <.0001
log_estimated_days      0.60152      0.10584      4.38632    32.30  <.0001

Bounds on condition number: 2.3439, 9.3758

Forward Selection: Step 3

Variable inc Entered: R-Square = 0.8052 and C(p) = 9.1051

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>3</td>
<td>29.45033</td>
<td>9.81678</td>
<td>73.02</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>53</td>
<td>7.12568</td>
<td>0.13445</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter     Standard
Variable               Estimate        Error   Type II SS  F Value  Pr > F
Intercept               0.37856      0.21940      0.40026     2.98  0.0903
Depth_ft             0.00006371   0.00001502      2.42031    18.00  <.0001
log_estimated_days      0.62660      0.10723      4.59127    34.15  <.0001
inc                     0.00219      0.00177      0.20784     1.55  0.2192

Bounds on condition number: 2.43, 17.702

Forward Selection: Step 4

Residual Plot: log Actual Days = Depth_ft log_Estimated_Days

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Forward Selection: Step 4

Variable Temp_BHT_F Entered: R-Square = 0.8257 and C(p) = 5.0000

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>4</td>
<td>30.19903</td>
<td>7.54976</td>
<td>61.56</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>52</td>
<td>6.37699</td>
<td>0.12263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter     Standard
Variable               Estimate        Error   Type II SS  F Value  Pr > F
All variables have been entered into the model.

Summary of Forward Selection

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>Number Entered</th>
<th>Partial Vars In</th>
<th>Partial R-Square</th>
<th>Model R-Square</th>
<th>C(p)</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>log_estimated_days</td>
<td>1</td>
<td>0.7232</td>
<td>0.7232</td>
<td>29.5573</td>
<td>143.70</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Depth_ft</td>
<td>2</td>
<td>0.0763</td>
<td>0.7995</td>
<td>8.7999</td>
<td>20.55</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>inc</td>
<td>3</td>
<td>0.0057</td>
<td>0.8052</td>
<td>9.1051</td>
<td>1.55</td>
<td>0.2192</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Temp_BHT_F</td>
<td>4</td>
<td>0.0205</td>
<td>0.8257</td>
<td>5.0000</td>
<td>6.11</td>
<td>0.0168</td>
<td></td>
</tr>
</tbody>
</table>

Residual Plot: log Actual Days = Depth_ft log_Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

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>4</td>
<td>30.19903</td>
<td>7.54976</td>
<td>61.56</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>52</td>
<td>6.37699</td>
<td>0.12263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.35019
R-Square 0.8257
Dependent Mean 3.11950
Adj R-Sq 0.8122
Coef Var 11.22590

Parameter Estimates

| Parameter Type | DF | Estimate | Standard Error | t Value | Pr > |t| |
|----------------|----|----------|----------------|---------|------|---|
| Intercept      | 1  | -0.72931 | 0.49492        | -1.47   | 0.1466|   |

Bounds on condition number: 134.39, 1080.4

----------------------------------------------------------
<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Coefficient 1</th>
<th>Standard Error 1</th>
<th>Coefficient 2</th>
<th>Standard Error 2</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Chi-Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth_ft</td>
<td>1</td>
<td>-0.00019360</td>
<td>0.00010512</td>
<td>-1.84</td>
<td>0.0712</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log_estimated_days</td>
<td>1</td>
<td>0.57459</td>
<td>0.10455</td>
<td>5.50</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inc</td>
<td>1</td>
<td>0.00743</td>
<td>0.00271</td>
<td>2.74</td>
<td>0.0083</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp_BHT_F</td>
<td>1</td>
<td>0.01549</td>
<td>0.00627</td>
<td>2.47</td>
<td>0.0168</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Residual Plot: log Actual Days = Depth_ft log_Estimated_Days

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second
Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>12.71</td>
<td>0.5495</td>
</tr>
</tbody>
</table>
Residual Plot: log Actual Days = Depth_ft log_Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

Backward Elimination: Step 0

All Variables Entered: R-Square = 0.8257 and C(p) = 5.0000

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>4</td>
<td>30.19903</td>
<td>7.54976</td>
<td>61.56</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>52</td>
<td>6.37699</td>
<td>0.12263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter Estimate Standard Error Type II SS F Value Pr > F
Intercept -0.72931 0.49492 0.26629 2.17 0.1466
Depth_ft -0.00019360 0.00010512 0.41594 3.39 0.0712
log_estimated_days 0.57459 0.10455 3.70414 30.20 <.0001
inc 0.00743 0.00271 0.92345 7.53 0.0083
Temp_BHT_F 0.01549 0.00627 0.74869 6.11 0.0168

Bounds on condition number: 134.39, 1080.4

All variables left in the model are significant at the 0.1000 level.
Root MSE              0.35019    R-Square     0.8257
Dependent Mean        3.11950    Adj R-Sq     0.8122
Coeff Var             11.22590

Parameter Estimates

| Parameter     | DF | Estimate    | Standard Error | t Value | Pr > |t| |
|---------------|----|-------------|----------------|---------|------|---|
| Inflation     |    |             |                |         |      |   |
| Intercept     | 1  | -0.72931    | 0.49492        | -1.47   | 0.1466 |
| Depth_ft      | 1  | -0.00019360 | 0.00010512     | -1.84   | 0.0712 |
| log_estimated_days | 1  | 0.57459     | 0.10455        | 5.50    | <.0001 |
| inc           | 1  | 0.00743     | 0.00271        | 2.74    | 0.0083 |
| Temp_BHT_F    | 1  | 0.01549     | 0.00627        | 2.47    | 0.0168 |

Residual Plot: log Actual Days = Depth_ft log_Estimated_Days

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>12.71</td>
<td>0.5495</td>
</tr>
</tbody>
</table>
The UNIVARIATE Procedure
Variable:  resid  (Residual)

Moments

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>57</td>
<td>Sum Weights</td>
<td>57</td>
</tr>
<tr>
<td>Mean</td>
<td>0</td>
<td>Sum Observations</td>
<td>0</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>10.6075792</td>
<td>Variance</td>
<td>112.520738</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.190869</td>
<td>Kurtosis</td>
<td>0.97192839</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>6301.1613</td>
<td>Corrected SS</td>
<td>6301.1613</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>.</td>
<td>Std Error Mean</td>
<td>1.40500819</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td></td>
<td>Variability</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.000000</td>
<td>Std Deviation</td>
<td>10.60758</td>
</tr>
<tr>
<td>Median</td>
<td>0.693429</td>
<td>Variance</td>
<td>112.52074</td>
</tr>
<tr>
<td>Mode</td>
<td>.</td>
<td>Range</td>
<td>55.21750</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interquartile Range</td>
<td>9.66674</td>
</tr>
</tbody>
</table>

Tests for Location:  \( \mu_0 = 0 \)

<table>
<thead>
<tr>
<th>Test</th>
<th>-Statistic-</th>
<th>-----( p ) Value------</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's t</td>
<td>t</td>
<td>0</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>2.5</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>11.5</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</td>
<td>0.969891</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.095845</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.113858</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.662165</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>23.579001</td>
</tr>
<tr>
<td>99%</td>
<td>23.579001</td>
</tr>
<tr>
<td>95%</td>
<td>18.667073</td>
</tr>
<tr>
<td>90%</td>
<td>15.385396</td>
</tr>
</tbody>
</table>
Q-Q Plot: log Actual Days = Depth_ft log Estimated_Days Temp_BHT_F inc

The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% Q3</td>
<td>4.333042</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.693429</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-5.333699</td>
</tr>
<tr>
<td>10%</td>
<td>-11.201434</td>
</tr>
<tr>
<td>5%</td>
<td>-21.058065</td>
</tr>
<tr>
<td>1%</td>
<td>-31.638498</td>
</tr>
<tr>
<td>0% Min</td>
<td>-31.638498</td>
</tr>
</tbody>
</table>

Extreme Observations

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-31.6385</td>
<td>29</td>
<td>16.5369</td>
<td>8</td>
</tr>
<tr>
<td>-22.9302</td>
<td>26</td>
<td>18.3696</td>
<td>55</td>
</tr>
<tr>
<td>-21.0581</td>
<td>56</td>
<td>18.6671</td>
<td>4</td>
</tr>
<tr>
<td>-12.6892</td>
<td>11</td>
<td>22.4639</td>
<td>5</td>
</tr>
<tr>
<td>-12.2253</td>
<td>51</td>
<td>23.5790</td>
<td>35</td>
</tr>
</tbody>
</table>
Model G

