Macy Chen December 13th, 2024

College Sleep Habits and Their Impact on GPA

Introduction.

This project explores the relationship between various factors and academic performance, with a focus on students' Grade Point Average (GPA). The analysis is based on data from a comprehensive sleep study, which examines connections between sleep patterns, cognitive function, mental health, and lifestyle habits. By analyzing these factors, the project aims to identify key predictors of GPA and provide insights into how student well-being impacts academic success.

Data.

The dataset used in this analysis is the "SleepStudy" dataset, sourced from Lock5Stat. It contains 253 observations and 27 variables, capturing a wide range of information on students' sleep behaviors, cognitive test scores, mental health indicators, and lifestyle factors. This project specifically focuses on a subset of variables hypothesized to affect GPA, including measures of class attendance, cognitive skills, sleep quality, mental health (e.g., depression, anxiety, and stress), and alcohol consumption. By building and refining a predictive model, the goal is to uncover the most influential factors contributing to GPA outcomes.

Model Building Process.

A random subset of 200 cases was drawn from the dataset to facilitate analysis. Initial predictors of GPA considered included: Classes Missed, Cognition Z-score, Poor Sleep Quality, Depression Score, Anxiety Score, Stress Score, DASScore, Drinks, Average Sleep and Weekday Sleep.

Scatterplots were used to visualize relationships between variables and GPA (see appendix A). Looking at the plots, only predictors ‘drinks’ and ‘cognition z-score’ suggest a strong linear correlation with GPA. However, the graphs for ‘classes missed’, ‘stress score’ and ‘DAS Score’ suggest that a specific nonlinear relationship may exist. For ‘classes missed’, GPA appears to decrease at first and then levels out as classes missed increases. For ‘stress score’ and ‘DAS Score’, the opposite occurs, where GPA increases sharply initially and then levels out as stress and DAS score increases. These behaviours suggest that applying a logarithmic transformation might reveal a better linear relationship. Log(variable + 1) is taken to avoid a log of 0. This is proven to be true when comparing the results of their simple linear regression.

To evaluate the relationships between GPA and the predictor variables, simple linear regression models were performed (see appendix B). Predictors were analyzed both in their original and logarithmic-transformed forms, with R^2 values compared to assess model fit. The results demonstrate that transformations improved linear relationships for certain predictors. The following variables showed a strong association with GPA:

- Drinks: The regression analysis showed a significant negative relationship between Drinks and GPA ($\beta = -0.030$, $p = 0.0001 < 0.05$, $R^2 = 0.06$). The untransformed data showed a statistically significant association, and no transformation was required.
- Cognition Z-score: The regression analysis showed a significant positive relationship between Cognition Z-score and GPA ($\beta = 0.142$, $p = 0.0005 < 0.05$, $R^2 = 0.05$). The untransformed data showed a statistically significant association, and no transformation was required.
- Classes Missed: When untransformed, the negative relationship between Classes Missed and GPA had a weak fit ($\beta = -0.019$, $p = 0.035 < 0.05$, $R^2 = 0.017$). The model improved

significantly after applying a logarithmic transformation ($\beta = -0.097$, $p = 0.009 < 0.05$, $R^2 = 0.029$). The data showed a statistically significant association with GPA, therefore, log of 'classes missed' is taken and used in further analysis.

- Stress Score: When untransformed, the positive relationship between Classes Missed and GPA had a weak fit ($\beta = 0.0095$, $p = 0.0085 < 0.05$, $R^2 = 0.029$). The model improved significantly after applying a logarithmic transformation ($\beta = 0.098$, $p = 0.0016 < 0.05$, $R^2 = 0.044$). The data showed a statistically significant association with GPA, therefore, log of 'stress score' is taken and used in further analysis.

To check for multicollinearity between variables, a correlation matrix was constructed (see Appendix C). There is a strong correlation between stress, anxiety, depression and DAS scores, as well as between 'weekday sleep' and 'average sleep'. Therefore, considering stress score's significant association and weekday sleep's slightly better association with GPA, 'DAS score' and 'average sleep' were longer considered for the final model.

A full linear regression model was built, incorporating all significant variables. Based on the ANOVA table (see appendix D), 'classes missed', 'cognition z-score', 'stress score' and 'drinks' have a statistically significant effect on the variability in GPA (p-values < 0.05), after accounting for the other variables in the model. However, the summary table showed a low adjusted R^2 of 0.19 (see appendix D), meaning that only 19% of the variability in GPA can be explained by the significant predictors in the model, after accounting for the number of predictors and sample size. This low adjusted R^2 suggests that although these variables have a statistically significant effect, there are likely other important factors influencing GPA that are not included in the model. Therefore, interaction terms were explored to improve the model's explanatory power.

Stepwise regression, guided by AIC, was employed to optimize the predictive model for GPA. Initially, all potential predictors, including interaction terms, were considered. Insignificant predictors were iteratively removed, and interaction terms were systematically tested and added if they improved the model fit. For full steps and process, see separate R code document. After refining the model through these iterative steps, the final model selected was:

GPA ~ log(ClassesMissed+1) + CognitionZscore + AnxietyScore + log(StressScore+1) + Happiness + Drinks + log(StressScore+1)×Drinks + Happiness×Drinks + CognitionZscore×log(StressScore+1).

The regression analysis reveals several key predictors of GPA. Looking at the summary table (appendix E, Figure 1.0), 'Cognition Z-score' and 'stress score' have the most significant effect on GPA with p-values < 0.001, then 'anxiety score', 'happiness' and 'drinks' with p-values < 0.01, and 'classes missed' with p-value < 0.05. Interaction terms showed that the positive effect of drinking diminishes with higher stress and that happiness's impact on GPA is moderated by drinking. The interaction between 'cognition z-score' and 'stress score' suggests cognitive benefits are weaker under stress. The model's R^2 of 0.26 highlights the importance of both main effects and interactions in predicting GPA as it is an improvement from the R^2 of the initial model with no interaction terms.

