Depression's Influence on Academic Performance
By Harley Clifton, Sara Bancale, and Matt McMahon

## I. Background

A common mental health issue among students is depression which may be influenced by many extraneous factors (e.g. comparison of academic achievements to peers, relationship issues, stressful life events, etc.). Many students who struggle with depression experience decreased energy, lack of motivation, and often end up struggling with school work and due dates (Mayo Clinic Health System, 2021).

The PHQ-9 form, where patients suspected of depression self-report their symptoms and severity of symptoms, consists of 9 questions, each scored 0 to 3, for a total severity score between 0 and 27. Scores for the PHQ-9 form of 5, 10, 15, and 20 represent the minimum scores that qualify as someone having mild, moderate, moderately severe, and severe symptoms of depression respectively (Eack, S. M., Greeno, C. G., & Lee, B.-J., 2006, Form 1). There is potential for response bias since patients may underreport the severity of symptoms in fear of consequences of being completely transparent with their health care provider. However, since depression symptoms are not visual, self-reporting is the only way to detect them.

We are curious whether the severity of depression is responsible for a change in students' academic performance, or if other confounding variables contribute to a change in performance.

## II. Data Set Description

The raw data set consists of 18 different variables. Sex is an indicator variable with two levels: female, and male. Age is a categorical indicator with three levels, "18 years or less", "19 to 24 years", and "25 years and above". Educational Level specifies the degree each subject is pursuing, with three levels, "High School", "College - Bachelors", and "Masters". Variables four

through twelve are the individual questions on the PHQ-9 form (Form 1). In this dataset, subjects ranked each of these questions on a scale of 1-4, 4 being the highest, as opposed to the normal 0-3 ranking scale. The Job status variable has three levels: "None", "Part-time", and "Full-time". Living Situation indicates where the student lived during the school term with three levels; "Home (with parents)", "Private rented accommodation", and "University hall or Residence". Study time measures the amount of time the student spends studying each day with three levels: "1-2 hours", "3-4 hours", and "More than 4 hours". SocialMedia measures the number of hours spent on social media daily. Its three levels are "1-2 hours", "2-4 hours", and "4 or more hours". Next is the students' GPA from the previous semester, measured quantitatively. The Electronics variable measures the number of electronics present in a student's living space - with three levels, "1-3", "4-6", and "More than 6" - but was not of particular interest to us.

We created and modified additional variables for our research purposes. 'DepressionScore' was created by summing all responses for each individual on the PHQ-9 questionnaire. The 'DepressionScore' variable was adjusted to account for the difference between the total points possible on our questionnaire and the official scale by subtracting one point for every question in the PHQ-9 (total of 9 points) from the overall score for each student. Subsequently, a categorical variable called 'DepressionLevel' was generated to provide a more descriptive assessment of the 'DepressionScore' variable. It was created according to the PHQ-9 guidelines of a Depression score of 0-4 as "Normal", 5-9 as "Mild", 10-14 as "Moderate", 15-19 as "Moderately Severe", and 20+ as "Severe". Additionally, 'Standing' was created by mutating the GPA variable; a GPA $< 2$ indicates the student qualifies for Academic Probation and a GPA $\geq 2$ indicates a student is in Good Standing. Lastly, a Binary version of the Standing variable was created and called 'Stand.Bin', with a 0 representing Good Standing and a 1 representing

Probation - which will be considered a ‘success’ in our analysis. The Stand.Bin variable will be our response since it allows us to measure the estimated probability of Academic Probation.

## III. Scientific Goals and Primary Questions of Interest

We seek to explore the relationship between depression severity, measured via an individual’s adjusted cumulative score on the PHQ-9 questionnaire, and academic performance. We aim to determine whether there is a measurable impact on overall passing grades. A GPA of 2.0 is the cutoff that most higher education institutions use to dictate whether a student is put on Academic Probation. If a student is placed on Academic Probation, typically a faculty member(s) will check in with the student and provide opportunities for additional support if needed. If a student fails to improve their performance over the Academic Probation period, they will face Academic Dismissal.

As students ourselves, we know that numerous factors may influence an individual’s academic performance. Therefore, we want to determine if Depression Index is a good predictor of Academic Standing, in addition to exploring potential confounding factors. Specifically, we are curious to discover how Sex, Age, Job Status, Education Level, Living Situation, daily Study Time, and daily Social Media use relate to our two main variables of interest.