Residual Plot: log Actual Days = Depth_ft log_Estimated_Days Temp_BHT_F inc

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

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>2</td>
<td>29.24001</td>
<td>14.62000</td>
<td>107.62</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>7.33601</td>
<td>0.13585</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.36858 R-Square 0.7994
Dependent Mean 3.11950 Adj R-Sq 0.7920
Coeff Var 11.81539

Parameter Estimates

| Parameter Variable | DF | Estimate | Error     | t Value | Pr > |t| |
|--------------------|----|----------|-----------|---------|-------|
| Inflation          |    |          |           |         |       |
| Intercept          | 1  | 0.18264  | 0.21158   | 0.86    | 0.3918|
| Temp_BHT_F         | 1  | 0.00408  | 0.00090007| 4.53    | <.0001|
| 2.50092            |    |          |           |         |       |
| log_estimated_days | 1  | 0.58106  | 0.10935   | 5.31    | <.0001|
| 2.50092            |    |          |           |         |       |

Residual Plot: log Actual Days = Depth_ft log_Estimated_Days Temp_BHT_F inc

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>7.69</td>
<td>0.1744</td>
</tr>
</tbody>
</table>
The UNIVARIATE Procedure

Variable: resid (Residual)

Moments

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>57</td>
<td>Sum Weights</td>
<td>57</td>
</tr>
<tr>
<td>Mean</td>
<td>0</td>
<td>Sum Observations</td>
<td>0</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>0.36193941</td>
<td>Variance</td>
<td>0.13100014</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.7239985</td>
<td>Kurtosis</td>
<td>0.67582857</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>7.33600764</td>
<td>Corrected SS</td>
<td>7.33600764</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>.</td>
<td>Std Error Mean</td>
<td>0.04794005</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.000000</td>
</tr>
<tr>
<td>Median</td>
<td>0.037210</td>
</tr>
<tr>
<td>Mode</td>
<td>.</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>0.36194</td>
</tr>
<tr>
<td>Variance</td>
<td>0.13100</td>
</tr>
<tr>
<td>Range</td>
<td>1.58922</td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>0.45505</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</td>
<td>0</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>2.5</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>64.5</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</td>
<td>0.954123</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.084488</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.06938</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.566471</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>0.5746979</td>
</tr>
<tr>
<td>99%</td>
<td>0.5746979</td>
</tr>
<tr>
<td>95%</td>
<td>0.5330984</td>
</tr>
<tr>
<td>90%</td>
<td>0.5299825</td>
</tr>
</tbody>
</table>
The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% Q3</td>
<td>0.2201254</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.0372096</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-0.2349214</td>
</tr>
<tr>
<td>10%</td>
<td>-0.4220720</td>
</tr>
<tr>
<td>5%</td>
<td>-0.8597671</td>
</tr>
<tr>
<td>1%</td>
<td>-1.0145186</td>
</tr>
<tr>
<td>0% Min</td>
<td>-1.0145186</td>
</tr>
</tbody>
</table>

 Extreme Observations

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.014519</td>
<td>29</td>
<td>0.530152</td>
<td>55</td>
</tr>
<tr>
<td>-0.910925</td>
<td>26</td>
<td>0.530468</td>
<td>40</td>
</tr>
<tr>
<td>-0.859767</td>
<td>56</td>
<td>0.533098</td>
<td>8</td>
</tr>
<tr>
<td>-0.674074</td>
<td>50</td>
<td>0.572543</td>
<td>32</td>
</tr>
<tr>
<td>-0.505949</td>
<td>45</td>
<td>0.574698</td>
<td>5</td>
</tr>
</tbody>
</table>
Model H

Residual Plot: log Actual Days = TEMP_BHT_F log_Estimated_Days

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

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>2</td>
<td>28.80412</td>
<td>14.40206</td>
<td>100.07</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>7.77190</td>
<td>0.14392</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.37937
R-Square 0.7875
Dependent Mean 3.11950
Adj R-Sq 0.7796
Coeff Var 12.16135

Parameter Estimates

| Parameter        | DF | Estimate | Error | t Value | Pr > |t|
|------------------|----|----------|-------|---------|-------|
| Intercept        | 1  | 1.30916  | 0.14239 | 9.19    | <.0001|
| Depth_ft         | 1  | 0.00010005 | 0.00001159 | 8.63    | <.0001|
| Estimated_Days   | 1  | 0.02003  | 0.00382 | 5.24    | <.0001|

Residual Plot: log Actual Days = TEMP_BHT_F log_Estimated_Days

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>9.81</td>
<td>0.0807</td>
</tr>
</tbody>
</table>
Q-Q Plot: Log Actual Days = Depth_ft Estimated_Days

The UNIVARIATE Procedure
Variable: resid (Residual)

Moments

<table>
<thead>
<tr>
<th>N</th>
<th>57</th>
<th>Sum Weights</th>
<th>57</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0</td>
<td>Sum Observations</td>
<td>0</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>0.37253708</td>
<td>Variance</td>
<td>0.13878398</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.2070043</td>
<td>Kurtosis</td>
<td>-0.1638041</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>7.77189712</td>
<td>Corrected SS</td>
<td>7.77189712</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>.</td>
<td>Std Error Mean</td>
<td>0.04934374</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.00000</td>
</tr>
<tr>
<td>Median</td>
<td>-0.06298</td>
</tr>
<tr>
<td>Mode</td>
<td>.</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</td>
<td>0</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>-3.5</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>11.5</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</td>
<td>0.973903</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.075896</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.045614</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.347221</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>0.6660560</td>
</tr>
<tr>
<td>99%</td>
<td>0.6660560</td>
</tr>
<tr>
<td>95%</td>
<td>0.6402173</td>
</tr>
<tr>
<td>90%</td>
<td>0.4977924</td>
</tr>
</tbody>
</table>
Q-Q Plot: Log Actual Days = Depth_ft Estimated_Days

The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% Q3</td>
<td>0.2581751</td>
</tr>
<tr>
<td>50% Median</td>
<td>-0.0629803</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-0.2160444</td>
</tr>
<tr>
<td>10%</td>
<td>-0.4350841</td>
</tr>
<tr>
<td>5%</td>
<td>-0.7915949</td>
</tr>
<tr>
<td>1%</td>
<td>-0.8455539</td>
</tr>
<tr>
<td>0% Min</td>
<td>-0.8455539</td>
</tr>
</tbody>
</table>

Extreme Observations

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.845554</td>
<td>26</td>
<td>0.605254</td>
<td>14</td>
</tr>
<tr>
<td>-0.837898</td>
<td>29</td>
<td>0.629209</td>
<td>8</td>
</tr>
<tr>
<td>-0.791595</td>
<td>56</td>
<td>0.640217</td>
<td>55</td>
</tr>
<tr>
<td>-0.680735</td>
<td>24</td>
<td>0.664484</td>
<td>40</td>
</tr>
<tr>
<td>-0.527839</td>
<td>45</td>
<td>0.666056</td>
<td>5</td>
</tr>
</tbody>
</table>
Model I

Residual Plot: Log Actual Days = Depth_ft Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

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>2</td>
<td>28.81429</td>
<td>14.40715</td>
<td>100.23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>7.76172</td>
<td>0.14374</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.37912    R-Square 0.7878
Dependent Mean 3.11950    Adj R-Sq 0.7799
Coeff Var 12.15338

Parameter Estimates

| Parameter    | DF | Estimate | Std Error | t Value | Pr > |t| |
|--------------|----|----------|-----------|---------|------|---|
| Inflation    |    |          |           |         |      |   |
| Intercept    | 1  | 0.88750  | 0.18191   | 4.88    | <.0001|
| Temp_BHT_F   | 1  | 0.00598  | 0.00069228| 8.64    | <.0001|
| Estimated_Days| 1 | 0.01893  | 0.00389   | 4.87    | <.0001|

Residual Plot: Log Actual Days = Depth_ft Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10.06</td>
<td>0.0734</td>
</tr>
</tbody>
</table>
The UNIVARIATE Procedure
Variable:  resid (Residual)