This relationships between 'classes missed', 'cognition z-score', 'stress score' and 'drinks' with GPA, as well as interaction terms between stress score and drinks, happiness and drinks, and cognition z-score and drinks with GPA are statistically significant, with the ANOVA results showing high F-values with p-values < 0.05 (see appendix E, Figure 2.0), indicating that these variables explain a significant amount of the variance in the GPA.

Best Model.

Using forward regression and backward elimination, two possible best models were determined (see Appendix F).

From forward regression, the best model is (no change to the current model):

GPA ~ log(ClassesMissed+1) + CognitionZscore + AnxietyScore + log(StressScore+1) + Happiness + Drinks + log(StressScore+1)×Drinks + Happiness×Drinks + CognitionZscore×log(StressScore+1).

From backward elimination, the best model is:

GPA ~ log(ClassesMissed+1) + CognitionZscore + AnxietyScore + log(StressScore+1) + Happiness + Drinks + log(StressScore+1)×Drinks + Happiness×Drinks.

To determine whether the additional predictor variable CognitionZscore×log(StressScore+1) is significant, the hypothesis is established as follows:

- Null Hypothesis (H_0): $\beta_{\text{CognitionZscore:log(StressScore + 1)}} = 0$
- Alternative Hypothesis (H_a): $\beta_{\text{CognitionZscore:log(StressScore + 1)}} \neq 0$

Performing the F-test using ANOVA to compare the two models, if the p-value is less than $\alpha = 0.05$, we reject the null hypothesis.

From the ANOVA table (see Appendix G), the p-value (0.03773) is less than 0.05, the null hypothesis is rejected. This means that the full model, including the interaction term CognitionZscore×log(StressScore+1), provides a significantly better fit to the data. So, based on the F-test, the interaction term indeed improves the regression model, suggesting it should be kept for a more accurate prediction of GPA.

Verification of Assumptions.

- Linearity: The residuals vs. fitted values plot show a random scatter around 0, with no curve clear patterns such as curves or clusters (appendix H, Figure 1.0). This confirms a linear relationship between predictors and GPA.
- Homoscedasticity: There is no increase or decrease of the spread of residuals in the residuals vs. fitted values plot. The residuals displayed constant variance across fitted values, showing no indication of the assumption being violated.
- Normality of Residuals: The normal Q-Q plot (appendix H, Figure 2.0) and histogram of residuals (appendix H, Figure 3.0) confirm that the residuals are approximately normally distributed, with only a few outliers at the extremes, notably at the lower end.

These diagnostics confirm that the assumptions of linear regression are largely met, reinforcing the validity of the model. The boxcox plot (Appendix H, Figure 4.0) shows that no transformation of the response variable is necessary.

Conclusion.

In this study, multiple linear regression models were developed to identify the predictors of GPA and explore their relationships. By incorporating transformations and interaction terms, the final model identified significant variables, Classes Missed, Cognition Z-Score, Anxiety Score, Stress Score, Happiness, and Drinks. Interaction effects between variables, such as Stress Score and Drinks, Cognition Z-Score and Stress Score, and Happiness and Drinks, further demonstrated how these factors interact to influence GPA.

Model diagnostics confirmed the validity of assumptions regarding linearity, homoscedasticity, and normality of residuals. Despite an improved R^2 value of 0.26, the model accounts for only a portion of the variability in GPA, highlighting the complexity of academic

performance and the potential influence of unmeasured factors. The inclusion of interaction terms, validated through ANOVA comparisons, underscores their importance in enhancing the explanatory power of the model.

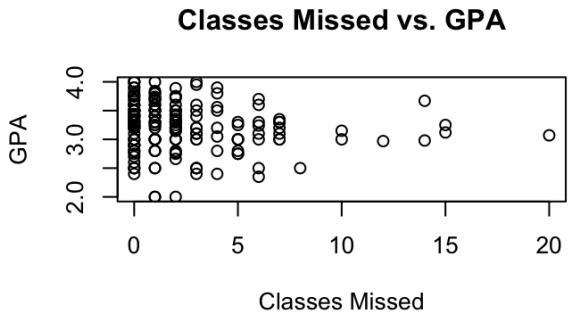
Future research could explore additional predictors and consider nonlinear modeling approaches to more fully capture the intricate dynamics underlying GPA performance.

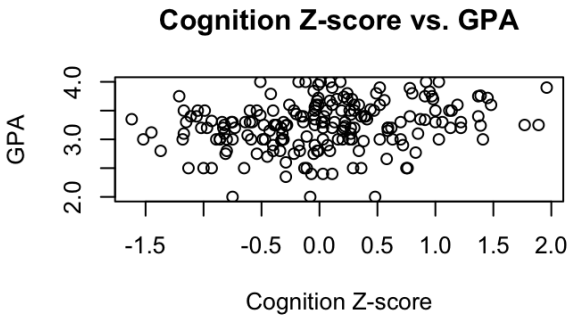
Citations.

