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 struggle 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, 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

This data set was sourced from *Kaggle*, a data-sharing platform open to the public. Our data in particular was shared by an individual named Kane Rudolph (2021). The data was collected via

surveying students at varying education levels across the United States. A total of 352 students' responses to the questionnaire - completed with informed consent - were included in the study. Of the 352 students, 124 were high schoolers, 204 were undergraduates pursuing their Bachelor's degree, and the remaining 24 were graduate students enrolled in Master's programs.

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 question 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 ≥ 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 and explore 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. Data Visualization

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 the 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 indicating an outlier in the probation group (Figure 5) (R Core Team, 2021). To visualize the relationship between our two main variables of interest, an Enhanced Violin Plot was created with the ‘yarrr’ package in R (Phillips, 2017).

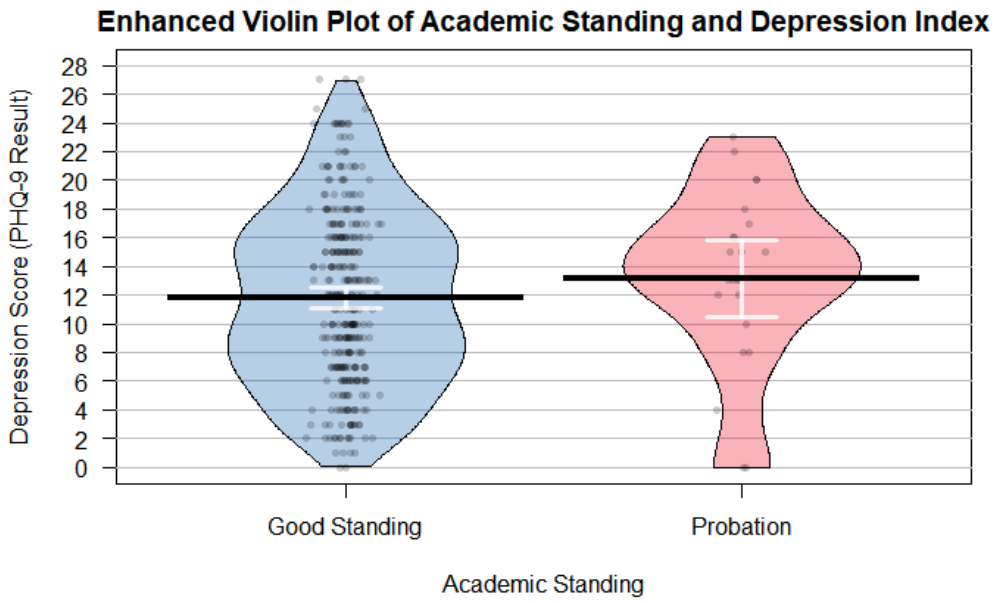


Figure 6: Enhanced Violin Plot of the Distribution of Depression Scores within each Academic Standing Category.

The plot above shows that, on average, depression scores for students with good standing is slightly lower compared to students on academic probation. Through further data exploration using the 'mosaic' package in R, the mean PHQ-9 score for those with good standing was 11.81 points compared to a mean of 13.17 points for those who qualify for academic probation (Pruim, 2017). It is also worth noting that the confidence interval for the probation group is much wider than for the good standing group; This is most likely due to the vast difference in the total number of students in each group.

V. 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 were generated with the effects package (Fox and Weisberg, 2019) (Figures 7, 8, & 9). The incongruity of the blue smoothing lines and the pink fitted lines in Figures 7 & 8, in addition to the converging nature of the lines in Figure 9, suggested potential for two-way interactions between our three predictors. Thus, additional forward stepwise regression was performed and 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:

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

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, but since we are dealing with a binary response, most of our typical diagnostic tools are not applicable (R Core Team) (Figure 10). The only relevant diagnostic plot is the Residual vs. Leverage plot, which is used to assess

whether the Pearson Residuals indicate a potential for influential points (R Core Team) (Figure 11). No points in our data set have a Cook's Distance greater than 0.5 in the Residual vs. Leverage plot, suggesting no evidence of influential points in our data set.

Tables were generated to get a preliminary idea of the distribution of students on academic probation within the sex, age, and depression level subgroups (Tables 2, 3, & 4). Table 2 indicated that there are notably more female participants than male participants. Surprisingly, none of the male participants in this sample qualified for academic probation. Table 3 explored the relationship between age groups and academic standing, and found that there were no students 18 and younger in this sample who qualified for probation. Table 4 suggests that there are a larger number of students who have "Mild", "Moderate", or "Moderately Severe" depression compared to those with "Normal/Minimal" and "Severe" depression.

VI. Methods of Analysis and Results

To evaluate how well our final model fits our data, we ran a Hosmer-Lemeshow Goodness of Fit Test using the 'ResourceSelection' package in R (Lele, 2019). This test is used to determine how well the observed number of students currently on academic probation matches the expected amount of students on academic probation. The results of the Hosmer-Lemeshow Test provide little to no evidence that a more complicated model would better explain our data ($\chi^2 = 11.752$, df = 8, p-value = 0.1626). Thus, we deem our model to fit the data well.

The following interpretations are the exponentiated coefficients from our model:

Intercept: The estimated odds of qualifying for academic probation for a female age 18 years old or younger with a depression score of zero points are 0.00000000288147 to 1.