## IV. Preliminary Data Exploration

The relationships between variables of interest were visualized in R via a Scatterplot and Correlation matrix from the psych package (Revelle, 2021) (Figure 1). Sex and daily hours spent on Social Media were correlated, although weekly ($r = -0.22$), as well as Job status and Age ($r = -0.22$). A barplot was created using the ggplot2 package to visualize the counts of students with

Good Standing and students on Academic Probation (Wickham, 2016) (Figure 2). Upon further examination, there are a total of 324 students that have Good Standing and 23 who qualify for Academic Probation in our sample. We suspect the disparity in number of subjects in each category may be attributed to the time-sensitivity of the Academic Dismissal Decision Process.

Additional plots were generated to inquire about other intriguing variables and relationships; these include a histogram showing the distribution of depression scores (Figure 3), a scatterplot displaying the relationship between depression scores and Academic Standing (Figure 4), and Boxplots comparing the same variables (Figure 5) (R Core Team, 2021).

## V. Analysis Plan and Modeling

Since Academic Standing is a Binary variable, modeling this data with Logistic Regression is appropriate. Initially, a full additive model was fit with Academic Standing as the response, and included Depression Score, Sex, Age, Job Status, Education Level, Living Situation, daily Study Time, and daily Social Media use as predictors. The summary of this model revealed large p-values for all of our predictors, indicating we have too many in our model (Table 1).

Model selection was conducted using the step function in R (R Core Team, 2021). The results suggested that the model with the lowest AIC had only Sex and Age as predictors. However, since our main research question is to see how Depression Score influences a student's Academic Standing, we need to include that variable in the model. Effects plots generated with the effects package and provided further evidence that there is no interaction between these three predictors (Fox and Weisberg, 2019) (Figures 6 & 7). Additional forward stepwise regression deemed all possible combinations of 2-way interactions between these variables to be unimportant. Therefore, all interaction terms were dropped from the final model.

The Final Estimated Model (Model 1) is as follows:

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = -19.66497 + 0.02288x_D + 17.05319 * I_{Age=19to24} + 16.85922 * I_{Age=25above} - 16.93007 * I_{Sex=Male}$$

Where $\hat{\pi}$ is the probability that an individual in our sample will be placed on academic probation, $x_D$ is the cumulative result of the PHQ-9 questionnaire (recorded as an individual's Depression Score) in points, $I_{Age=19to24}$ is an indicator variable that is 1 when the individual is between 19 and 24 years old and 0 otherwise, $I_{Age=25above}$ is an indicator variable that is 1 when the individual is 25 years old or more and 0 otherwise, and $I_{Sex=Male}$ is an indicator variable that is 1 when the individual's Sex is Male and 0 otherwise. The standard suite of diagnostic plots was created in R, and since we are dealing with a Binary response, we expect most of our assumptions to be violated (R Core Team) (Figure 8).

# Appendix

## References

H. Wickham. The ggplot2 Package: Elegant Graphics for Data Analysis (2016). Springer-Verlag New York. https://ggplot2.tidyverse.org.

Introduction - ons.org. (n.d.). Introduction In 2013, the Oncology Nursing Society (ONS) published Statement on the Scope and Standards of Oncology Nursing Practice: Generalist and Advanced Practice. https://www.ons.org/sites/default/files//publication_pdfs/0657_Sample.pdf

J. Fox and S. Weisberg. The effects Package: Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals (2018). Journal of Statistical Software, 87(9):1-27. https://doi.org/10.18637/jss.v087.i09.

R Core Team. R: A language and environment for statistical computing (2021). R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

W. Revelle. The psych Package: Procedures for Personality and Psychological Research (2021). Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych.

Mayo Clinic Health System. (2021). College students and Depression. Mayo Clinic Health System. Retrieved April 8, 2022. https://www.mayoclinichealthsystem.org/hometown-health/speaking-of-health/college-students-and-depression.

Eack, S. M., Greeno, C. G., & Lee, B.-J. (2006). Limitations of the patient health questionnaire in identifying anxiety and depression: Many cases are undetected. Research on social work practice. Retrieved April 8, 2022. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3899353/.

## Figures

Figure 1: Scatterplot and Correlation Matrix for a Subset of Interesting Variables.