Moments

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>57</td>
<td>Sum Weights</td>
<td>57</td>
</tr>
<tr>
<td>Mean</td>
<td>0</td>
<td>Sum Observations</td>
<td>0</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>0.3722931</td>
<td>Variance</td>
<td>0.13860216</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.248246</td>
<td>Kurtosis</td>
<td>-0.0810346</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>7.76172071</td>
<td>Corrected SS</td>
<td>7.76172071</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>.</td>
<td>Std Error Mean</td>
<td>0.04931143</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td></td>
<td>Variability</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.00000</td>
<td>Std Deviation</td>
<td>0.37229</td>
</tr>
<tr>
<td>Median</td>
<td>-0.03973</td>
<td>Variance</td>
<td>0.13860</td>
</tr>
<tr>
<td>Mode</td>
<td>.</td>
<td>Range</td>
<td>1.52890</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interquartile Range</td>
<td>0.47775</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</td>
<td>0</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>-2.5</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>28.5</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</td>
<td>0.975171</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.059124</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.037815</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.306774</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>0.6614564</td>
</tr>
<tr>
<td>99%</td>
<td>0.6614564</td>
</tr>
<tr>
<td>95%</td>
<td>0.6540181</td>
</tr>
<tr>
<td>90%</td>
<td>0.4713515</td>
</tr>
</tbody>
</table>
The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

Quantile        Estimate
75% Q3         0.2530500
50% Median    -0.0397291
25% Q1        -0.2247016
10%            -0.4659558
5%            -0.8154419
1%            -0.8674445
0% Min        -0.8674445

Extreme Observations

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.867444</td>
<td>26</td>
<td>0.625395</td>
<td>55</td>
</tr>
<tr>
<td>-0.837098</td>
<td>29</td>
<td>0.625649</td>
<td>8</td>
</tr>
<tr>
<td>-0.815442</td>
<td>56</td>
<td>0.654018</td>
<td>5</td>
</tr>
<tr>
<td>-0.625418</td>
<td>24</td>
<td>0.660744</td>
<td>14</td>
</tr>
<tr>
<td>-0.563790</td>
<td>45</td>
<td>0.661456</td>
<td>40</td>
</tr>
</tbody>
</table>
Model J

Residual Plot: Log Actual Days = TEMP_BHT_F  Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

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>2</td>
<td>28.37448</td>
<td>14.18724</td>
<td>93.41</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>8.20153</td>
<td>0.15188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.38972    R-Square 0.7758
Dependent Mean 3.11950    Adj R-Sq 0.7675
Coeff Var 12.49297

Parameter Estimates

| Parameter       | DF | Estimate      | Standard Error | t Value | Pr > |t| |
|-----------------|----|---------------|----------------|---------|------|---|
| Intercept       | 1  | -4.13960      | 0.78034        | -5.30   | <.0001 |
| logdepth_ft     | 1  | 0.72552       | 0.08811        | 8.23    | <.0001 |
| Estimated_Days  | 1  | 0.02007       | 0.00395        | 5.09    | <.0001 |

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6.63</td>
<td>0.2500</td>
</tr>
</tbody>
</table>
Q-Q Plot: log Actual Days - logDepth_ft Estimated_Days

10:08 Thursday, August 12, 2010

The UNIVARIATE Procedure
Variable: resid (Residual)

Moments

N Sum Weights
57 57
Mean Sum Observations
0 0
Std Deviation Variance
0.38269556 0.14645589
Skewness Kurtosis
-0.4134658 -0.0197204
Uncorrected SS Corrected SS
8.20152993 8.20152993
Coeff Variation Std Error Mean
. 0.05068927

Basic Statistical Measures

Location Variability
Mean Std Deviation
0.000000 0.38270
Median Variance
0.033453 0.14646
Mode Range
. 1.66504
Interquartile Range 0.53982

Tests for Location: Mu0=0

Test -Statistic- -----p Value------
Student's t t 0 Pr > |t| 1.0000
Sign M 2.5 Pr >= |M| 0.5966
Signed Rank S 26.5 Pr >= |S| 0.8354

Tests for Normality

Test --Statistic-- -----p Value------
Shapiro-Wilk W 0.969709 Pr < W 0.1627
Kolmogorov-Smirnov D 0.071677 Pr > D >0.1500
Cramer-von Mises W-Sq 0.047175 Pr > W-Sq >0.2500
Anderson-Darling A-Sq 0.385826 Pr > A-Sq >0.2500

Quantiles (Definition 5)

Quantile Estimate
100% Max 0.6181359
99% 0.6181359
95% 0.5775079
90% 0.5254011
The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% Q3</td>
<td>0.2640075</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.0334534</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-0.2758166</td>
</tr>
<tr>
<td>10%</td>
<td>-0.4524619</td>
</tr>
<tr>
<td>5%</td>
<td>-0.7773676</td>
</tr>
<tr>
<td>1%</td>
<td>-1.0469089</td>
</tr>
<tr>
<td>0% Min</td>
<td>-1.0469089</td>
</tr>
</tbody>
</table>

Extreme Observations

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.046909</td>
<td>29</td>
<td>0.556112</td>
<td>40</td>
</tr>
<tr>
<td>-0.823864</td>
<td>26</td>
<td>0.561783</td>
<td>35</td>
</tr>
<tr>
<td>-0.777368</td>
<td>56</td>
<td>0.577508</td>
<td>55</td>
</tr>
<tr>
<td>-0.679165</td>
<td>50</td>
<td>0.615893</td>
<td>49</td>
</tr>
<tr>
<td>-0.474580</td>
<td>45</td>
<td>0.618136</td>
<td>5</td>
</tr>
</tbody>
</table>
Model K

Residual Plot: log Actual Days = logDepth_ft Estimated_Days
63
10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

Analysis of Variance

Source DF Sum of Squares Mean Square F Value Pr > F
Model 2 29.17926 14.58963 106.51 <.0001
Error 54 7.39676 0.13698
Corrected Total 56 36.57601

Root MSE 0.37010 R-Square 0.7978
Dependent Mean 3.11950 Adj R-Sq 0.7903
Coeff Var 11.86421

Parameter Estimates

Parameter DF Estimate Standard Error t Value Pr > |t|
Inflation
Intercept 1 -4.98991 0.80789 -6.18 <.0001
logTEMP_BHT_F 1 1.36169 0.15125 9.00 <.0001
1.41380
Estimated_Days 1 0.01828 0.00382 4.79 <.0001
1.41380

Test of First and Second Moment Specification

DF Chi-Square Pr > ChiSq
5 8.88 0.1138
Q-Q Plot: \( \log \text{Actual Days} = \log \text{TEMP\_BHT\_F} \)  Estimated_Days

The UNIVARIATE Procedure

Variable: resid (Residual)

Moments

<table>
<thead>
<tr>
<th>Moment</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>57</td>
</tr>
<tr>
<td>Sum Weights</td>
<td>57</td>
</tr>
<tr>
<td>Mean</td>
<td>0.000000</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>0.36343491</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.4378214</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.18934783</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>7.39675609</td>
</tr>
<tr>
<td>Corrected SS</td>
<td>7.39675609</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>0.04813813</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Variability</td>
</tr>
<tr>
<td>Mean</td>
<td>0.000000</td>
</tr>
<tr>
<td>Median</td>
<td>0.002799</td>
</tr>
<tr>
<td>Mode</td>
<td></td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>0.44930</td>
</tr>
<tr>
<td>Range</td>
<td>1.57006</td>
</tr>
<tr>
<td>Variance</td>
<td>0.13208</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>0.36343</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td></td>
</tr>
</tbody>
</table>

Tests for Location: \( \mu_0=0 \)

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's t</td>
<td>t</td>
<td>0</td>
</tr>
<tr>
<td>Sign</td>
<td>M</td>
<td>0.5</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S</td>
<td>37.5</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</td>
<td>0.971689</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.071257</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.030583</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.30949</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>0.62106691</td>
</tr>
<tr>
<td>99%</td>
<td>0.62106691</td>
</tr>
<tr>
<td>95%</td>
<td>0.58773308</td>
</tr>
<tr>
<td>90%</td>
<td>0.52869277</td>
</tr>
</tbody>
</table>
Q-Q Plot: log Actual Days = logTEMP_BHT_F, Estimated_Days

The UNIVARIATE Procedure
Variable: resid (Residual)

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>75% Q3</td>
<td>0.23482423</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.00279929</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-0.21447374</td>
</tr>
<tr>
<td>10%</td>
<td>-0.44734575</td>
</tr>
<tr>
<td>5%</td>
<td>-0.81615983</td>
</tr>
<tr>
<td>1%</td>
<td>-0.94898981</td>
</tr>
<tr>
<td>0% Min</td>
<td>-0.94898981</td>
</tr>
</tbody>
</table>