DepressionScore: A one point increase in depression score is associated with an estimated 2.31% increase in the odds of qualifying for academic probation, controlling for age group and sex of the individual.

Age19to24: Among individuals of the same sex who have the same depression score, the odds ratio of qualifying for academic probation for students between 19 to 24 years old is 25474570.

Age25above: Among individuals of the same sex who have the same depression score, the odds ratio of qualifying for academic probation for students who are 25 and older is 20982960.

SexMale: The estimated odds of qualifying for academic probation for males are 99.99% lower than for females, controlling for depression score and age group. This extreme difference in odds is likely due to the disparity we previously noted in Table 2; We had a much larger proportion of female respondents than male, and zero male students in our sample qualified for academic probation.

The primary purpose of this model is to predict how likely a student is to be placed on academic probation based on their depression score, sex, and age group. To assess the predictive ability of our model, we created a ROC curve using the 'ROCR' package in R (Sing, 2005) (Figure 12). In our context, the 'true positive rate' would translate to how often a student on academic probation was accurately predicted to be on academic probation; The 'false positive rate' would represent when a student is predicted to qualify for probation but actually has good standing. When visually gauging prediction accuracy, we noted our ROC curve is similar to what we would expect to see if the model were making predictions by random chance (Figure 12). After creating the plot, the area under the curve was calculated to quantify the model's prediction ability (Sing, 2005). The area under the ROC curve is about 0.6622, indicating that the model does not predict academic probation very well. Specifically, the probability that a randomly

chosen student on academic probation has a higher estimated probability of probation than a randomly chosen student with good standing is 0.6622.

VII. Discussion

This was an observational study and the data were obtained through a survey, so any statistical findings cannot be assumed to have a causal association. Since the data were not collected from a random sample, findings may be generalized only to individuals similar to those in the study (i.e. those in a position to be given a PHQ-9 form who consent to complete it).

In addition, using surveys for data collection introduces the potential for sampling biases. Not only is there potential for response bias, such as patients downplaying the severity of symptoms to avoid involuntary hospitalization, but individuals who fill out the PHQ-9 form may be fundamentally different than the overall population of students with depression. It is likely that students who are prompted to complete a PHQ-9 form are actively seeking out medical help. Whether it be for symptoms of depression or other reasons, students who actively seek medical aid may differ from students who avoid medical attention for the purposes of this study.

One observed oddity is that the source advertised the data set as exploring the relationship between academic performance and symptoms of depression and anxiety, suggesting students who participated took the General Anxiety Disorder assessment (GAD-7) as well, but results from this form were not included in the data. Unfortunately, this could be an unaccounted for confounding variable as depression and anxiety are often comorbidities; Anxiety disorders are associated with a unique set of debilitating symptoms that may contribute to a student's performance in school.

Should further research be conducted, some form of random sampling would be preferable for the generalizability of findings and independence of observations. We recommend collecting data regarding each students' cumulative GAD-7 score in addition to the existing variables. Including anxiety as a predictor in our model may better explain academic performance. Once in the model, there is potential for an interaction between depression and anxiety score, meaning the relationship between depression score and academic performance may differ with different levels of anxiety. Thus, the interaction between symptoms of depression and anxiety on academic performance is important to explore.

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.

K. Rudolph. “Depression and Academice Performance of Students” (2021). *Kaggle*. https://www.kaggle.com/datasets/kanerudolph/depression-and-academic-performance-of-students.

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

N. Phillips. The yarrr Package: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R" (2017). R package version 0.1.5. https://CRAN.R-project.org/package=yarrr

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

R. Pruim, D. Kaplan, and N. Horton. The mosaic Package: Helping Students to 'Think with Data' Using R (2017). The R Journal, 9(1):77-102. https://journal.r-project.org/archive/2017/RJ-2017-024/index.html

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

S. Lele, J. Kiem, and P. Solymos. The ResourceSelection Package: Resource Selection (Probability) Functions for Use-Availability Data (2019). R package version 0.3-5. https://CRAN.R-project.org/package=ResourceSelection

T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer. The ROCR Package: Visualizing Classifier Performance in R (2005). Bioinformatics, 21(20):3940-3941. http://rocr.bioinf.mpi-sb.mpg.de

