Depression's Influence on Academic Performance

STAT 439 - Data Project Draft Report

Harley Clifton, Sara Bancale, & Matt McMahon

April 4, 2022

Project Scaffolding

```
Acad <- read_csv("AcademicPerformance.csv")
```

```
## Rows: 352 Columns: 18
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (8): Gender:, Age:, Educational Level, Do you have part-time or full-ti...
## dbl (10): Little interest or pleasure in doing things, Feeling down, depress...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
Acad <- Acad %>% rename(Age='Age:',
            Pleasure=`Little interest or pleasure in doing things`,
            Depressed=`Feeling down, depressed, or hopeless`,
            Sleep=`Trouble falling or staying asleep, or sleeping too much`,
            Tired=`Feeling tired or having little energy`,
            Appetite=`Poor appetite or overeating`,
            Feeling=`Feeling bad about yourself or that you are a failure or not have let yourself or y
            Concentrating=`Trouble concentrating on things, such as reading the newspaper or watching t
            Slow=`Moving or speaking so slowly that other people could have noticed Or being so restless
            Thoughts=`Thoughts that you would be better off dead or of hurting yourself in some way`,
            Job=`Do you have part-time or full-time job?`,
            LivingSituation=`Which of the following best describes your term-time accommodation?`,
            Study=`How many hours do you spend studying each day?`,
            Electronics=`How many of the electronic gadgets (e.g. mobile phone, computer, laptop, PSP, I
            SocialMedia=`How many hours do you spend on social media per day?`,
            GPA=`Your Last Semester GPA:`,
            Sex=`Gender:`,
            EduLevel='Educational Level')
```

Data Wrangling

```
# Data Wrangling

Acad <- Acad %>%
  mutate(Standing = case_when(GPA < 2 ~"Probation",
                   GPA >= 2 ~"Good Standing"))

Acad$DepressionScore = Acad$Depressed + Acad$Sleep + Acad$Tired + Acad$Appetite +
  Acad$Feeling + Acad$Concentrating + Acad$Slow + Acad$Thoughts + Acad$Pleasure - 9

Acad <- Acad %>%
  mutate(DepressionLevel= case_when(
    DepressionScore <= 4 ~ "Normal",
    DepressionScore >= 5 & DepressionScore <= 9 ~ "Mild",
    DepressionScore >= 10 & DepressionScore <= 14 ~ "Moderate",
    DepressionScore >= 15 & DepressionScore <= 19 ~ "Moderately Severe",
    DepressionScore >= 20 ~ "Severe"))

Acad <- Acad %>%
  mutate(EduLevel = case_when(EduLevel == "High School" ~ "HS",
                              EduLevel == "College - Bachelor's" ~ "Bach",
                              EduLevel == "Master" ~ "Mast"))

Acad <- Acad %>%
  mutate(Study = case_when(Study == "1 - 2 hours" ~ "1to2hrs",
                           Study == "2 - 4 hours" ~ "2to4hrs",
                           Study == "More than 4 hours" ~ "4above"))

Acad <- Acad %>%
  mutate(SocialMedia = case_when(SocialMedia == "1 - 2 Hours" ~ "1to2hrs",
                           SocialMedia == "2 - 4 Hours" ~ "2to4hrs",
                           SocialMedia == "More than 4 Hours" ~ "4above"))

Acad <- Acad %>%
  mutate(Age = case_when(Age == "18 years or less" ~ "18less",
                         Age == "19 to 24 years" ~ "19to24",
                         Age == "25 years and above" ~ "25above"))

Acad <- Acad %>%
  mutate(Job = case_when(Job == "No" ~ "None",
                         Job == "Part time" ~ "PartTime",
                         Job == "Part Time" ~ "PartTime",
                         Job == "part Time" ~ "PartTime",
                         Job == "part time" ~ "PartTime",
                         Job == "Full time" ~ "FullTime",
                         Job == "Full Time" ~ "FullTime",
                         Job == "full Time" ~ "FullTime",
                         Job == "Full Time" ~ "FullTime"))

Acad <- Acad %>%
  mutate(LivingSituation = case_when(LivingSituation == "Home (with parents)" ~ "wParents",
                         LivingSituation == "University hall of residence" ~ "Dorm",
```

```
                        LivingSituation == "Private rented accommodation" ~ "offCampus"))

# Create a new binary Standing variable with 'Probation' as a success
Acad.new <- Acad %>% mutate(Stand.Bin = case_when(GPA < 2 ~"1",
                                                  GPA >= 2 ~"0"))
Acad.new$Stand.Bin <- as.numeric(Acad.new$Stand.Bin)

Acad.new <- na.omit(Acad.new)
dim(Acad.new)
```