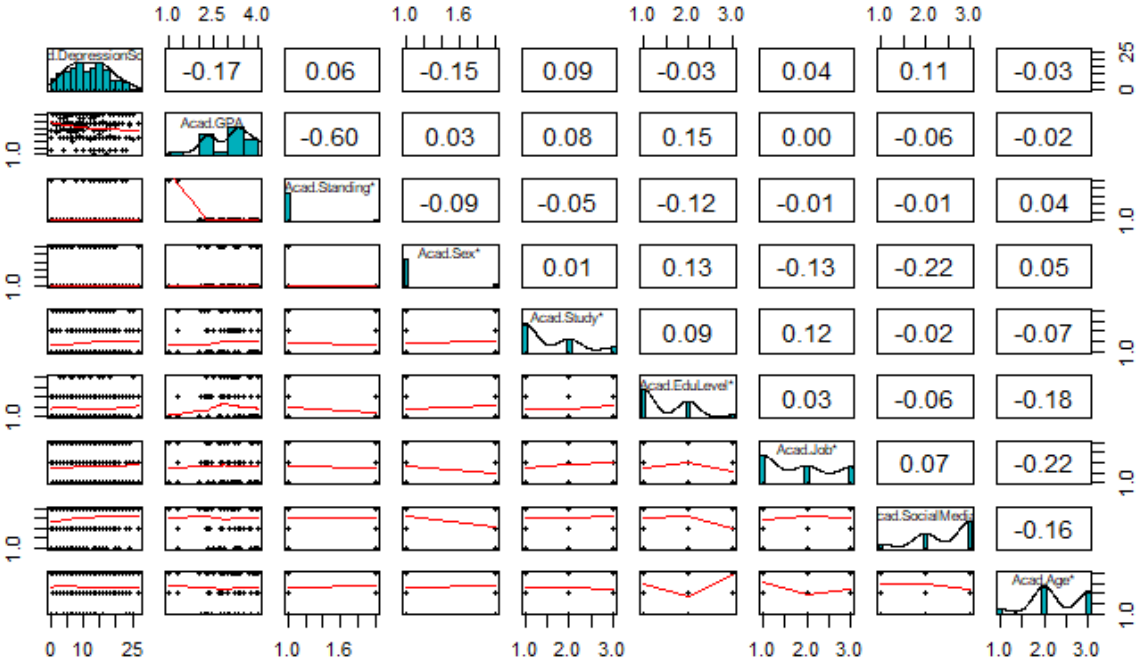



Figure 2: Bar Plot of Academic Standing.