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

Figures

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

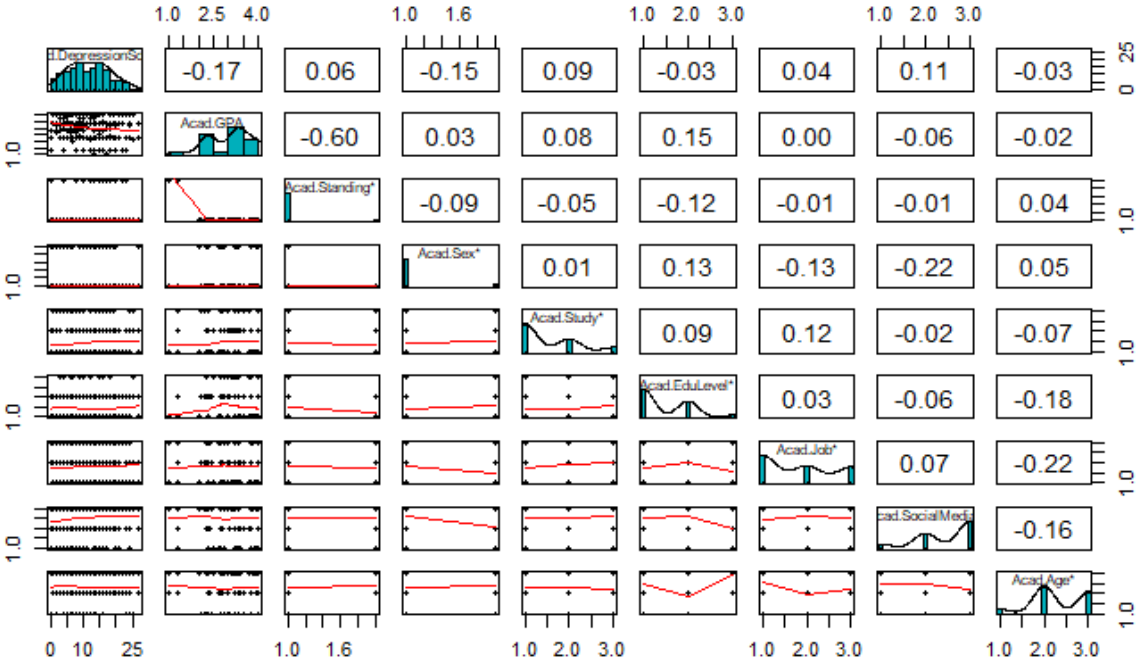


Figure 2: Bar Plot of Academic Standing.

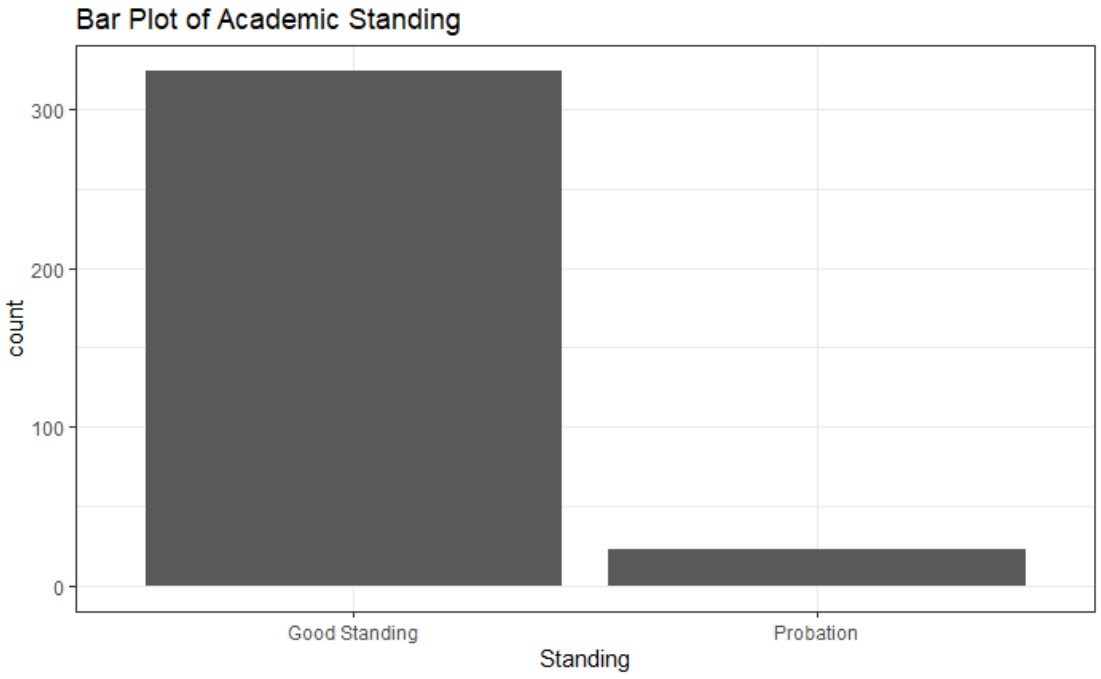


Figure 3: Histogram of Depression Score Index.