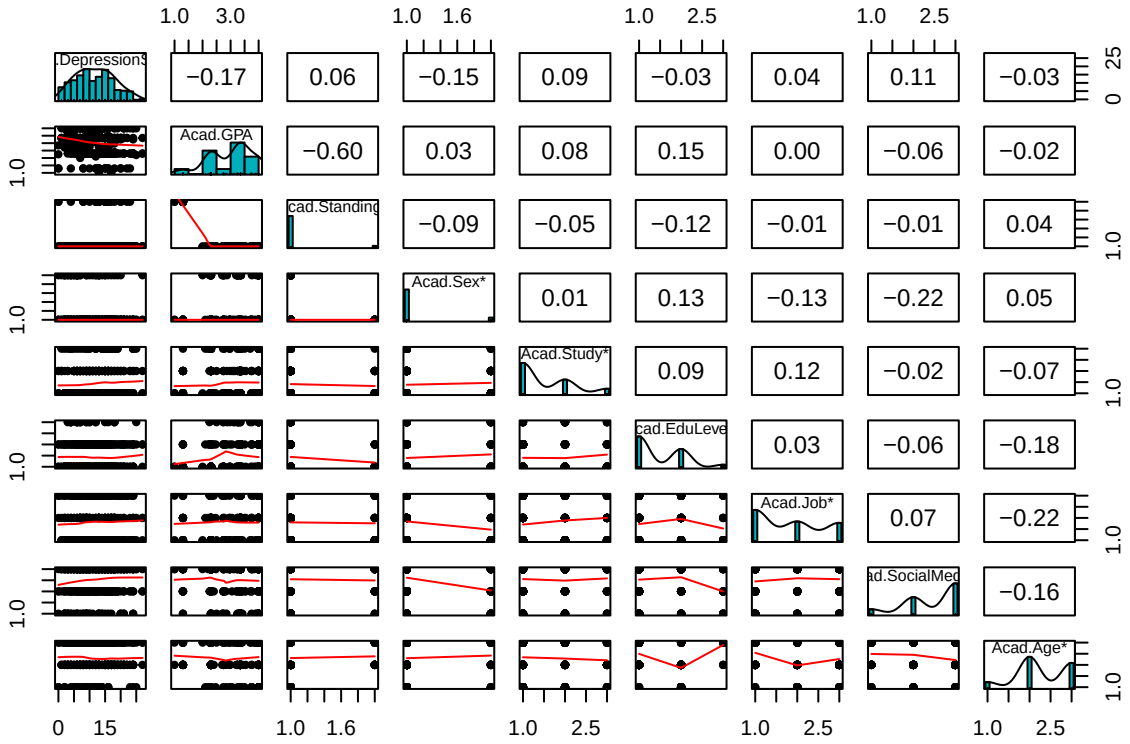
```
## [1] 347  22
```

Data Visualization

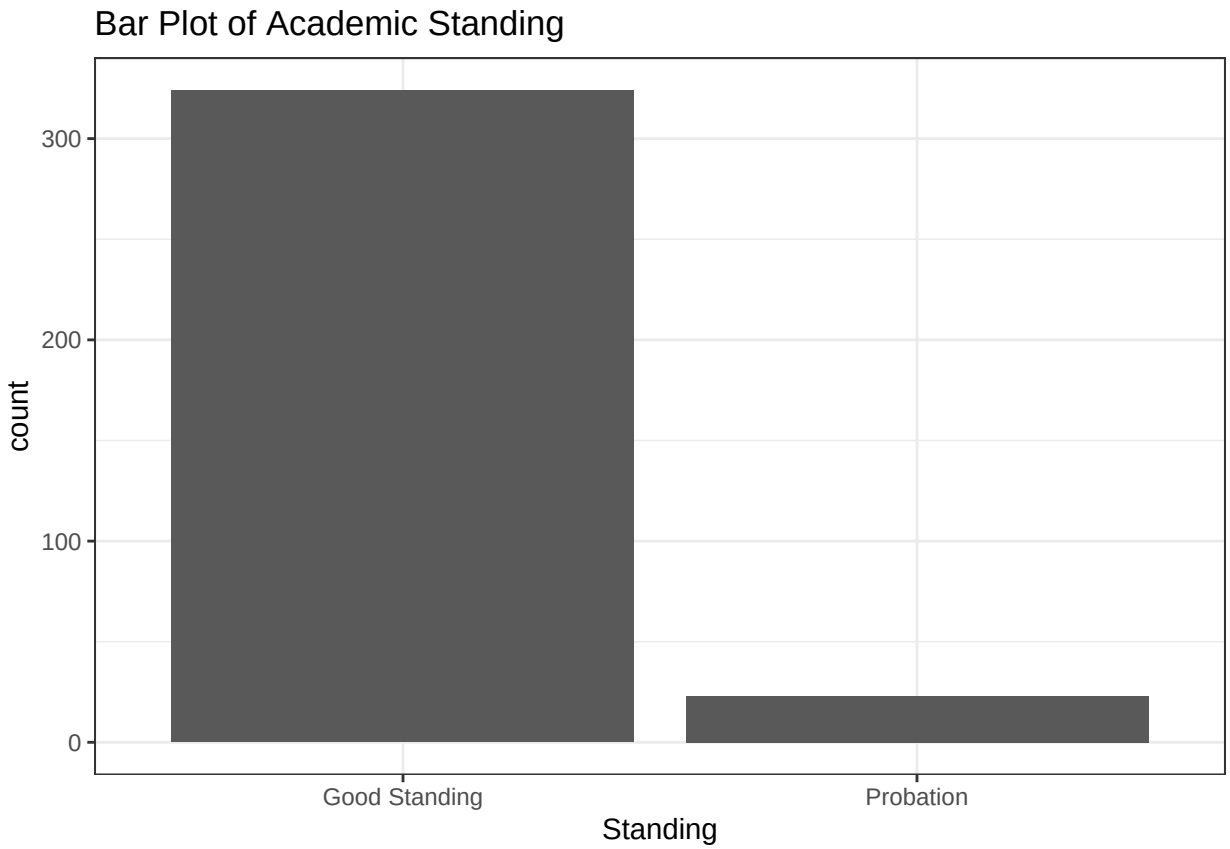
```
# Creating a Scatterplot Matrix and Correlation Matrix

Acad.sub <- data.frame(Acad$DepressionScore, Acad$GPA, Acad$Standing, Acad$Sex,
                       Acad$Study, Acad$EduLevel, Acad$Job, Acad$SocialMedia, Acad$Age)

library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
suppressWarnings(pairs.panels(Acad.sub,
             method = "pearson", # correlation method
             hist.col = "#00AFBB",
             density = TRUE,  # show density plot
             ellipses = FALSE # do not show correlation ellipses
             ))
```




```
# Univariate Distributions

Acad.new %>% ggplot(aes(x = Standing)) +
  geom_bar() +
  ggtitle("Bar Plot of Academic Standing")
```