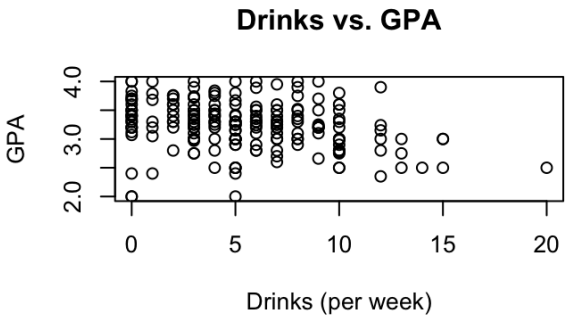
Lock5. (2023). *Dataset documentation for Statistics: Unlocking the power of data, 3rd edition*. Lock5. https://www.lock5stat.com/datasets3e/Lock5DataGuide3e.pdf.

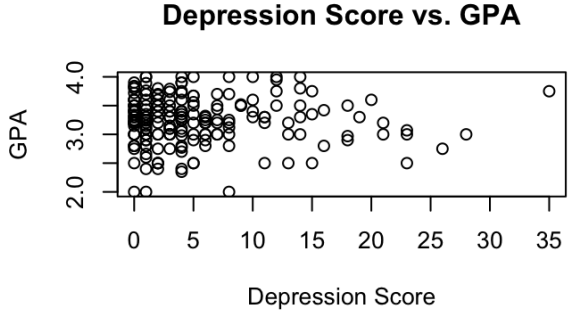
Onyper, S. V., Thacher, P. V., Gilbert, J. W., & Gradess, S. G. (2012). Class start times, sleep, and academic performance in college: a path analysis. *Chronobiology international*, *29*(3), 318–335. https://doi.org/10.3109/07420528.2012.655868.

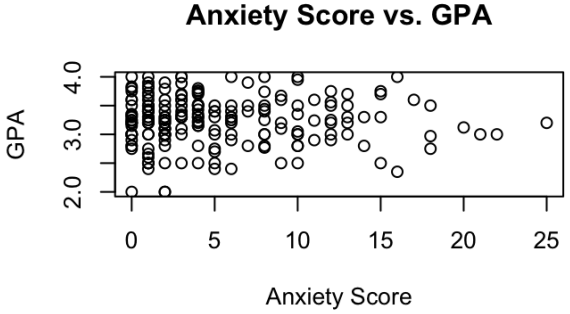
Appendix A - Scatterplots.

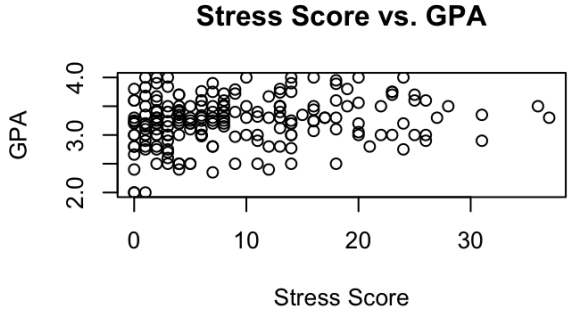


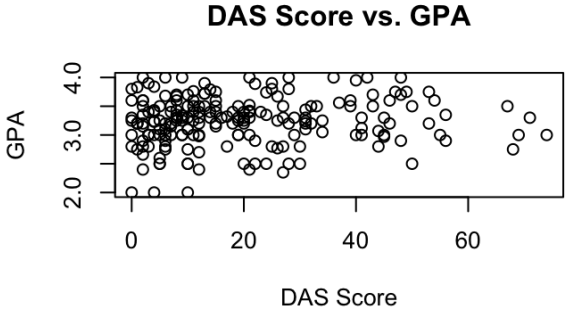




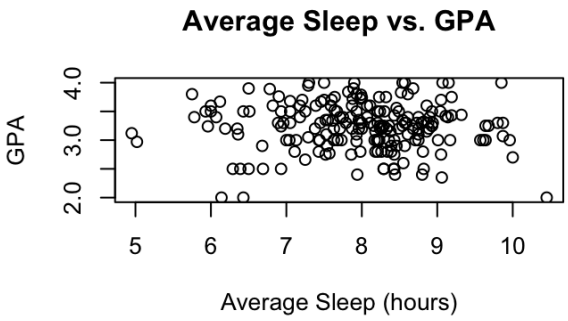


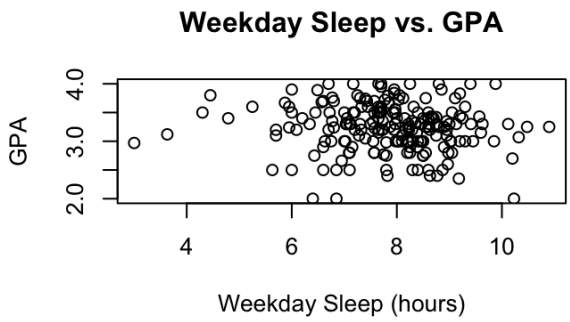







Poor Sleep Quality vs. GPA
GPA
2.0
3.0
4.0
5
10
15
Poor Sleep Quality


Average Sleep vs. GPA
GPA
2.0
3.0
4.0
5
6
7
8
9
10
Average Sleep (hours)


Weekday Sleep vs. GPA
GPA
2.0
3.0
4.0
4
6
8
10
Weekday Sleep (hours)