Extreme Observations

<table>
<thead>
<tr>
<th>Value</th>
<th>Obs</th>
<th>Value</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.948990</td>
<td>29</td>
<td>0.564515</td>
<td>8</td>
</tr>
<tr>
<td>-0.862051</td>
<td>26</td>
<td>0.581635</td>
<td>55</td>
</tr>
<tr>
<td>-0.816160</td>
<td>56</td>
<td>0.587733</td>
<td>14</td>
</tr>
<tr>
<td>-0.548144</td>
<td>50</td>
<td>0.588273</td>
<td>40</td>
</tr>
<tr>
<td>-0.546199</td>
<td>45</td>
<td>0.621067</td>
<td>5</td>
</tr>
</tbody>
</table>

10:08 Thursday, August 12, 2010
Testing Chosen Models with a New Variable that equals the Chosen Regressor Variables Multiplied Together (No Additional Significance was Found)

Residual Plot: $\log \text{Actual Days} = \log \text{TEMP_BHT_F} \cdot \text{Estimated_Days}$

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Number of Observations Read 57
Number of Observations Used 57

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>3</td>
<td>16189</td>
<td>5396.18685</td>
<td>47.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>53</td>
<td>6019.68506</td>
<td>113.57896</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>22208</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 10.65734
Dependent Mean 29.49123
Coeff Var 36.13733
R-Square 0.7289
Adj R-Sq 0.7136

Parameter Estimates

| Parameter        | DF | Estimate | Standard Error | t Value | Pr > |t| |
|------------------|----|----------|----------------|---------|------|----|
| Intercept        | 1  | 0.18262  | 5.10120        | 0.04    | 0.9716 |
| Depth_ft         | 1  | 0.00075006 | 0.00036517    | 2.05    | 0.0449 |
| Estimated_Days   | 1  | 0.26043  | 0.37292        | 0.70    | 0.4880 |
| DepthbyEst_days  | 1  | 0.00003660 | 0.00002120    | 1.73    | 0.0901 |

Residual Plot: $\log \text{Actual Days} = \log \text{TEMP_BHT_F} \cdot \text{Estimated_Days}$

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: ActualDays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>13.29</td>
<td>0.1024</td>
</tr>
</tbody>
</table>

Residual Plot: $\log \text{Actual Days} = \log \text{TEMP_BHT_F} \cdot \text{Estimated_Days}$
The REG Procedure

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations Read</td>
<td>57</td>
</tr>
<tr>
<td>Number of Observations Used</td>
<td>0</td>
</tr>
<tr>
<td>Number of Observations with Missing Values</td>
<td>57</td>
</tr>
</tbody>
</table>
The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

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>3</td>
<td>28.49568</td>
<td>9.49856</td>
<td>62.30</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>53</td>
<td>8.08033</td>
<td>0.15246</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.39046
R-Square 0.7791
Dependent Mean 3.11950
Adj R-Sq 0.7666
Coeff Var 12.51675

Parameter Estimates

| Parameter          | DF | Estimate | Standard Error | t Value | Pr > |t| |
|--------------------|----|----------|----------------|---------|-------|
| Intercept          | 1  | -3.69290 | 0.92858        | -3.98   | 0.0002|
| logdepth_ft        | 1  | 0.68257  | 0.10057        | 6.79    | <.0001|
| Estimated_Days     | 1  | -0.07912 | 0.11132        | -0.71   | 0.4804|
| logdepthbyEstDays  |    | 0.01017  | 0.01140        | 0.89    | 0.3766|
| 1081.45512         |    |          |                |         |        |
| 1103.14258         |    |          |                |         |        |

Residual Plot: log Actual Days = logTEMP_BHT_F Estimated_Days

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>11.14</td>
<td>0.1940</td>
</tr>
</tbody>
</table>

Residual Plot: log Actual Days = logTEMP_BHT_F Estimated_Days

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 57

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>3</td>
<td>29.18034</td>
<td>9.72678</td>
<td>69.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>53</td>
<td>7.39568</td>
<td>0.13954</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.37355
R-Square 0.7978
Dependent Mean 3.11950
Adj R-Sq 0.7864
Coeff Var 11.97474

Parameter Estimates

| Variance Variable | DF | Estimate  | Standard Error | t Value | Pr > |t| |
|-------------------|----|-----------|----------------|---------|------|---|
| Inflation         |    | -5.04516  | 1.02928        | -4.90   | <.0001|   |
| logTEMP_BHT_F     | 1  | 1.37080   | 0.18444        | 7.43    | <.0001|   |
| 2.06368           |    | 0.02819   | 0.11266        | 0.25    | 0.8034|   |
| Estimated_Days    | 1  | -0.00102  | 0.01157        | -0.09   | 0.9302|   |
| 1210.24741        |    |           |                |         |       |   |
| logdepthbyEstDays | 1  | -0.00102  | 0.01157        | -0.09   | 0.9302|   |
| 1240.81068        |    |           |                |         |       |   |

Residual Plot: log Actual Days = logTEMP_BHT_F  Estimated_Days
10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>11.73</td>
<td>0.2290</td>
</tr>
</tbody>
</table>

Residual Plot: log Actual Days = logTEMP_BHT_F  Estimated_Days
10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Analysis of Variance
### Source Table

<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>3</td>
<td>29.27033</td>
<td>9.75678</td>
<td>70.78</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>53</td>
<td>7.30569</td>
<td>0.13784</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Root MSE: 0.37127
- R-Square: 0.8003
- Dependent Mean: 3.11950
- Adj R-Sq: 0.7890
- Coeff Var: 11.90166

### Parameter Estimates

| Parameter                  | DF | Estimate     | Standard Error | t Value | Pr > |t| |
|----------------------------|----|--------------|----------------|---------|------|---|
| Intercept                  | 1  | 1.10633      | 0.17771        | 6.23    | <.0001|
| Depth_ft                   | 1  | 0.00011065   | 0.00001272     | 8.70    | <.0001|
| Estimated_Days             | 1  | 0.04291      | 0.01299        | 3.30    | 0.0017|
| DepthbyEst_days            | 1  | -0.00000136  | 7.385924E-7    | -1.84   | 0.0715|
Residual Plot: log Actual Days = logTEMP_BHT_F Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second
Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>6.88</td>
<td>0.5495</td>
</tr>
</tbody>
</table>

Residual Plot: log Actual Days = logTEMP_BHT_F Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Number of Observations Read 57
Number of Observations Used 0
Number of Observations with Missing Values 57

Residual Plot: log Actual Days = logTEMP_BHT_F Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

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>2</td>
<td>29.17926</td>
<td>14.58963</td>
<td>106.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>54</td>
<td>7.39676</td>
<td>0.13698</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>56</td>
<td>36.57601</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.37010 R-Square 0.7978
Dependent Mean 3.11950 Adj R-Sq 0.7903
Coeff Var 11.86421

Parameter Estimates

| Parameter     | Standard Error | t Value | Pr > |t| |
|---------------|----------------|---------|------|---|
| Inflation     |                |         |      |   |
| Intercept     | 1              | -4.98991| 0.80789| -6.18| <.0001|
| logTEMP_BHT_F | 1              | 1.36169 | 0.15125| 9.00 | <.0001|

1.41380
<table>
<thead>
<tr>
<th>Estimated_Days</th>
<th>1</th>
<th>0.01828</th>
<th>0.00382</th>
<th>4.79</th>
<th>&lt;.0001</th>
</tr>
</thead>
</table>

Residual Plot: log Actual Days = logTEMP_BHT_F * Estimated_Days

10:08 Thursday, August 12, 2010

The REG Procedure
Model: MODEL1
Dependent Variable: logactualdays

Test of First and Second Moment Specification

<table>
<thead>
<tr>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>8.88</td>
<td>0.1138</td>
</tr>
</tbody>
</table>
APPENDIX F