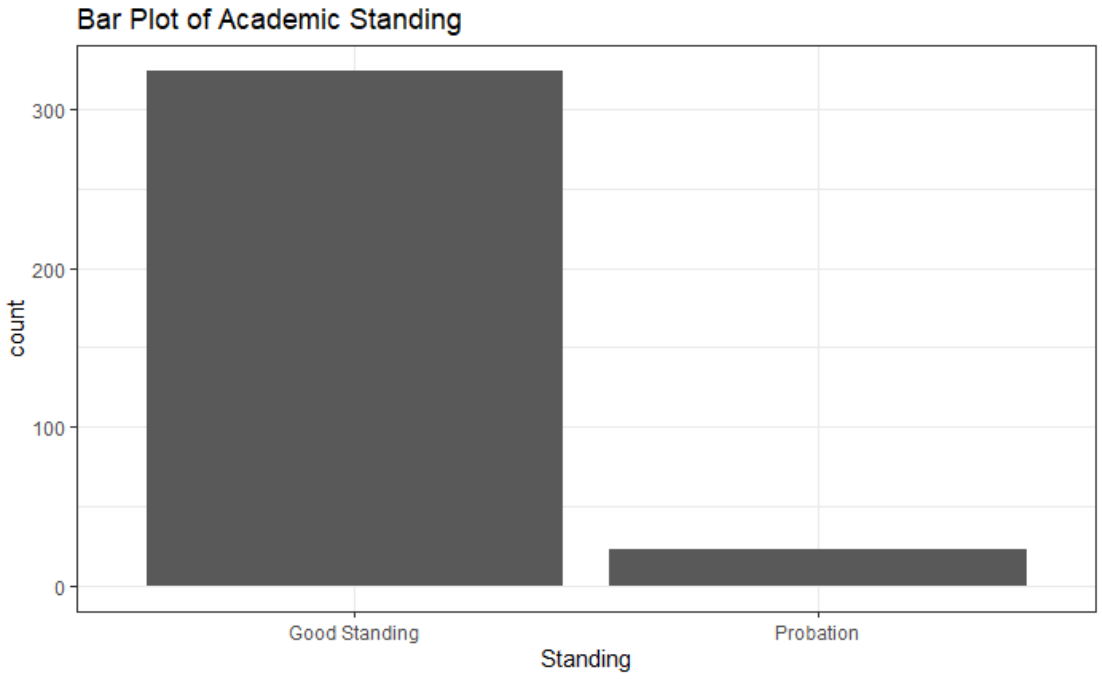


Figure 3: Histogram of Depression Score Index.

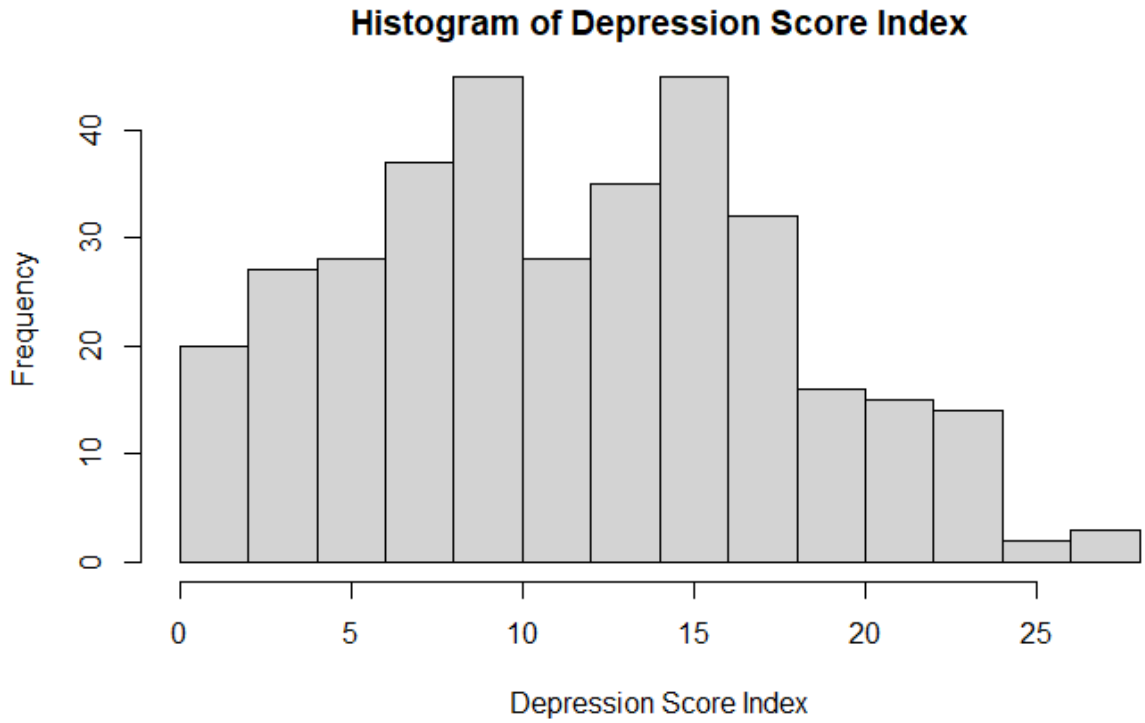



Figure 4: Scatterplot of Academic Standing and Depression Score.