Appendix B - Simple Linear Regression Tables.

```
Call:
lm(formula = GPA ~ log(StressScore + 1), data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.11665 -0.24879  0.02687  0.24465  0.88335

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          3.04879    0.06632  45.973   <2e-16 ***
log(StressScore + 1)  0.09790    0.03060   3.199   0.0016 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4064 on 198 degrees of freedom
Multiple R-squared:  0.04915,   Adjusted R-squared:  0.04435
F-statistic: 10.24 on 1 and 198 DF,  p-value: 0.001605
```

```
Call:
lm(formula = GPA ~ StressScore, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.16203 -0.23750  0.03239  0.25411  0.83797

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.15255    0.04385  71.891  < 2e-16 ***
StressScore  0.00948    0.00357   2.656  0.00856 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4095 on 198 degrees of freedom
Multiple R-squared:  0.03439,   Adjusted R-squared:  0.02951
F-statistic: 7.052 on 1 and 198 DF,  p-value: 0.008563
```

```
Call:
lm(formula = GPA ~ DASScore, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.21955 -0.23835  0.03276  0.27657  0.79715

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.198668   0.045221  70.734   <2e-16 ***
DASScore    0.002089   0.001738   1.202    0.231
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4152 on 198 degrees of freedom
Multiple R-squared:  0.007242,  Adjusted R-squared:  0.002228
F-statistic: 1.444 on 1 and 198 DF,  p-value: 0.2309
```

```
Call:
lm(formula = GPA ~ log(DASScore + 1), data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.22974 -0.25666  0.03253  0.26023  0.82876

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        3.12177    0.08138  38.361   <2e-16 ***
log(DASScore + 1)  0.04503    0.02892   1.557    0.121
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4142 on 198 degrees of freedom
Multiple R-squared:  0.0121,    Adjusted R-squared:  0.007108
F-statistic: 2.425 on 1 and 198 DF,  p-value: 0.121
```

```
Call:
lm(formula = GPA ~ ClassesMissed, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.26273 -0.24305  0.02743  0.25695  0.77663

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.28241    0.03533  92.909   <2e-16 ***
ClassesMissed -0.01968    0.00927  -2.123    0.035 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4121 on 198 degrees of freedom
Multiple R-squared:  0.02225,   Adjusted R-squared:  0.01732
F-statistic: 4.507 on 1 and 198 DF,  p-value: 0.035
```

```
Call:
lm(formula = GPA ~ log(ClassesMissed + 1), data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.25104 -0.21335  0.02672  0.25396  0.81620

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           3.31828    0.04157  79.824  < 2e-16 ***
log(ClassesMissed + 1) -0.09700    0.03695  -2.626  0.00933 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4097 on 198 degrees of freedom
Multiple R-squared:  0.03365,   Adjusted R-squared:  0.02877
F-statistic: 6.894 on 1 and 198 DF,  p-value: 0.009325
```

```
Call:
lm(formula = GPA ~ CognitionZscore, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.30292 -0.24321  0.05479  0.29270  0.83789

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     3.23465    0.02863 112.980  < 2e-16 ***
CognitionZscore  0.14224    0.04045   3.516 0.000542 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4043 on 198 degrees of freedom
Multiple R-squared:  0.05878,   Adjusted R-squared:  0.05402
F-statistic: 12.36 on 1 and 198 DF,  p-value: 0.0005425
```

```
Call:
lm(formula = GPA ~ PoorSleepQuality, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.25146 -0.24111  0.01406  0.26130  0.78477

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.272166   0.071067  46.044   <2e-16 ***
PoorSleepQuality -0.005176   0.010407  -0.497     0.62
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4165 on 198 degrees of freedom
Multiple R-squared:  0.001248,  Adjusted R-squared:  -0.003797
F-statistic: 0.2473 on 1 and 198 DF,  p-value: 0.6195
```

```
Call:
lm(formula = GPA ~ DepressionScore, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.24093 -0.24041  0.01036  0.26036  0.76148

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.2409266  0.0392035  82.669   <2e-16 ***
DepressionScore -0.0001716  0.0047880  -0.036    0.971
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4168 on 198 degrees of freedom
Multiple R-squared:  6.487e-06, Adjusted R-squared:  -0.005044
F-statistic: 0.001284 on 1 and 198 DF,  p-value: 0.9714
```

```
Call:
lm(formula = GPA ~ AnxietyScore, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.24407 -0.24250  0.01381  0.26164  0.76854

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.2440729  0.0416420  77.904   <2e-16 ***
AnxietyScore -0.0007886  0.0056966  -0.138     0.89
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4167 on 198 degrees of freedom
Multiple R-squared:  9.677e-05, Adjusted R-squared:  -0.004953
F-statistic: 0.01916 on 1 and 198 DF,  p-value: 0.89
```

```
Call:
lm(formula = GPA ~ Happiness, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.23370 -0.24352  0.02952  0.25943  0.78201

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.188528   0.144860  22.011   <2e-16 ***
Happiness   0.001964   0.005411   0.363    0.717
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4166 on 198 degrees of freedom
Multiple R-squared:  0.0006647, Adjusted R-squared:  -0.004382
F-statistic: 0.1317 on 1 and 198 DF,  p-value: 0.7171
```

```
Call:
lm(formula = GPA ~ Drinks, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.40076 -0.25287  0.01755  0.25839  0.86545

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.400765   0.049651  68.493  < 2e-16 ***
Drinks      -0.029580   0.007497  -3.946  0.00011 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4013 on 198 degrees of freedom
Multiple R-squared:  0.0729,    Adjusted R-squared:  0.06821
F-statistic: 15.57 on 1 and 198 DF,  p-value: 0.0001105
```

```
Call:
lm(formula = GPA ~ AverageSleep, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.26905 -0.23662  0.03657  0.26378  0.78958

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.36609    0.24213  13.902   <2e-16 ***
AverageSleep -0.01580    0.03012  -0.525      0.6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4165 on 198 degrees of freedom
Multiple R-squared:  0.001388,  Adjusted R-squared:  -0.003655
F-statistic: 0.2752 on 1 and 198 DF,  p-value: 0.6004
```