SAS CODE NECESSARY FOR REGRESSIONS AND UNIVARIATE CALCULATIONS

```
proc contents data=drilling;
run;

proc reg data=drilling;
model actualdays = Depth_ft Estimated_Days inc TEMP_BHT_F
/output out=results r=resid p=fits;
run;
quit;

proc reg data=drilling;
model actualdays = Depth_ft Estimated_Days inc TEMP_BHT_F
/output out=results r=resid p=fits;
run;
quit;

proc reg data=drilling;
model actualdays = Depth_ft Estimated_Days inc TEMP_BHT_F
/output out=results r=resid p=fits;
run;
quit;

proc reg data=drilling;
model actualdays = TEMP_BHT_F  Estimated_Days/spec vif;
run;
quit;

proc univariate data=results normaltest ;
var resid;
title 'Q-Q Plot: Actualdays = TEMP_BHT_F  Estimated_Days';
probpplot resid / normal(mu=est sigma=est);
run;

proc gplot data=results;
title 'Residual Plot: Actualdays = TEMP_BHT_F  Estimated_Days';
plot resid*fits;
run;

proc reg data=drilling;
model actualdays = Depth_ft Estimated_Days/spec vif;
output out=results r=resid p=fits;
run;
quit;

proc univariate data=results normaltest ;
var resid;
```
title 'Q-Q Plot: Actualdays = Depth_ft Estimated_Days';
probplot resid / normal(mu=est sigma=est);
run;

proc gplot data=results;
title 'Residual Plot: Actualdays = Depth_ft Estimated_Days';
plot resid*fits;
run;
/* removal of outlier */
data drilling2;
set drilling;
where actualdays ne 174;
run;

proc reg data=drilling2;
model actualdays = Depth_ft Estimated_Days /spec vif;
output out=results r=resid p=fits;
run;
quit;

proc univariate data=results normaltest ;
var resid;
title 'Q-Q Plot: Actualdays = Depth_ft Estimated_Days';
probplot resid / normal(mu=est sigma=est);
run;

proc gplot data=results;
title 'Residual Plot: Actualdays = Depth_ft Estimated_Days';
plot resid*fits;
run;

proc reg data=drilling2;
model actualdays = TEMP_BHT_F  Estimated_Days /spec vif ;
output out=results r=resid p=fits;
run;
quit;

proc univariate data=results normaltest ;
var resid;
title 'Q-Q Plot: Actualdays = TEMP_BHT_F  Estimated_Days';
probplot resid / normal(mu=est sigma=est);
run;

proc gplot data=results;
title 'Residual Plot: Actualdays = TEMP_BHT_F  Estimated_Days';
plot resid*fits;
run;

proc gplot data=drilling2;
title 'Actual Days vs. Estimated Days';
plot actualdays*estimated_days;
run;
proc gplot data=drilling2;
title 'Actual Days vs. Depth_ft';
plot actualdays*depth_ft;
run;
proc gplot data=drilling2;
title 'Actual Days vs. TEMP_BHT_F';
plot actualdays*TEMP_BHT_F;
run;

data drilling3;
set drilling2;
logactualdays=log(actualdays);
exp_estimated_days=exp(estimated_days);
estimated_days_sq=estimated_days**2;
log_estimated_days=log(estimated_days);
logdepth_ft=log(Depth_ft);
logTEMP_BHT_F=log(TEMP_BHT_F);
run;

proc gplot data=drilling3;
title 'logactualdays*log_estimated_days';
plot logactualdays*log_estimated_days;
run;
proc gplot data=drilling3;
title 'logactualdays*estimated_days_sq';
plot logactualdays*estimated_days_sq;
run;
title 'logactualdays*depth_ft';
plot logactualdays*depth_ft;
proc gplot data=drilling3;
title 'logactualdays*TEMP_BHT_F';
plot logactualdays*TEMP_BHT_F;
run;
proc gplot data=drilling3;
title 'logactualdays*inc';
plot logactualdays*inc;
run;
proc gplot data=drilling3;
title 'logactualdays*logDepth_ft';
plot logactualdays*logDepth_ft;
run;

proc reg data=drilling3;
model logactualdays = Depth_ft log_Estimated_Days inc TEMP_BHT_F
/selection=stepwise spec vif;
run;
quit;
proc univariate data=results normaltest ;
var resid;
title 'Q-Q Plot: log Actual Days = Depth_ft log_Estimated_Days';
probplot resid / normal(mu=est sigma=est);
run;

proc gplot data=results;
title 'Residual Plot: log Actual Days = Depth_ft log_Estimated_Days';
plot resid*fits;
proc reg data=drilling3;
model logactualdays = Depth_ft log_Estimated_Days inc TEMP_BHT_F /selection=forward spec vif;
run;
quit;

proc reg data=drilling3;
model logactualdays = Depth_ft log_Estimated_Days inc TEMP_BHT_F /selection=backward spec vif;
run;
quit;

proc univariate data=results normaltest ;
var resid;
title 'Q-Q Plot: log Actual Days = Depth_ft log_Estimated_Days Temp_BHT_F inc';
probplot resid / normal(mu=est sigma=est);
run;

proc gplot data=results;
title 'Residual Plot: log Actual Days = Depth_ft log_Estimated_Days Temp_BHT_F inc';
plot resid*fits;
run;

proc reg data=drilling3;
model logactualdays = TEMP_BHT_F log_Estimated_Days /spec vif ;
output out=results r=resid p=fits;
run;
quit;

proc univariate data=results normaltest ;
var resid;
title 'Q-Q Plot: log Actual Days = TEMP_BHT_F log_Estimated_Days';
probplot resid / normal(mu=est sigma=est);
run;

proc gplot data=results;
title 'Residual Plot: log Actual Days = TEMP_BHT_F log_Estimated_Days';
plot resid*fits;
run;

proc reg data=drilling3;
model logactualdays = depth_ft Estimated_Days /spec vif ;
output out=results r=resid p=fits;
run;
quit;

proc univariate data=results normaltest ;
var resid;
title 'Q-Q Plot: Log Actual Days = Depth_ft Estimated_Days';
probplot resid / normal(mu=est sigma=est);
run;
PROC GPLOT DATA=RESULTS;
   TITLE 'Residual Plot: Log Actual Days = Depth_ft Estimated_Days';
   PLOT RESID*FITS;
RUN;
PROC REG DATA=DRILLING3;
   MODEL LOGACTUALDAYS = TEMP_BHT_F  ESTIMATED_DAYS /SPEC VIF ;
   OUTPUT OUT=RESULTS R=RESID P=FITS;
RUN;
QUIT;

PROC UNIVARIATE DATA=RESULTS NORMALTEST;
   VAR RESID;
   TITLE 'Q-Q Plot: Log Actual Days = TEMP_BHT_F  Estimated_Days';
   PROBPLOT RESID / NORMAL(MU=EST SIGMA=EST);
RUN;
PROC GPLOT DATA=RESULTS;
   TITLE 'Residual Plot: Log Actual Days = TEMP_BHT_F  Estimated_Days';
   PLOT RESID*FITS;
RUN;
PROC REG DATA=DRILLING3;
   MODEL LOGACTUALDAYS = LOGDEPTH_FT ESTIMATED_DAYS /SPEC VIF ;
   OUTPUT OUT=RESULTS R=RESID P=FITS;
RUN;
QUIT;

PROC UNIVARIATE DATA=RESULTS NORMALTEST;
   VAR RESID;
   TITLE 'Q-Q Plot: log Actual Days = logDepth_ft Estimated_Days';
   PROBPLOT RESID / NORMAL(MU=EST SIGMA=EST);
RUN;
PROC GPLOT DATA=RESULTS;
   TITLE 'Residual Plot: log Actual Days = logDepth_ft Estimated_Days';
   PLOT RESID*FITS;
RUN;
PROC REG DATA=DRILLING3;
   MODEL LOGACTUALDAYS = LOGTEMP_BHT_F  ESTIMATED_DAYS /SPEC VIF ;
   OUTPUT OUT=RESULTS R=RESID P=FITS;
RUN;
QUIT;

PROC UNIVARIATE DATA=RESULTS NORMALTEST;
   VAR RESID;
   TITLE 'Q-Q Plot: log Actual Days = logTEMP_BHT_F  Estimated_Days';
   PROBPLOT RESID / NORMAL(MU=EST SIGMA=EST);
RUN;
PROC GPLOT DATA=RESULTS;
   TITLE 'Residual Plot: log Actual Days = logTEMP_BHT_F  Estimated_Days';
   PLOT RESID*FITS;
run;

data drilling3wint;
set drilling3;
DepthbyEst_days=Depth_ft*Estimated_Days;
EstDaysbyBHT=Estimated_Days*Temp_bht;
logTempBHTbyEst=logTEMP_BHT_F*Estimated_Days;
logdepthbyEstDays=logDepth_ft*Estimated_Days;
run;

proc reg data=drilling3wint;
model actualdays = Depth_ft Estimated_Days DepthbyEst_days/spec vif;
output out=results r=resid p=fits;
run;
quit;

proc reg data=drilling3wint;
model actualdays = TEMP_BHT_F Estimated_Days EstDaysbyBHT /spec vif;
output out=results r=resid p=fits;
run;
quit;

proc reg data=drilling3wint;
model logactualdays = logDepth_ft Estimated_Days logdepthbyEstDays/spec vif;
output out=results r=resid p=fits;
run;
quit;

proc reg data=drilling3wint;
model logactualdays = logTEMP_BHT_F Estimated_Days logdepthbyEstDays/spec vif;
output out=results r=resid p=fits;
run;
quit;

proc reg data=drilling3wint;
model logactualdays = depth_ft Estimated_Days DepthbyEst_days/spec vif;
output out=results r=resid p=fits;
run;
quit;

proc reg data=drilling3wint;
model logactualdays = TEMP_BHT_F Estimated_Days EstDaysbyBHT/spec vif;
output out=results r=resid p=fits;
run;
quit;

proc reg data=drilling3wint;
model logactualdays = logTEMP_BHT_F Estimated_Days /spec vif;
output out=results r=resid p=fits;
run;
ods graphics on;

proc corr data=drilling plots=matrix;
var actualdays estimated_days TEMP_bht_f Depth_ft inc;
run;
APPENDIX G