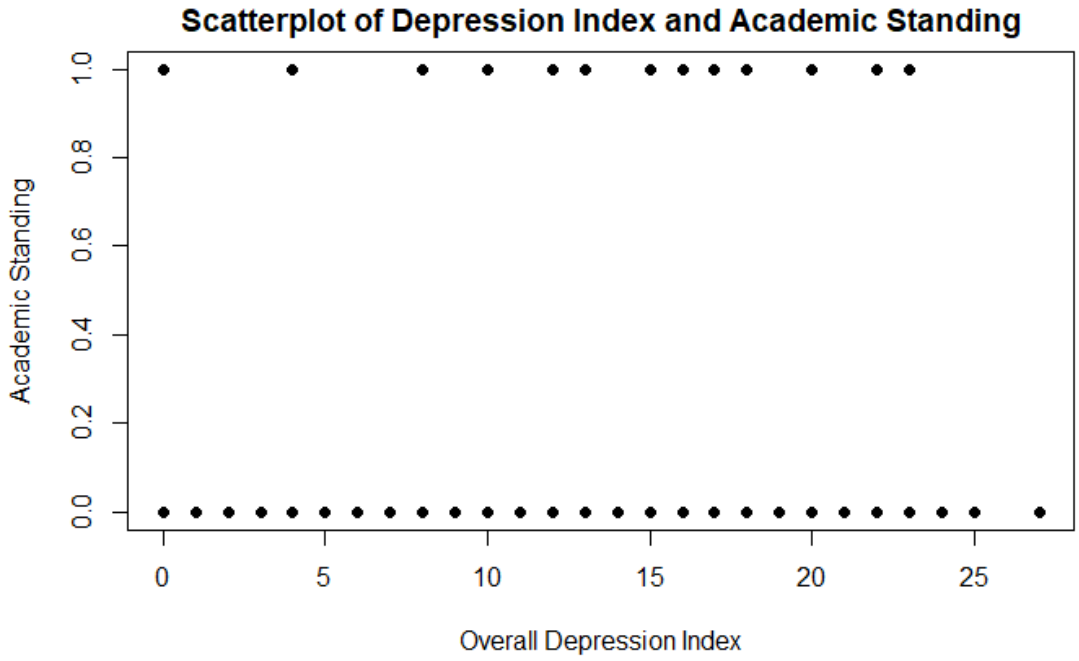


Figure 5: Boxplot of Depression Score and Academic Standing.