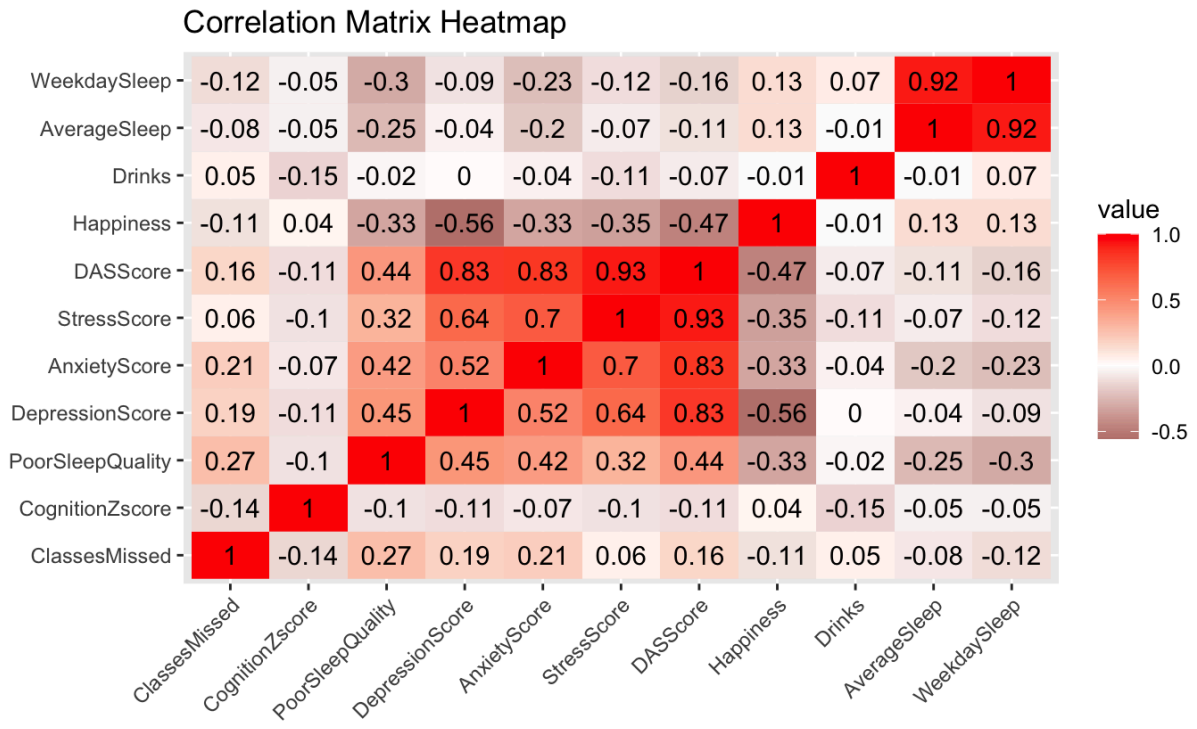
```
Call:
lm(formula = GPA ~ WeekdaySleep, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.27897 -0.23325  0.03827  0.25767  0.81494

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.45168    0.19559  17.647   <2e-16 ***
WeekdaySleep -0.02699    0.02465  -1.095    0.275
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4155 on 198 degrees of freedom
Multiple R-squared:  0.006016,  Adjusted R-squared:  0.0009955
F-statistic: 1.198 on 1 and 198 DF,  p-value: 0.275
```

Appendix C - Correlation Matrix.



Appendix D - Regression Tables for Initial Model.

```
Call:
lm(formula = GPA ~ log(ClassesMissed + 1) + CognitionZscore +
    PoorSleepQuality + DepressionScore + AnxietyScore + log(StressScore +
    1) + Happiness + Drinks + WeekdaySleep, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.08827 -0.20209  0.01601  0.22974  0.89046

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            3.3013646  0.2789578  11.835  < 2e-16 ***
log(ClassesMissed + 1) -0.0663142  0.0357058  -1.857 0.064825 .
CognitionZscore         0.1306336  0.0383707   3.405 0.000809 ***
PoorSleepQuality       -0.0002043  0.0114181  -0.018 0.985743
DepressionScore        -0.0057431  0.0062802  -0.914 0.361623
AnxietyScore           -0.0179742  0.0072304  -2.486 0.013785 *
log(StressScore + 1)    0.1922596  0.0406912   4.725 4.47e-06 ***
Happiness               0.0025530  0.0058914   0.433 0.665259
Drinks                 -0.0192095  0.0072000  -2.668 0.008290 **
WeekdaySleep           -0.0287684  0.0237418  -1.212 0.227125
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3732 on 190 degrees of freedom
Multiple R-squared:  0.2306,    Adjusted R-squared:  0.1942
F-statistic: 6.327 on 9 and 190 DF,  p-value: 7.953e-08
```

Figure 1.0: Summary table

```
Analysis of Variance Table

Response: GPA
                        Df  Sum Sq Mean Sq F value    Pr(>F)
log(ClassesMissed + 1)   1  1.1571  1.1571  8.3086  0.004400 **
CognitionZscore          1  1.7441  1.7441 12.5242  0.000505 ***
PoorSleepQuality         1  0.0370  0.0370  0.2658  0.606731
DepressionScore          1  0.0468  0.0468  0.3362  0.562711
AnxietyScore             1  0.0056  0.0056  0.0405  0.840786
log(StressScore + 1)     1  3.6621  3.6621 26.2969 7.184e-07 ***
Happiness                1  0.0225  0.0225  0.1614  0.688325
Drinks                   1  1.0505  1.0505  7.5432  0.006603 **
WeekdaySleep             1  0.2045  0.2045  1.4683  0.227125
Residuals              190 26.4593  0.1393
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2.0: ANOVA table

Appendix E - Regression Tables for Improved Model.