AN EXAMPLE OF R’S CODE AND OUTPUT FOR REGRESSIONS AND UNIVARIATE CALCULATIONS

```r
> drilling = read.csv(file.choose(), header=T)
> attach(drilling)
>
> Reg1<-lm(ActualDays~Estimated.Days+Depth.ft,drilling)
> full<-lm(ActualDays~,drilling)
> Reg2=step(full)
Start:  AIC=353.78
ActualDays ~ Estimated.Days + Depth.ft + Temp.BHT.F + inc

Df Sum of Sq RSS   AIC
- Depth.ft 1   298.1 22054 352.57
- inc      1   311.5 22067 352.60
- Temp.BHT.F 1  501.5 22257 353.10
<none>      21755 353.78
- Estimated.Days 1  7079.6 28835 368.12

Step:  AIC=352.57
ActualDays ~ Estimated.Days + Temp.BHT.F + inc

Df Sum of Sq RSS   AIC
- inc      1   43.3 22097 350.68
<none>    22054 352.57
- Temp.BHT.F 1  2590.8 24644 357.01
- Estimated.Days 1  7321.1 29375 367.19

Step:  AIC=350.68
ActualDays ~ Estimated.Days + Temp.BHT.F

Df Sum of Sq RSS   AIC
<none>    22097 350.68
- Temp.BHT.F 1  2610.8 24708 355.16
- Estimated.Days 1  7303.2 29400 365.24
> null<-lm(ActualDays~1,drilling)
> Reg3=step(null,scope=list(lower=null,upper=full,direction="forward"))
Start:  AIC=384.93
ActualDays ~ 1
```
```
+ Estimated.Days  1   18023   24708   355.16
+ Temp.BHT.F     1    13331   29400   365.24
+ Depth.ft      1    12476   30255   366.90
<none>                        42731   384.93
+ inc            1    129    42602   386.75

Step:  AIC=355.16
ActualDays ~ Estimated.Days

+ Temp.BHT.F     1   2610.8  22097   350.68
+ Depth.ft      1   2450.7  22257   351.10
<none>                        24708   355.16
+ inc            1    63.3  24644   357.01
- Estimated.Days  1   18023.4 42731   384.93

Step:  AIC=350.68
ActualDays ~ Estimated.Days + Temp.BHT.F

<none>                        22097   350.68
+ inc            1   43.3  22054   352.57
+ Depth.ft      1    29.9  22067   352.60
- Temp.BHT.F     1   2610.8  24708   355.16
- Estimated.Days  1   7303.2 29400   365.24
> Reg4=step(full,direction="backward")
Start:  AIC=353.78
ActualDays ~ Estimated.Days + Depth.ft + Temp.BHT.F + inc

- Depth.ft      1    298.1  22054   352.57
- inc            1  311.5  22067   352.60
- Temp.BHT.F     1    501.5  22257   353.10
<none>                        21755   353.78
- Estimated.Days  1   7079.6 28835   368.12

Step:  AIC=352.57
ActualDays ~ Estimated.Days + Temp.BHT.F + inc

- inc            1   43.3  22097   350.68
<none>                        22054   352.57
- Temp.BHT.F     1   2590.8  24644   357.01
```
- Estimated.Days  1 7321.1 29375 367.19

Step:  AIC=350.68
ActualDays ~ Estimated.Days + Temp.BHT.F

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>RSS</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;none&gt;</td>
<td>22097</td>
<td>350.68</td>
<td></td>
</tr>
<tr>
<td>- Temp.BHT.F</td>
<td>1</td>
<td>2610.8</td>
<td>24708</td>
</tr>
<tr>
<td>- Estimated.Days</td>
<td>1</td>
<td>7303.2</td>
<td>29400</td>
</tr>
</tbody>
</table>

> Reg5=step(full,scale=(summary(full)$sigma)^2)
Start:  AIC=5
ActualDays ~ Estimated.Days + Depth.ft + Temp.BHT.F + inc

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>RSS</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Depth.ft</td>
<td>1</td>
<td>298.1</td>
<td>22054</td>
</tr>
<tr>
<td>- inc</td>
<td>1</td>
<td>311.5</td>
<td>22067</td>
</tr>
<tr>
<td>- Temp.BHT.F</td>
<td>1</td>
<td>501.5</td>
<td>22257</td>
</tr>
<tr>
<td>&lt;none&gt;</td>
<td>21755</td>
<td>5.0000</td>
<td></td>
</tr>
<tr>
<td>- Estimated.Days</td>
<td>1</td>
<td>7079.6</td>
<td>28835</td>
</tr>
</tbody>
</table>

Step:  AIC=3.73
ActualDays ~ Estimated.Days + Temp.BHT.F + inc

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>RSS</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>- inc</td>
<td>1</td>
<td>43.3</td>
<td>22097</td>
</tr>
<tr>
<td>&lt;none&gt;</td>
<td>22054</td>
<td>3.7263</td>
<td></td>
</tr>
<tr>
<td>- Temp.BHT.F</td>
<td>1</td>
<td>2590.8</td>
<td>24644</td>
</tr>
<tr>
<td>- Estimated.Days</td>
<td>1</td>
<td>7321.1</td>
<td>29375</td>
</tr>
</tbody>
</table>

Step:  AIC=1.83
ActualDays ~ Estimated.Days + Temp.BHT.F

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>RSS</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;none&gt;</td>
<td>22097</td>
<td>1.8317</td>
<td></td>
</tr>
<tr>
<td>- Temp.BHT.F</td>
<td>1</td>
<td>2610.8</td>
<td>24708</td>
</tr>
<tr>
<td>- Estimated.Days</td>
<td>1</td>
<td>7303.2</td>
<td>29400</td>
</tr>
</tbody>
</table>

> Reg6=step(null,scope=list(lower=null,upper=full,direction="forward"),scale=(summary(full)$sigma)^2)
Start:  AIC=48.1
ActualDays ~ 1

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>RSS</th>
<th>Cp</th>
</tr>
</thead>
</table>
+ Estimated.Days 1 18023 24708 6.1922
+ Temp.BHT.F 1 13331 29400 17.6236
+ Depth.ft 1 12476 30255 19.7055
<none> 42731 48.1004
+ inc 1 129 42602 49.7862

Step: AIC=6.19
ActualDays ~ Estimated.Days

<table>
<thead>
<tr>
<th>Df Sum of Sq</th>
<th>RSS</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp.BHT.F</td>
<td>2610.8</td>
<td>22097</td>
</tr>
<tr>
<td>Depth.ft</td>
<td>2450.7</td>
<td>22257</td>
</tr>
<tr>
<td>inc</td>
<td>63.3</td>
<td>24644</td>
</tr>
<tr>
<td>Estimated.Days</td>
<td>18023.4</td>
<td>42731</td>
</tr>
</tbody>
</table>

Step: AIC=1.83
ActualDays ~ Estimated.Days + Temp.BHT.F

<table>
<thead>
<tr>
<th>Df Sum of Sq</th>
<th>RSS</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>inc</td>
<td>43.3</td>
<td>22054</td>
</tr>
<tr>
<td>Depth.ft</td>
<td>29.9</td>
<td>22067</td>
</tr>
<tr>
<td>Temp.BHT.F</td>
<td>2610.8</td>
<td>24708</td>
</tr>
<tr>
<td>Estimated.Days</td>
<td>7303.2</td>
<td>29400</td>
</tr>
</tbody>
</table>

> Reg7=step(full,direction="backward",scale=(summary(full)$sigma)^2)
Start: AIC=5
ActualDays ~ Estimated.Days + Depth.ft + Temp.BHT.F + inc