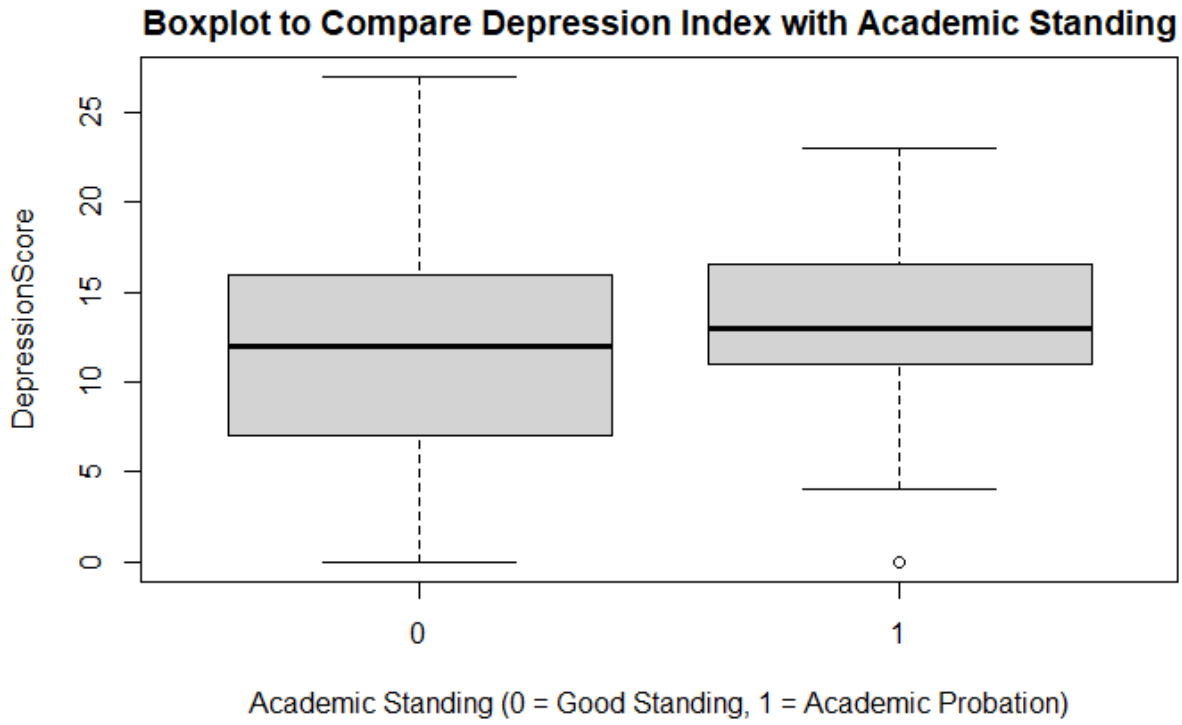



Figure 6: Effects Plot of Depression Score and Age on Academic Standing

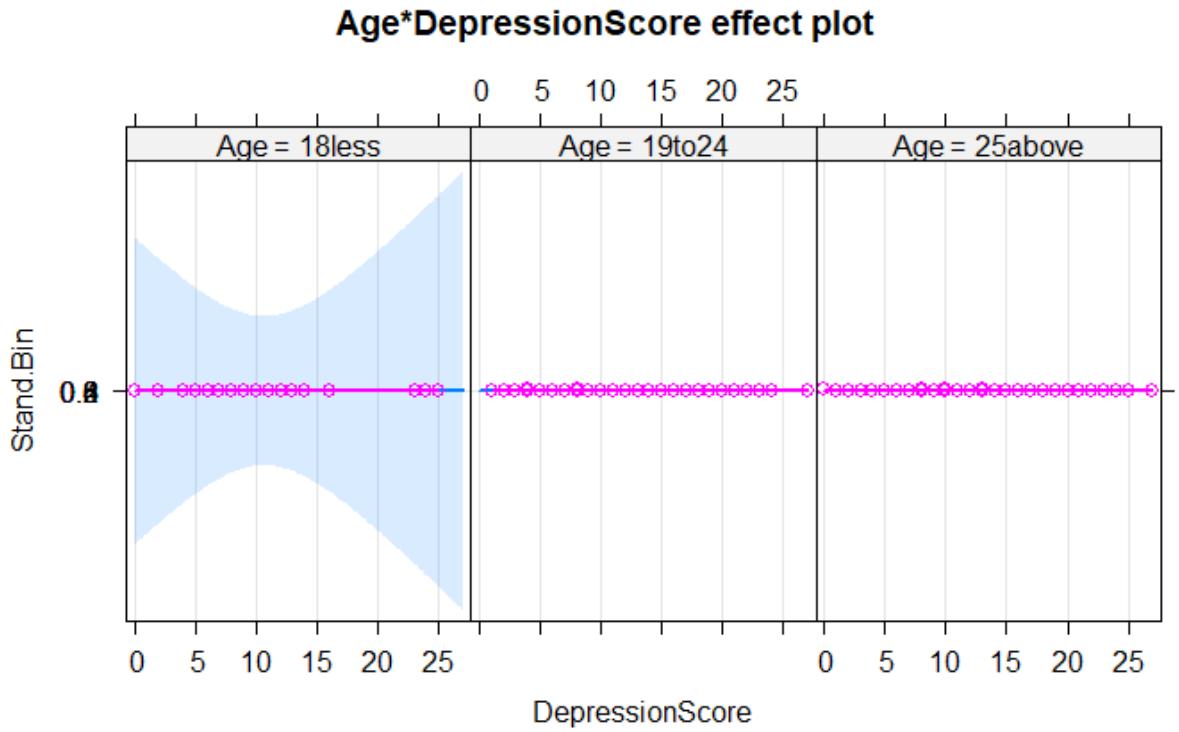


Figure 7: Effects Plot of Depression