```
Call:
lm(formula = GPA ~ log(ClassesMissed + 1) + CognitionZscore +
    AnxietyScore + log(StressScore + 1) + Happiness + Drinks +
    log(StressScore + 1) * Drinks + Happiness * Drinks + CognitionZscore *
    log(StressScore + 1), data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-1.07830 -0.22166 -0.00135  0.22343  0.88019

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                         2.251254   0.272064   8.275 2.22e-14 ***
log(ClassesMissed + 1)             -0.076682   0.033573  -2.284 0.023477 *
CognitionZscore                     0.293067   0.083936   3.492 0.000597 ***
AnxietyScore                       -0.020574   0.006646  -3.096 0.002261 **
log(StressScore + 1)                0.352816   0.058539   6.027 8.51e-09 ***
Happiness                           0.021890   0.008364   2.617 0.009583 **
Drinks                              0.129115   0.041826   3.087 0.002325 **
log(StressScore + 1):Drinks        -0.031139   0.008372  -3.720 0.000263 ***
Happiness:Drinks                   -0.003523   0.001238  -2.846 0.004916 **
CognitionZscore:log(StressScore + 1) -0.080133   0.038297  -2.092 0.037730 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3577 on 190 degrees of freedom
Multiple R-squared:  0.2929,    Adjusted R-squared:  0.2594
F-statistic: 8.745 on 9 and 190 DF,  p-value: 5.813e-11
```

Figure 1.0: Summary table

```
Analysis of Variance Table

Response: GPA
                                   Df  Sum Sq Mean Sq F value    Pr(>F)
log(ClassesMissed + 1)              1  1.1571  1.1571  9.0408 0.0029971 **
CognitionZscore                     1  1.7441  1.7441 13.6279 0.0002911 ***
AnxietyScore                        1  0.0479  0.0479  0.3743 0.5414138
log(StressScore + 1)                1  3.3890  3.3890 26.4803 6.609e-07 ***
Happiness                           1  0.1318  0.1318  1.0301 0.3114181
Drinks                              1  1.1009  1.1009  8.6024 0.0037702 **
log(StressScore + 1):Drinks         1  0.7646  0.7646  5.9743 0.0154284 *
Happiness:Drinks                    1  1.1773  1.1773  9.1993 0.0027596 **
CognitionZscore:log(StressScore + 1)  1  0.5603  0.5603  4.3782 0.0377305 *
Residuals                         190 24.3163  0.1280
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2.0: ANOVA table

Appendix F - Forward Regression and Backward Elimination.

```
Subset selection object
Call: regsubsets.formula(GPA ~ log(ClassesMissed + 1) + CognitionZscore +
    AnxietyScore + log(StressScore + 1) + Happiness + Drinks +
    log(StressScore + 1) * Drinks + Happiness * Drinks + CognitionZscore *
    log(StressScore + 1), data = dataset, method = "forward")
9 Variables  (and intercept)
                                   Forced in Forced out
log(ClassesMissed + 1)                 FALSE      FALSE
CognitionZscore                        FALSE      FALSE
AnxietyScore                           FALSE      FALSE
log(StressScore + 1)                   FALSE      FALSE
Happiness                              FALSE      FALSE
Drinks                                 FALSE      FALSE
log(StressScore + 1):Drinks            FALSE      FALSE
Happiness:Drinks                       FALSE      FALSE
CognitionZscore:log(StressScore + 1)   FALSE      FALSE
1 subsets of each size up to 8
Selection Algorithm: forward
         log(ClassesMissed + 1) CognitionZscore AnxietyScore log(StressScore + 1) Happiness
1  ( 1 ) " "                    " "             " "          " "                   " "
2  ( 1 ) " "                    "*"             " "          " "                   " "
3  ( 1 ) " "                    "*"             " "          "*"                   " "
4  ( 1 ) " "                    "*"             "*"          "*"                   " "
5  ( 1 ) " "                    "*"             "*"          "*"                   " "
6  ( 1 ) " "                    "*"             "*"          "*"                   " "
7  ( 1 ) "*"                    "*"             "*"          "*"                   " "
8  ( 1 ) "*"                    "*"             "*"          "*"                   " "
         Drinks log(StressScore + 1):Drinks Happiness:Drinks
1  ( 1 ) "*"    " "                        " "
2  ( 1 ) "*"    " "                        " "
3  ( 1 ) "*"    " "                        " "
4  ( 1 ) "*"    " "                        " "
5  ( 1 ) "*"    "*"                        " "
6  ( 1 ) "*"    "*"                        " "
7  ( 1 ) "*"    "*"                        " "
8  ( 1 ) "*"    "*"                        "*"
         CognitionZscore:log(StressScore + 1)
1  ( 1 ) " "
2  ( 1 ) " "
3  ( 1 ) " "
4  ( 1 ) " "
5  ( 1 ) " "
6  ( 1 ) "*"
7  ( 1 ) "*"
8  ( 1 ) "*"
```