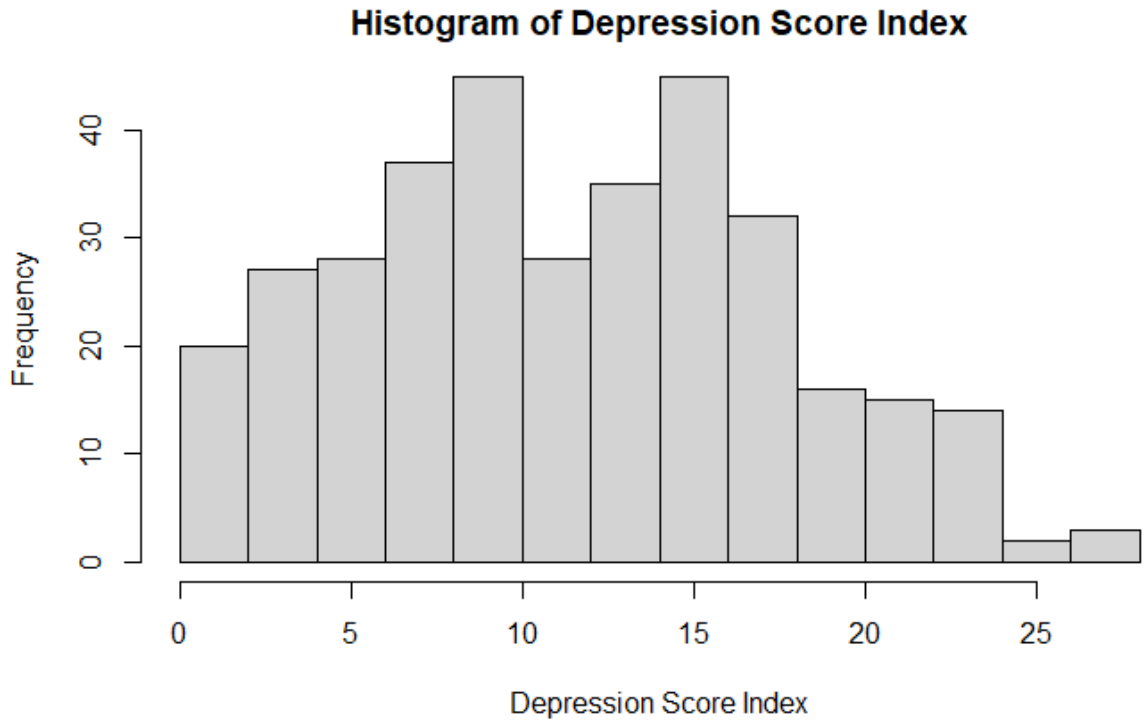


Figure 4: Scatterplot of Academic Standing and Depression Score.

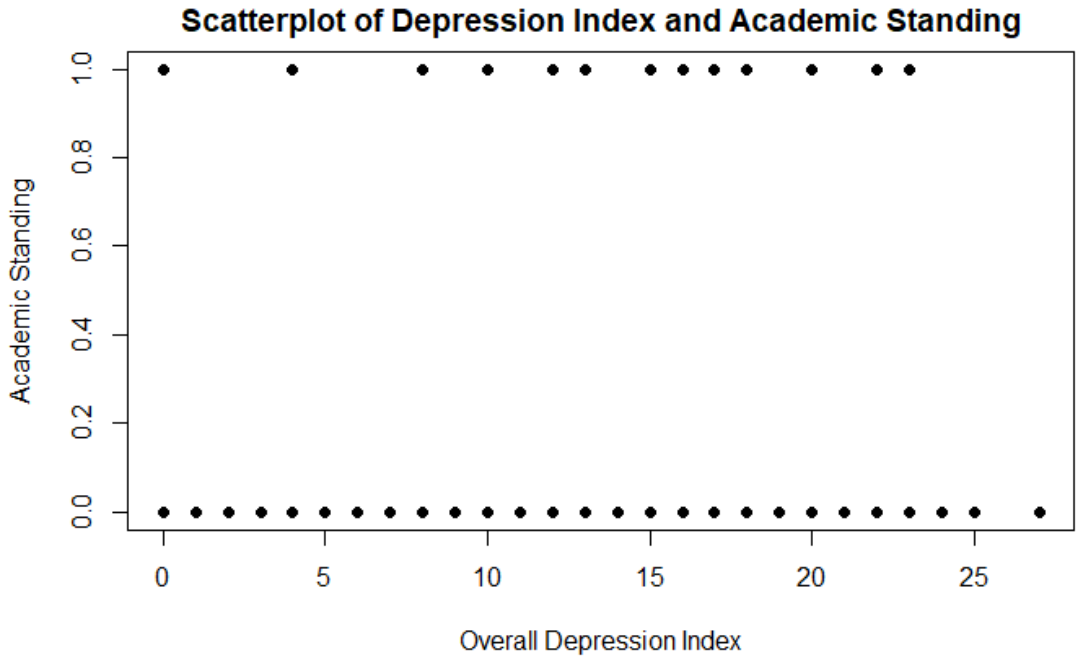


Figure 5: Boxplot of Depression Score and Academic Standing.