Score and Sex on Academic Standing

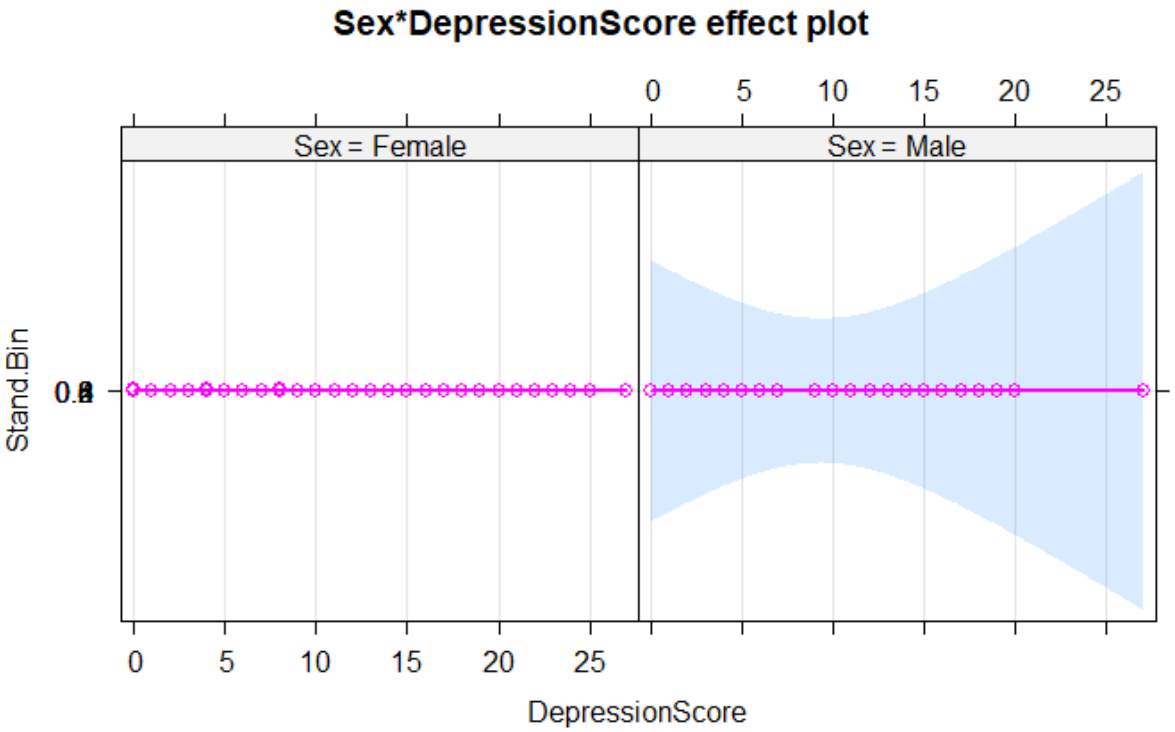
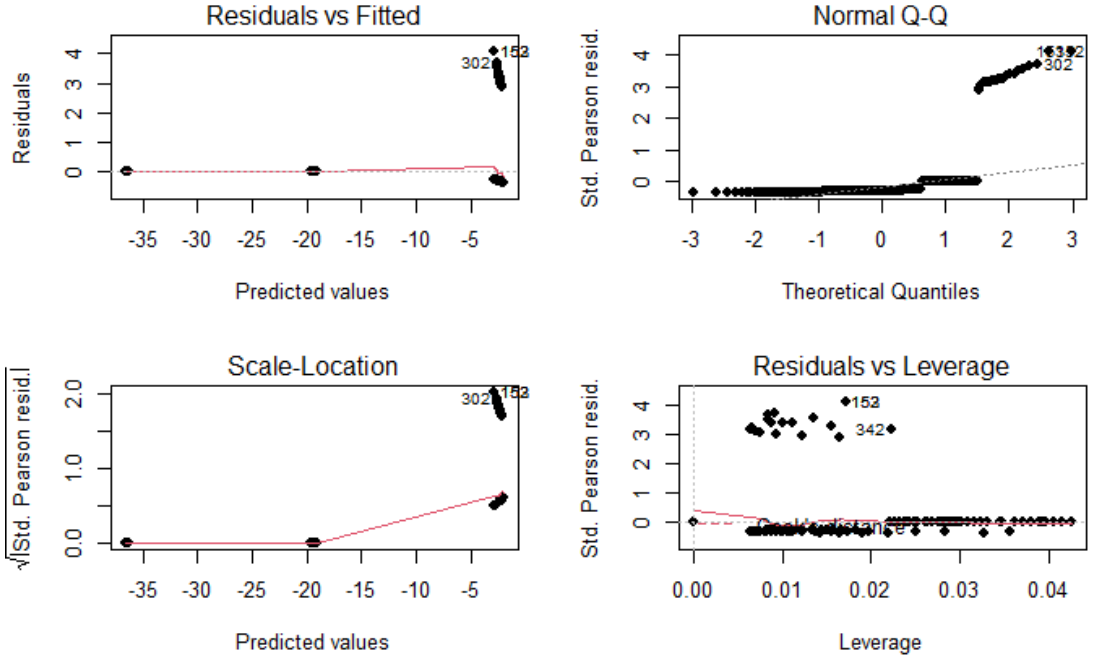



Figure 8: Standard Suite of Diagnostic Plots for the Final Model.