Figure 1.0: Forward Regression

```
Subset selection object
Call: regsubsets.formula(GPA ~ log(ClassesMissed + 1) + CognitionZscore +
    AnxietyScore + log(StressScore + 1) + Happiness + Drinks +
    log(StressScore + 1) * Drinks + Happiness * Drinks + CognitionZscore *
    log(StressScore + 1), data = dataset, method = "backward")
9 Variables  (and intercept)
                                  Forced in Forced out
log(ClassesMissed + 1)                FALSE      FALSE
CognitionZscore                       FALSE      FALSE
AnxietyScore                          FALSE      FALSE
log(StressScore + 1)                  FALSE      FALSE
Happiness                             FALSE      FALSE
Drinks                                FALSE      FALSE
log(StressScore + 1):Drinks           FALSE      FALSE
Happiness:Drinks                      FALSE      FALSE
CognitionZscore:log(StressScore + 1)  FALSE      FALSE
1 subsets of each size up to 8
Selection Algorithm: backward
         log(ClassesMissed + 1) CognitionZscore AnxietyScore log(StressScore + 1) Happiness
1  ( 1 ) " "                    " "             " "          "*"                   " "
2  ( 1 ) " "                    " "             " "          "*"                   " "
3  ( 1 ) " "                    "*"             " "          "*"                   " "
4  ( 1 ) " "                    "*"             "*"          "*"                   " "
5  ( 1 ) " "                    "*"             "*"          "*"                   " "
6  ( 1 ) " "                    "*"             "*"          "*"                   " "
7  ( 1 ) " "                    "*"             "*"          "*"                   "*"
8  ( 1 ) "*"                    "*"             "*"          "*"                   "*"
         Drinks log(StressScore + 1):Drinks Happiness:Drinks
1  ( 1 ) " "    " "                        " "
2  ( 1 ) " "    "*"                        " "
3  ( 1 ) " "    "*"                        " "
4  ( 1 ) " "    "*"                        " "
5  ( 1 ) "*"    "*"                        " "
6  ( 1 ) "*"    "*"                        "*"
7  ( 1 ) "*"    "*"                        "*"
8  ( 1 ) "*"    "*"                        "*"
         CognitionZscore:log(StressScore + 1)
1  ( 1 ) " "
2  ( 1 ) " "
3  ( 1 ) " "
4  ( 1 ) " "
5  ( 1 ) " "
6  ( 1 ) " "
7  ( 1 ) " "
8  ( 1 ) " "
```