<table>
<thead>
<tr>
<th>Df Sum of Sq</th>
<th>RSS</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth.ft</td>
<td>298.1</td>
<td>22054</td>
</tr>
<tr>
<td>inc</td>
<td>311.5</td>
<td>22067</td>
</tr>
<tr>
<td>Temp.BHT.F</td>
<td>501.5</td>
<td>22257</td>
</tr>
<tr>
<td>Estimated.Days</td>
<td>7079.6</td>
<td>28835</td>
</tr>
</tbody>
</table>

Step: AIC=3.73
ActualDays ~ Estimated.Days + Temp.BHT.F + inc

<table>
<thead>
<tr>
<th>Df Sum of Sq</th>
<th>RSS</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>inc</td>
<td>43.3</td>
<td>22097</td>
</tr>
<tr>
<td>Temp.BHT.F</td>
<td>2590.8</td>
<td>24644</td>
</tr>
<tr>
<td>Estimated.Days</td>
<td>7321.1</td>
<td>29375</td>
</tr>
</tbody>
</table>
Step: AIC=1.83
ActualDays ~ Estimated.Days + Temp.BHT.F

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>RSS</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;none&gt;</td>
<td>22097</td>
<td>1.8317</td>
<td></td>
</tr>
<tr>
<td>- Temp.BHT.F</td>
<td>1</td>
<td>2610.8</td>
<td>24708</td>
</tr>
<tr>
<td>- Estimated.Days</td>
<td>1</td>
<td>7303.2</td>
<td>29400</td>
</tr>
</tbody>
</table>

```r
par(mfrow=c(2,2))
plot.lm(Reg7)
plot(Reg7$res,ylab="Residual",xlab="Predicted value of ActualDays",main="Residual Plot: ActualDays = Depth_ft Estimated_Days")

a<-Reg7$res
u<-(1:length(a)-.5)/length(a)
Q<-quantile(a,u,type=5)
plot(qnorm(u),Q)
abline(lm(Q~qnorm(u)))

SWReg7=shapiro.test(Reg7$res)
SWReg7

Shapiro-Wilk normality test
data:  Reg7$res
W = 0.6315, p-value = 8.311e-11

PTReg7=pearson.test(Reg7$res)
PTReg7

Pearson chi-square normality test
data:  Reg7$res
P = 34.1724, p-value = 3.781e-05

summary(Reg7)
```

Call:
`lm(formula = ActualDays ~ Estimated.Days + Temp.BHT.F, data = drilling)`

Residuals:
```
 Min  1Q Median   3Q  Max
```
-32.965 -7.853 -2.730 6.002 123.031

Coefficients:

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | -16.84055| 9.55474    | -1.763  | 0.0835   |
| Estimated.Days | 0.87604  | 0.20547    | 4.264   | 7.96e-05 |
| Temp.BHT.F     | 0.09255  | 0.03630    | 2.549   | 0.0136   |

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.04 on 55 degrees of freedom
Multiple R-squared: 0.4829, Adjusted R-squared: 0.4641
F-statistic: 25.68 on 2 and 55 DF, p-value: 1.329e-08

> plot(Depth.ft,ActualDays)
> lndepth<-log(Depth.ft)
> logdepth<-log10(Depth.ft)
> depthsqr<-(Depth.ft)^2

> drilling2<-(cbind(drilling,lndepth))
> drilling2

<table>
<thead>
<tr>
<th>ActualDays</th>
<th>Estimated.Days</th>
<th>Depth.ft</th>
<th>Temp.BHT.F</th>
<th>inc</th>
<th>lndepth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>10260</td>
<td>249.420</td>
<td>24.600000</td>
<td>9.236008</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>14496</td>
<td>321.432</td>
<td>2.150000</td>
<td>9.581628</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>16025</td>
<td>347.425</td>
<td>2.000000</td>
<td>9.681905</td>
</tr>
<tr>
<td>4</td>
<td>54</td>
<td>19250</td>
<td>402.250</td>
<td>1.500000</td>
<td>9.865266</td>
</tr>
<tr>
<td>5</td>
<td>64</td>
<td>14620</td>
<td>323.540</td>
<td>33.850000</td>
<td>9.590146</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>16350</td>
<td>352.950</td>
<td>2.500000</td>
<td>9.701983</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>18700</td>
<td>392.900</td>
<td>2.990000</td>
<td>9.836279</td>
</tr>
<tr>
<td>8</td>
<td>62</td>
<td>13470</td>
<td>303.990</td>
<td>21.370000</td>
<td>9.508220</td>
</tr>
<tr>
<td>9</td>
<td>28</td>
<td>15449</td>
<td>337.633</td>
<td>3.750000</td>
<td>9.645300</td>
</tr>
<tr>
<td>10</td>
<td>32</td>
<td>18100</td>
<td>382.700</td>
<td>7.000000</td>
<td>9.803667</td>
</tr>
<tr>
<td>11</td>
<td>17</td>
<td>10820</td>
<td>258.940</td>
<td>1.250000</td>
<td>9.289152</td>
</tr>
<tr>
<td>12</td>
<td>32</td>
<td>13460</td>
<td>301.066</td>
<td>76.800000</td>
<td>9.507478</td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>16102</td>
<td>301.236</td>
<td>92.070000</td>
<td>9.686699</td>
</tr>
<tr>
<td>14</td>
<td>52</td>
<td>13953</td>
<td>300.369</td>
<td>81.780000</td>
<td>9.543450</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>15618</td>
<td>300.811</td>
<td>88.720000</td>
<td>9.656179</td>
</tr>
<tr>
<td>16</td>
<td>41</td>
<td>13255</td>
<td>299.995</td>
<td>1.250000</td>
<td>9.492130</td>
</tr>
<tr>
<td>17</td>
<td>33</td>
<td>15499</td>
<td>337.820</td>
<td>3.000000</td>
<td>9.648531</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>16550</td>
<td>355.670</td>
<td>2.000000</td>
<td>9.714141</td>
</tr>
<tr>
<td>19</td>
<td>39</td>
<td>14700</td>
<td>322.537</td>
<td>23.210000</td>
<td>9.595603</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>16950</td>
<td>359.410</td>
<td>21.020000</td>
<td>9.738023</td>
</tr>
<tr>
<td>21</td>
<td>40</td>
<td>13350</td>
<td>301.950</td>
<td>1.700000</td>
<td>9.499272</td>
</tr>
<tr>
<td>ActualDays</td>
<td>Estimated.Days</td>
<td>Depth.ft</td>
<td>Temp.BHT.F</td>
<td>inc</td>
<td>lndepth</td>
</tr>
<tr>
<td>------------</td>
<td>----------------</td>
<td>----------</td>
<td>------------</td>
<td>-----</td>
<td>---------</td>
</tr>
<tr>
<td>15</td>
<td>23</td>
<td>10260</td>
<td>249.420</td>
<td>24.600000</td>
<td>9.236008</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>14496</td>
<td>321.432</td>
<td>2.150000</td>
<td>9.581628</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>16025</td>
<td>347.425</td>
<td>2.000000</td>
<td>9.681905</td>
</tr>
<tr>
<td>4</td>
<td>54</td>
<td>19200</td>
<td>401.400</td>
<td>2.510000</td>
<td>9.862666</td>
</tr>
</tbody>
</table>