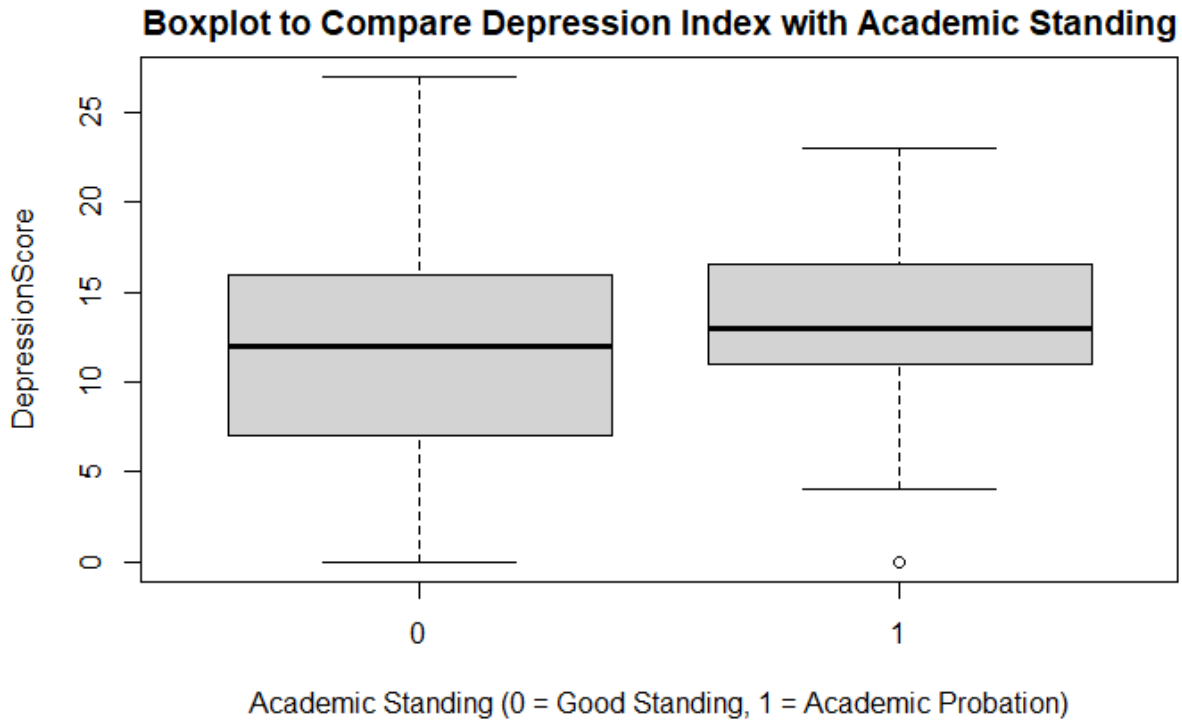


Figure 7: Effects Plot of Depression Score and Age on Academic Standing

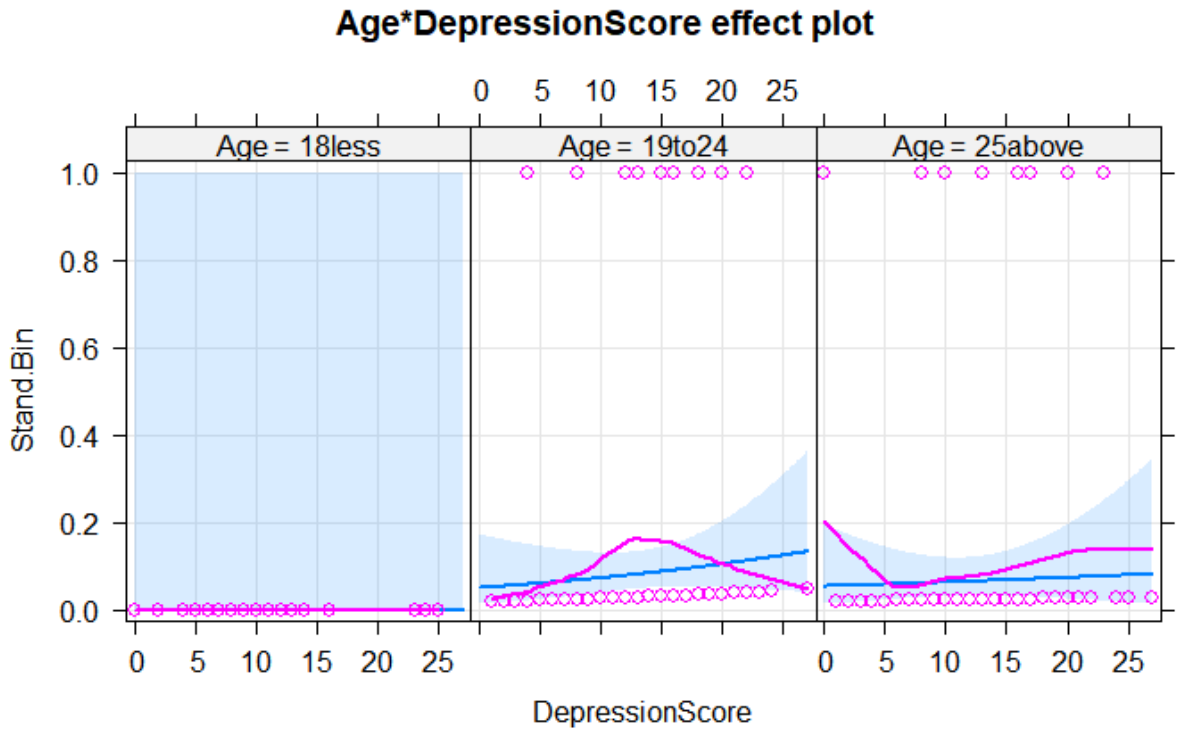


Figure 8: Effects Plot of Depression Score and Sex on Academic Standing