Figure 2.0: Backward Elimination

Appendix G - ANOVA Comparison Table.

```
Analysis of Variance Table

Model 1: GPA ~ log(ClassesMissed + 1) + CognitionZscore + AnxietyScore +
    log(StressScore + 1) + Happiness + Drinks + log(StressScore +
    1) * Drinks + Happiness * Drinks
Model 2: GPA ~ log(ClassesMissed + 1) + CognitionZscore + AnxietyScore +
    log(StressScore + 1) + Drinks + log(StressScore + 1) * Drinks +
    Happiness * Drinks + CognitionZscore * log(StressScore +
    1)
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1    191 24.877
2    190 24.316  1   0.56032 4.3782 0.03773 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix H - Residuals vs. Fitted Values Plots, Q-Q Plot, Boxcox Plot.

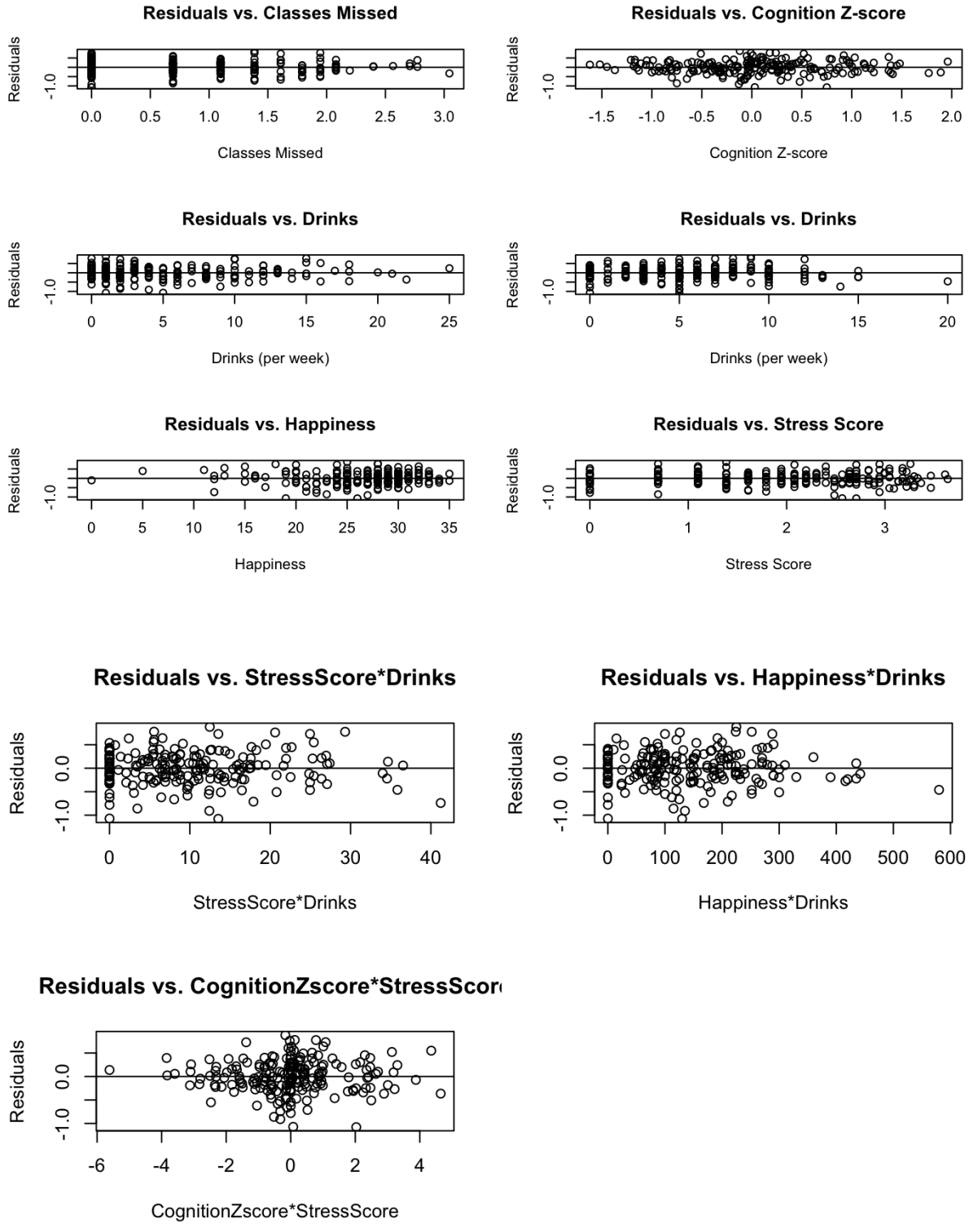


Figure 1.0: Residuals vs. individual predictors plots

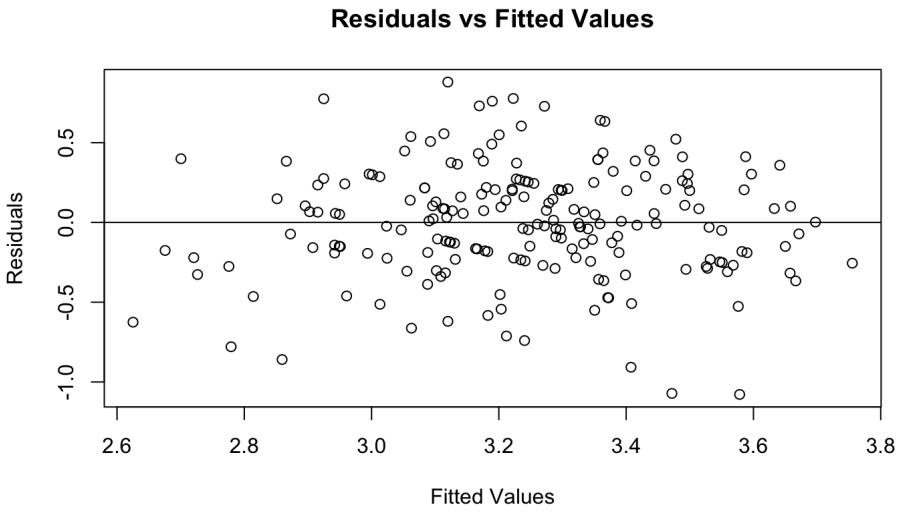


Figure 1.0: Residuals vs. fitted values plot

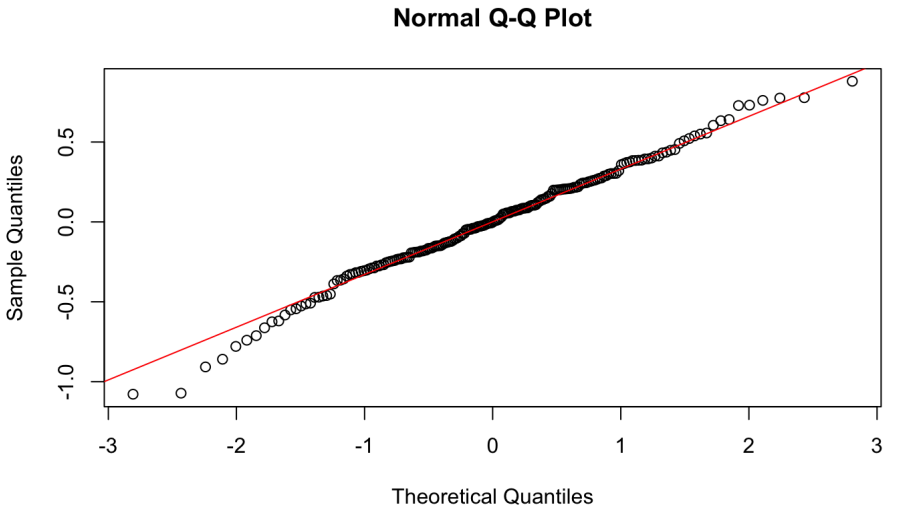


Figure 2.0: Normal Q-Q plot

Figure 3.0: Histogram of residuals plot

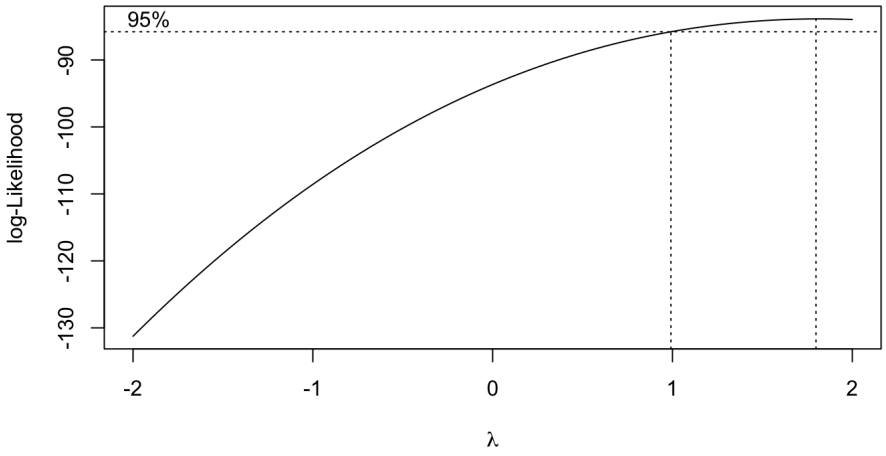


Figure 4.0: Boxcox plot