> drilling3 <- drilling2[which(ActualDays != 174),]

```r
> drilling3

  ActualDays Estimated.Days Depth.ft Temp.BHT.F inc lndepth
  1      15       23     10260   249.420    24.600000   9.236008
  2      32       26     14496   321.432    2.150000    9.581628
  3      32       17     16025   347.425    2.000000    9.681905
  4      54       23     19200   401.400    2.510000    9.862666
```
<table>
<thead>
<tr>
<th>#</th>
<th>5</th>
<th>64</th>
<th>36</th>
<th>14620</th>
<th>323.540</th>
<th>33.850000</th>
<th>9.590146</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>22</td>
<td>21</td>
<td>16350</td>
<td>352.950</td>
<td>2.500000</td>
<td>9.701983</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>21</td>
<td>18700</td>
<td>392.900</td>
<td>2.990000</td>
<td>9.836279</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>62</td>
<td>42</td>
<td>13470</td>
<td>303.990</td>
<td>21.370000</td>
<td>9.508220</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>28</td>
<td>24</td>
<td>15449</td>
<td>337.633</td>
<td>3.750000</td>
<td>9.686699</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>32</td>
<td>27</td>
<td>18100</td>
<td>382.700</td>
<td>7.000000</td>
<td>9.803667</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>17</td>
<td>17</td>
<td>10820</td>
<td>258.940</td>
<td>1.250000</td>
<td>9.289152</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>32</td>
<td>22</td>
<td>13460</td>
<td>301.066</td>
<td>76.800000</td>
<td>9.507478</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>11</td>
<td>16102</td>
<td>301.236</td>
<td>92.070000</td>
<td>9.686699</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>52</td>
<td>32</td>
<td>13953</td>
<td>299.970</td>
<td>81.780000</td>
<td>9.543450</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>27</td>
<td>10850</td>
<td>259.450</td>
<td>1.500000</td>
<td>9.291920</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>64</td>
<td>22</td>
<td>13350</td>
<td>301.950</td>
<td>1.700000</td>
<td>9.499272</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>33</td>
<td>24</td>
<td>15499</td>
<td>337.820</td>
<td>3.000000</td>
<td>9.648531</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>18</td>
<td>16550</td>
<td>355.670</td>
<td>2.000000</td>
<td>9.714141</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>39</td>
<td>26</td>
<td>14700</td>
<td>322.537</td>
<td>23.210000</td>
<td>9.595603</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>14</td>
<td>16950</td>
<td>359.410</td>
<td>5.250000</td>
<td>9.846917</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>40</td>
<td>42</td>
<td>13350</td>
<td>301.950</td>
<td>1.700000</td>
<td>9.499272</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>20</td>
<td>20</td>
<td>15573</td>
<td>339.741</td>
<td>76.800000</td>
<td>9.507478</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>34</td>
<td>20</td>
<td>18200</td>
<td>384.400</td>
<td>2.000000</td>
<td>9.809177</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>108</td>
<td>105</td>
<td>19500</td>
<td>406.500</td>
<td>2.184211</td>
<td>9.878170</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>47</td>
<td>42</td>
<td>14410</td>
<td>319.970</td>
<td>1.250000</td>
<td>9.575678</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>14</td>
<td>29</td>
<td>15939</td>
<td>345.963</td>
<td>2.000000</td>
<td>9.676524</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>40</td>
<td>29</td>
<td>18920</td>
<td>396.300</td>
<td>5.250000</td>
<td>9.846917</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>4</td>
<td>5</td>
<td>2885</td>
<td>124.045</td>
<td>1.750000</td>
<td>9.767280</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>11</td>
<td>42</td>
<td>10850</td>
<td>259.450</td>
<td>1.500000</td>
<td>9.291920</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>29</td>
<td>23</td>
<td>15750</td>
<td>342.750</td>
<td>2.000000</td>
<td>9.664596</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>46</td>
<td>37</td>
<td>19000</td>
<td>398.000</td>
<td>4.440000</td>
<td>9.852194</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>9</td>
<td>5</td>
<td>2900</td>
<td>124.300</td>
<td>0.000000</td>
<td>9.792466</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>57</td>
<td>42</td>
<td>14200</td>
<td>316.400</td>
<td>0.000000</td>
<td>9.560997</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>27</td>
<td>29</td>
<td>16077</td>
<td>348.309</td>
<td>3.780000</td>
<td>9.685145</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>64</td>
<td>29</td>
<td>19177</td>
<td>401.009</td>
<td>4.030000</td>
<td>9.861467</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>5</td>
<td>4</td>
<td>2900</td>
<td>124.300</td>
<td>0.500000</td>
<td>9.792466</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>15</td>
<td>16</td>
<td>11900</td>
<td>262.000</td>
<td>0.000000</td>
<td>9.305651</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>27</td>
<td>30</td>
<td>15750</td>
<td>342.750</td>
<td>0.000000</td>
<td>9.664596</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>6</td>
<td>5</td>
<td>2400</td>
<td>115.800</td>
<td>0.000000</td>
<td>9.783224</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>52</td>
<td>32</td>
<td>13361</td>
<td>300.250</td>
<td>43.180000</td>
<td>9.500095</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>20</td>
<td>11</td>
<td>15388</td>
<td>300.930</td>
<td>90.660000</td>
<td>9.641343</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>4</td>
<td>5</td>
<td>2450</td>
<td>116.650</td>
<td>0.530000</td>
<td>7.803843</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>35</td>
<td>36</td>
<td>13400</td>
<td>302.800</td>
<td>0.700000</td>
<td>9.503010</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>28</td>
<td>19</td>
<td>15000</td>
<td>330.000</td>
<td>4.840000</td>
<td>9.615805</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>16</td>
<td>17</td>
<td>16500</td>
<td>355.500</td>
<td>5.250000</td>
<td>9.711116</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>4</td>
<td>5</td>
<td>2850</td>
<td>123.450</td>
<td>2.810000</td>
<td>7.955074</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>16</td>
<td>10</td>
<td>8896</td>
<td>219.092</td>
<td>86.960000</td>
<td>9.093357</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>73</td>
<td>21</td>
<td>12306</td>
<td>219.092</td>
<td>90.460000</td>
<td>9.417842</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>3</td>
<td>800</td>
<td>88.600</td>
<td>1.250000</td>
<td>6.684612</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>7</td>
<td>17</td>
<td>7000</td>
<td>194.000</td>
<td>0.750000</td>
<td>8.853665</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>21</td>
<td>27</td>
<td>14100</td>
<td>314.700</td>
<td>1.000000</td>
<td>9.553930</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>38</td>
<td>20</td>
<td>15300</td>
<td>335.100</td>
<td>1.000000</td>
<td>9.635608</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>33</td>
<td>23</td>
<td>16300</td>
<td>352.100</td>
<td>4.800000</td>
<td>9.698920</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>5</td>
<td>5</td>
<td>2800</td>
<td>122.600</td>
<td>1.250000</td>
<td>7.937375</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>57</td>
<td>33</td>
<td>14321</td>
<td>318.457</td>
<td>1.000000</td>
<td>9.569482</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>14</td>
<td>27</td>
<td>15800</td>
<td>343.600</td>
<td>2.000000</td>
<td>9.667765</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>46</td>
<td>23</td>
<td>19200</td>
<td>401.400</td>
<td>8.000000</td>
<td>9.862666</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX H
A TEMPLATE OF R’s CODE FOR REGRESSION AND UNIVARIATE CALCULATIONS

```r
# Load data
drilling = read.csv(file.choose(), header=T)
attach(drilling)

# Regression models
Reg1 <- lm(ActualDays~Estimated.Days+Depth.ft,drilling)
full <- lm(ActualDays~-1,drilling)
Reg2 <- step(full)
null <- lm(ActualDays~1,drilling)
Reg3 <- step(null, scope=list(lower=null,upper=full,direction="forward"))
Reg4 <- step(full,direction="backward")

Reg5 <- step(full,scale=(summary(full)$sigma)^2)
Reg6 <- step(null,scope=list(lower=null,upper=full,direction="forward"),scale=(summary(full)$sigma)^2)
Reg7 <- step(full,direction="backward",scale=(summary(full)$sigma)^2)

# Plotting and analysis
par(mfrow=c(2,2))
plot.lm(Reg7)
plot(Reg7$res,ylab="Residual",xlab="Predicted value of ActualDays",main="Residual Plot: ActualDays = Depth_ft Estimated_Days")

a <- Reg7$res
u <- (1:length(a)-.5)/length(a)
Q <- quantile(a,u,type=5)
plot(qnorm(u),Q)
abline(lm(Q~qnorm(u)))

SWReg7 = shapiro.test(Reg7$res)
SWReg7
PTReg7 = pearson.test(Reg7$res)
PTReg7

summary(Reg7)

plot(Depth.ft,ActualDays)
lndepth <- log(Depth.ft)
logdepth <- log10(Depth.ft)
depthsqr <- -(Depth.ft)^2
```

drilling2 <- cbind(drilling, lndepth)
drilling2
drilling3 <- drilling2[which(ActualDays != 174),]
drilling3
VITA

Name: Jose Alejandro De Almeida

Email: jdealmeida@live.com

Education: B.S., Chemical Engineering, Colorado State University, 2005
M.S., Petroleum Engineering, Texas A&M University, 2010

Languages: English, Spanish, Portuguese

Work Experience: 3.5 years with M-I SWACO USA

Countries Lived: Brazil, Ecuador, France, Indonesia, Mexico, United Arab Emirates, United States of America.

Address: Department of Petroleum Engineering
        c/o Dr. F. E. Beck
        Texas A&M University
        College Station, TX, 77843-3116