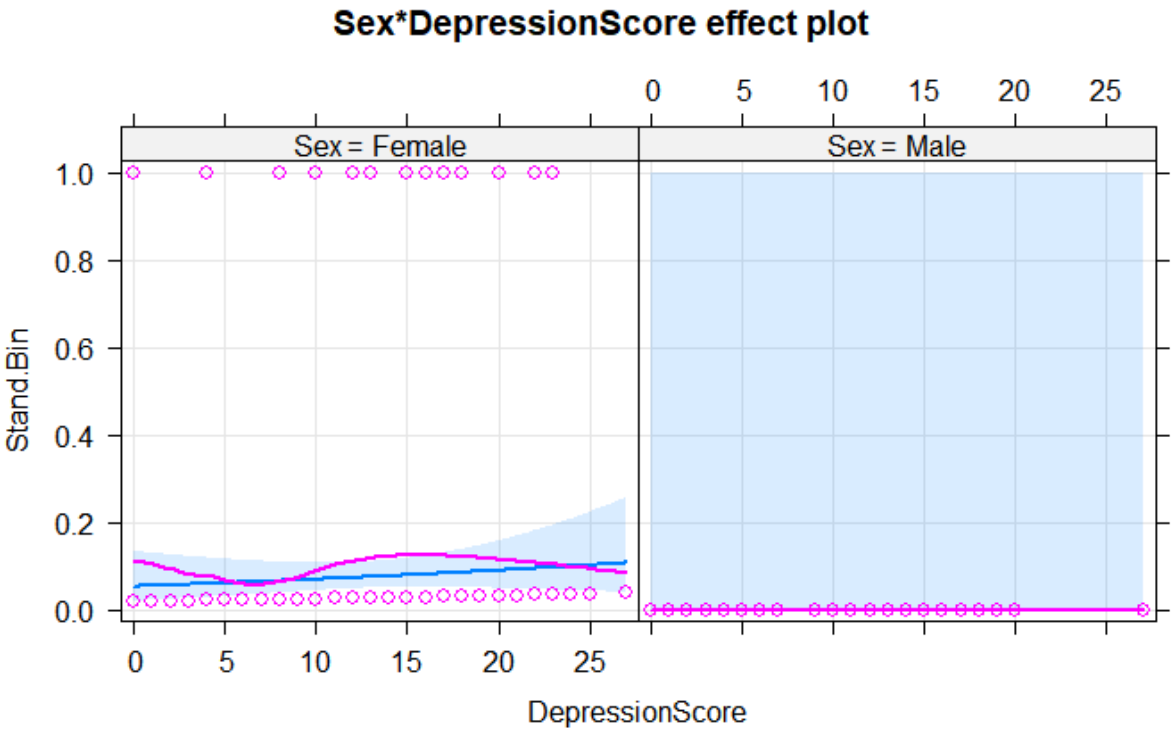


Figure 9: Effects Plot of Age Group and Sex on Academic Standing

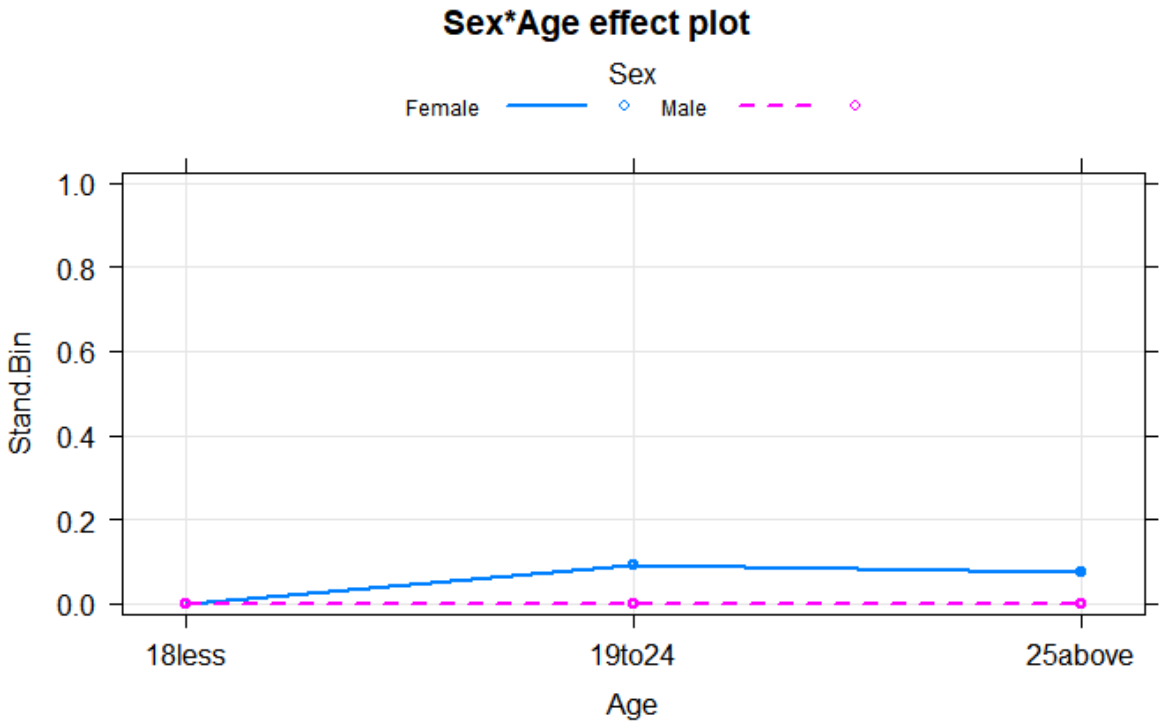


Figure 10: Standard Suite of Diagnostic Plots for the Final Model.