## Model Summaries

Model 1: Estimated Coefficients for the Model estimating the Probability of a student being placed on Academic Probation with Sex, Age, and Depression Score Index as predictors.

```
glm(formula = Stand.Bin ~ DepressionScore + Age + Sex, family = binomial(link = "logit"),
    data = Acad.new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5052  -0.4294  -0.3934  -0.0001   2.3935

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      -19.66497 1751.86154  -0.011    0.991
DepressionScore    0.02288    0.03586   0.638    0.523
Age19to24         17.05319 1751.86151   0.010    0.992
Age25above        16.85922 1751.86152   0.010    0.992
SexMale          -16.93007 1705.22367  -0.010    0.992

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 169.28  on 346  degrees of freedom
Residual deviance: 157.85  on 342  degrees of freedom
AIC: 167.85

Number of Fisher Scoring iterations: 18
```

## Tables

Table 1: Summary of the Model that includes All Interesting Predictors Additively.

```
glm(formula = Stand.Bin ~ DepressionScore + Sex + Age + EduLevel +
    Job + LivingSituation + Study + SocialMedia, family = binomial(link = "logit"),
    data = Acad.new)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.67449  -0.45479  -0.36246  -0.00011   2.42675

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)             -1.844e+01  7.292e+03  -0.003    0.998
DepressionScore          3.240e-02  3.702e-02   0.875    0.381
SexMale                 -1.683e+01  1.664e+03  -0.010    0.992
Age19to24                1.658e+01  1.729e+03   0.010    0.992
Age25above               1.619e+01  1.729e+03   0.009    0.993
EduLevelHS              -6.789e-01  5.732e-01  -1.184    0.236
EduLevelMast            -1.670e+01  2.065e+03  -0.008    0.994
JobNone                 -9.914e-03  5.623e-01  -0.018    0.986
JobPartTime             -5.356e-02  5.453e-01  -0.098    0.922
LivingSituationoffCampus -3.677e-01  7.084e+03   0.000    1.000
LivingSituationwParents  -5.546e-01  7.084e+03   0.000    1.000
Study2to4hrs             2.619e-02  5.052e-01   0.052    0.959
Study4above             -1.173e+00  1.076e+00  -1.090    0.276
SocialMedia2to4hrs       2.189e-01  8.312e-01   0.263    0.792
SocialMedia4above       -1.656e-01  8.093e-01  -0.205    0.838

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 169.28  on 346  degrees of freedom
Residual deviance: 151.24  on 332  degrees of freedom
AIC: 181.24

Number of Fisher Scoring iterations: 18
```

## Forms

Form 1: Medical PHQ-9 Depression Screening Questionnaire.

**PATIENT HEALTH QUESTIONNAIRE (PHQ-9)**

**ID #:** _____ **DATE:** _____

| Over the last *2 weeks*, how often have you been bothered by any of the following problems? *(use "✓" to indicate your answer)* | Not at all | Several days | More than half the days | Nearly every day |
|---|---|---|---|---|
| **1.** Little interest or pleasure in doing things | 0 | 1 | 2 | 3 |
| **2.** Feeling down, depressed, or hopeless | 0 | 1 | 2 | 3 |
| **3.** Trouble falling or staying asleep, or sleeping too much | 0 | 1 | 2 | 3 |
| **4.** Feeling tired or having little energy | 0 | 1 | 2 | 3 |
| **5.** Poor appetite or overeating | 0 | 1 | 2 | 3 |
| **6.** Feeling bad about yourself—or that you are a failure or have let yourself or your family down | 0 | 1 | 2 | 3 |
| **7.** Trouble concentrating on things, such as reading the newspaper or watching television | 0 | 1 | 2 | 3 |
| **8.** Moving or speaking so slowly that other people could have noticed. Or the opposite — being so figety or restless that you have been moving around a lot more than usual | 0 | 1 | 2 | 3 |
| **9.** Thoughts that you would be better off dead, or of hurting yourself | 0 | 1 | 2 | 3 |
| | | add columns | + | + |

*(Healthcare professional: For interpretation of TOTAL, please refer to accompanying scoring card).* TOTAL:

**10.** If you checked off *any problems*, how *difficult* have these problems made it for you to do your work, take care of things at home, or get along with other people?

- Not difficult at all _____
- Somewhat difficult _____
- Very difficult _____
- Extremely difficult _____



A2663B 10-04-2005