**Regression Analysis**

Take Home Final Exam

Due Thursday April 23, 2020 by 11:59 PM

The following data consists (from Norman Draper and Harry Smith) of 25 observations recorded at a steam plant. The data set has ten variables defined as follows:

1. Pounds of Steam Used Monthly (Y) in coded form
2. Pounds of real fatty acid (X1)in storage per month
3. Pounds of crude glycerin made (X2)
4. Average wind velocity (mph) (X3)
5. Calendar days per month (X4)
6. Operating Days per month (X5)
7. Days below 32F (X6)
8. Average atmospheric Temperature (F) (X7)
9. Average wind velocity squared (X8)
10. Number of startups (X9)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Index | Y | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 |
| 1 | 10.98 | 5.20 | 0.61 | 7.4 | 31 | 20 | 22 | 35.3 | 54.8 | 4 |
| 2 | 11.1 | 5.12 | 0.64 | 8.0 | 29 | 20 | 25 | 29.7 | 64 | 5 |
| 3 | 12.51 | 6.19 | 0.78 | 7.4 | 31 | 23 | 17 | 30.8 | 54.8 | 4 |
| 4 | 8.40 | 3.89 | 0.49 | 7.5 | 30 | 20 | 22 | 58.8 | 56.3 | 4 |
| 5 | 9.27 | 6.28 | 0.84 | 5.5 | 31 | 21 | 0 | 61.4 | 30.3 | 5 |
| 6 | 8.73 | 5.76 | 0.74 | 8.9 | 30 | 22 | 0 | 71.3 | 79.2 | 4 |
| 7 | 6.36 | 3.45 | 0.42 | 4.1 | 31 | 11 | 0 | 74.4 | 16.8 | 2 |
| 8 | 8.50 | 6.57 | 0.87 | 4.1 | 31 | 23 | 0 | 76.7 | 16.8 | 5 |
| 9 | 7.82 | 5.69 | 0.75 | 4.1 | 30 | 21 | 0 | 70.7 | 16.8 | 4 |
| 10 | 9.14 | 6.14 | 0.76 | 4.5 | 31 | 20 | 0 | 57.5 | 20.3 | 5 |
| 11 | 8.24 | 4.84 | 0.65 | 10.3 | 30 | 20 | 11 | 46.4 | 106.1 | 4 |
| 12 | 12.19 | 4.88 | 0.62 | 6.9 | 31 | 21 | 12 | 28.9 | 47.6 | 4 |
| 13 | 11.88 | 6.03 | 0.79 | 6.6 | 31 | 21 | 25 | 28.1 | 43.6 | 5 |
| 14 | 9.57 | 4.55 | 0.60 | 7.3 | 28 | 19 | 18 | 39.1 | 53.3 | 5 |
| 15 | 10.94 | 5.71 | 0.70 | 8.1 | 31 | 23 | 5 | 46.8 | 65.6 | 4 |
| 16 | 9.58 | 5.67 | 0.74 | 8.4 | 30 | 20 | 7 | 48.5 | 70.6 | 4 |
| 17 | 10.09 | 6.72 | 0.85 | 6.1 | 31 | 22 | 0 | 59.3 | 37.2 | 6 |
| 18 | 8.11 | 4.95 | 0.67 | 4.9 | 30 | 22 | 0 | 70.0 | 24 | 4 |
| 19 | 6.83 | 4.62 | 0.45 | 4.6 | 31 | 11 | 0 | 70.0 | 21.2 | 3 |
| 20 | 8.88 | 6.60 | 0.95 | 3.7 | 31 | 23 | 0 | 74.5 | 13.7 | 4 |
| 21 | 7.68 | 5.01 | 0.64 | 4.7 | 30 | 20 | 0 | 72.1 | 22.1 | 4 |
| 22 | 8.47 | 5.68 | 0.75 | 5.3 | 31 | 21 | 1 | 58.1 | 28.1 | 6 |
| 23 | 8.86 | 5.28 | 0.70 | 6.2 | 30 | 20 | 14 | 44.6 | 38.4 | 4 |
| 24 | 10.36 | 5.36 | 0.67 | 6.8 | 31 | 20 | 22 | 33.4 | 46.2 | 4 |
| 25 | 11.08 | 5.87 | 0.70 | 7.5 | 31 | 22 | 28 | 28.6 | 56.3 | 5 |

Turn to the next page for the questions regarding the data:

1. Prepare box plots/stem and leaf for each of the variables. What information do the plots provide?
2. Plot a scatter plot of Y against each predictor variable. What do the plots tell you about the nature of the relationship between Y and each of the independent variables? Note that the scatter plot will be more informative than the matrix plot because of the number of data points and the variables.
3. Fit a linear regression model containing the all nine independent variables. Run all model diagnostics on this model (this will include VIF’s, partial regression plots, residual analysis and influence diagnostics)
4. Discuss the results of the model diagnostics and if any transformations are needed. Do the appropriate transformations if they are needed.
5. Use an appropriate model building strategy from Chapter 10 to edit the model with the transformed variables (or the original model if there was no transformation needed) and reduce the number of regressors