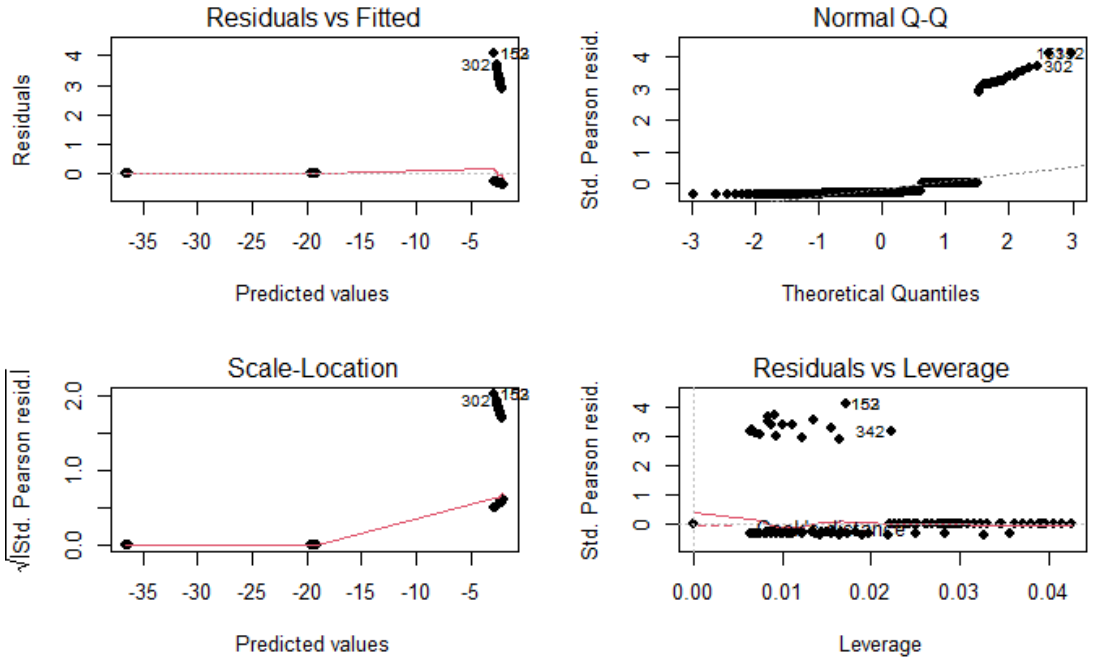


Figure 11: Residual vs. Leverage Diagnostic Plot for the Final Model.

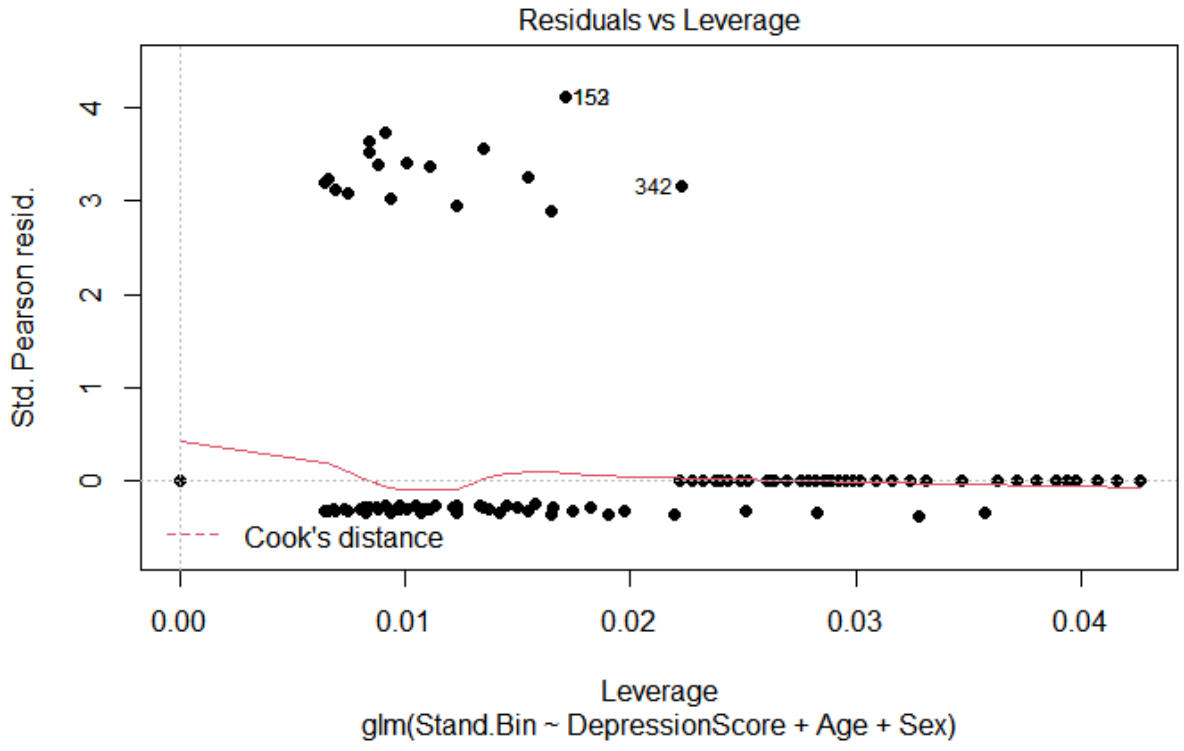
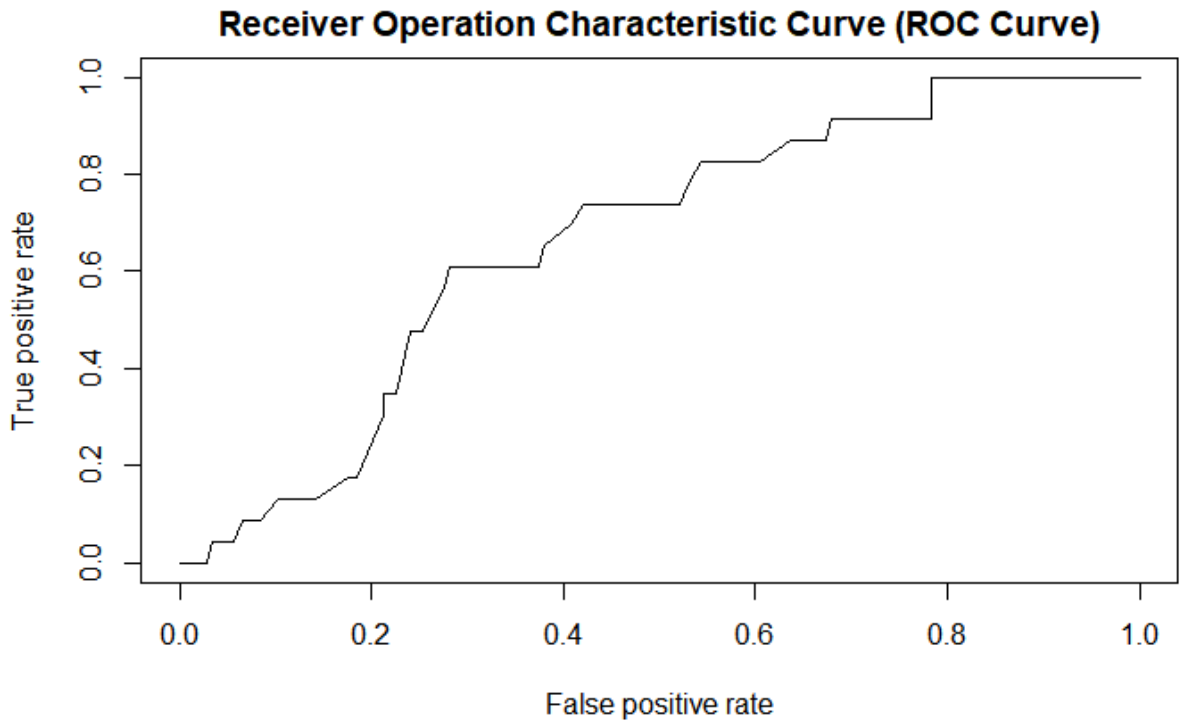


Figure 12: Receiver Operation Characteristic Curve (ROC)



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 (on the log-scale).

	Estimate	Standard Error	z value	p-value
(Intercept)	-19.66497	1751.86154	-0.011	0.991
DepressionScore	0.02288	0.03586	0.638	0.523
Age = 19 to 24	17.05319	1751.86151	0.010	0.992
Age = 25+	16.85922	1751.86152	0.010	0.992
Sex = Male	-16.93007	1705.22367	-0.010	0.992

95% Confidence Intervals for the Estimated Model Coefficients on the Response Scale.

	95% CI Lower Bound	95% CI Lower Bound
(Intercept)		
DepressionScore		
Age = 19 to 24		
Age = 25+		
Sex = Male		

Tables

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

	Estimate	Standard Error	z value	p-value
(Intercept)	-18.44	729.2	-0.003	0.998
DepressionScore	0.03240	0.03702	0.875	0.381
Sex = Male	-16.83	166.4	-0.010	0.992
Age = 19 to 24	16.58	172.9	0.010	0.992
Age = 25+	16.19	172.9	0.009	0.993
Education Level = High School	-0.6789	0.5732	-1.184	0.236
Education Level = Masters	-16.70	206.5	-0.008	0.994
Job = None	-0.009914	0.5623	-0.018	0.986
Job = Part Time	-0.05356	0.5453	-0.098	0.922
Housing = Off Campus	-0.3677	708.4	0.000	1.00
Housing = With Parents	-0.5546	708.4	0.000	1.00
Study Time = 2-4 Hours Daily	0.02619	0.5052	0.052	0.959
Study Time = 4+ Hours Daily	-1.173	1.076	-1.090	0.276
Social Media = 2-4 Hours Daily	0.2189	0.8312	0.263	0.792
Social Media = 4+ Hours Daily	-0.1656	0.8093	-0.205	0.838

Table 2: Counts of Students within the combinations of Academic Standing and Sex.

		Sex	
		Female	**Male**
Academic Standing	**Good Standing**	286	38
	Probation	23	0

Table 3: Counts of Students within the combinations of Academic Standing and Age Group.

		Age Group		
		18 and younger	**19 to 24 years old**	**25 and older**
Academic Standing	**Good Standing**	36	157	131
	Probation	0	14	9

Table 4: Counts of Students within the combinations of Academic Standing and Depression Level.

		Depression Level				
		Normal	**Mild**	**Moderate**	**Moderately Severe**	**Severe**
PHQ-9 Score Range		**0-4 points**	**5-9 points**	**10-14 points**	**15-19 points**	**20+ points**
Academic Standing	**Good Standing**	44	86	78	78	38
	Probation	3	2	7	7	4

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

Relevant R Output

Output 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
```

Output 2: 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
```

Output 3: Results of the Hosmer-Lemeshow Test for our Final Model.

```
        Hosmer and Lemeshow goodness of fit (GOF) test

data:  obs, expected
X-squared = 11.752, df = 8, p-value = 0.1626
```