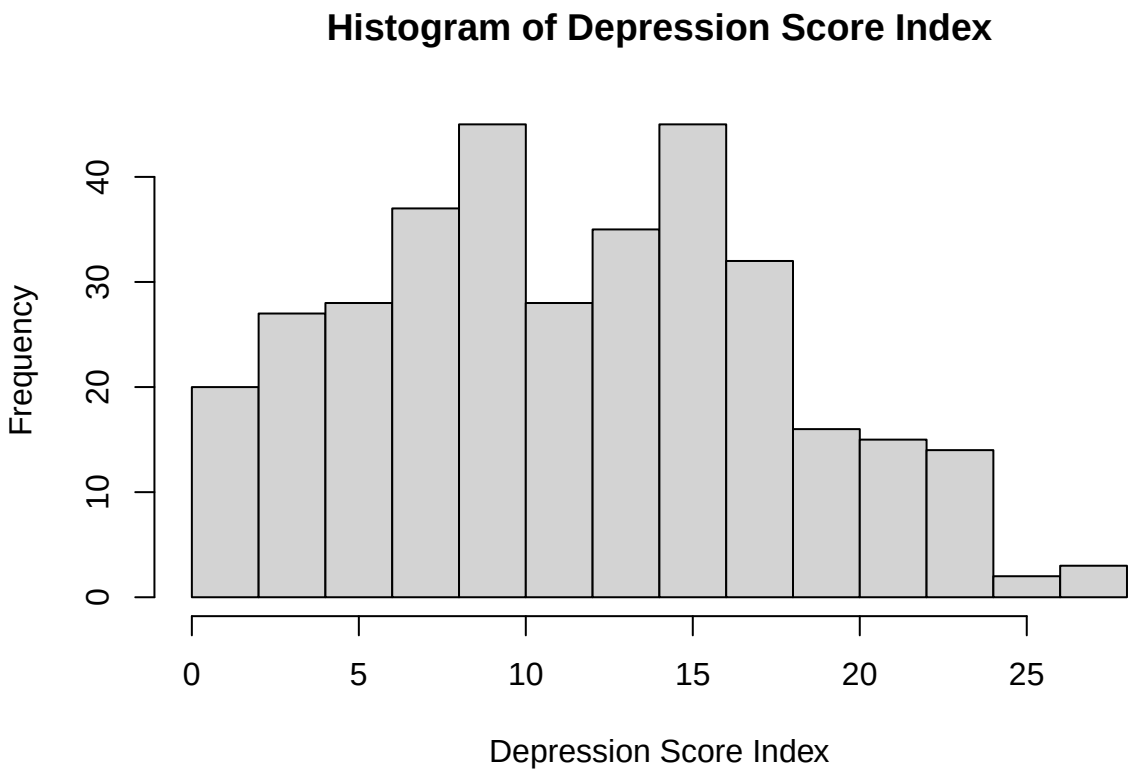



```
table(Acad.new$Standing)
```

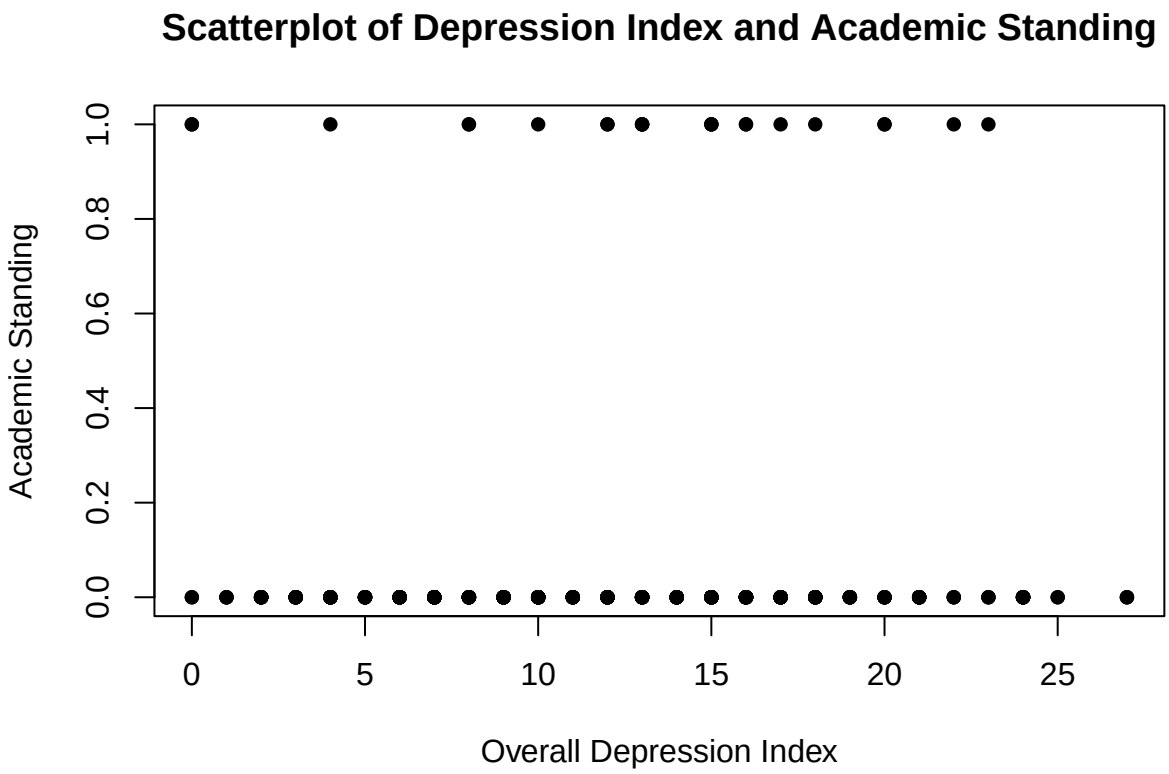
```
##
## Good Standing     Probation
##           324            23
```

Initially, we want to acknowledge that there are much fewer subjects who are on Academic Probation which may be attributed to the time sensitivity of the Academic Dismissal Decision Process.

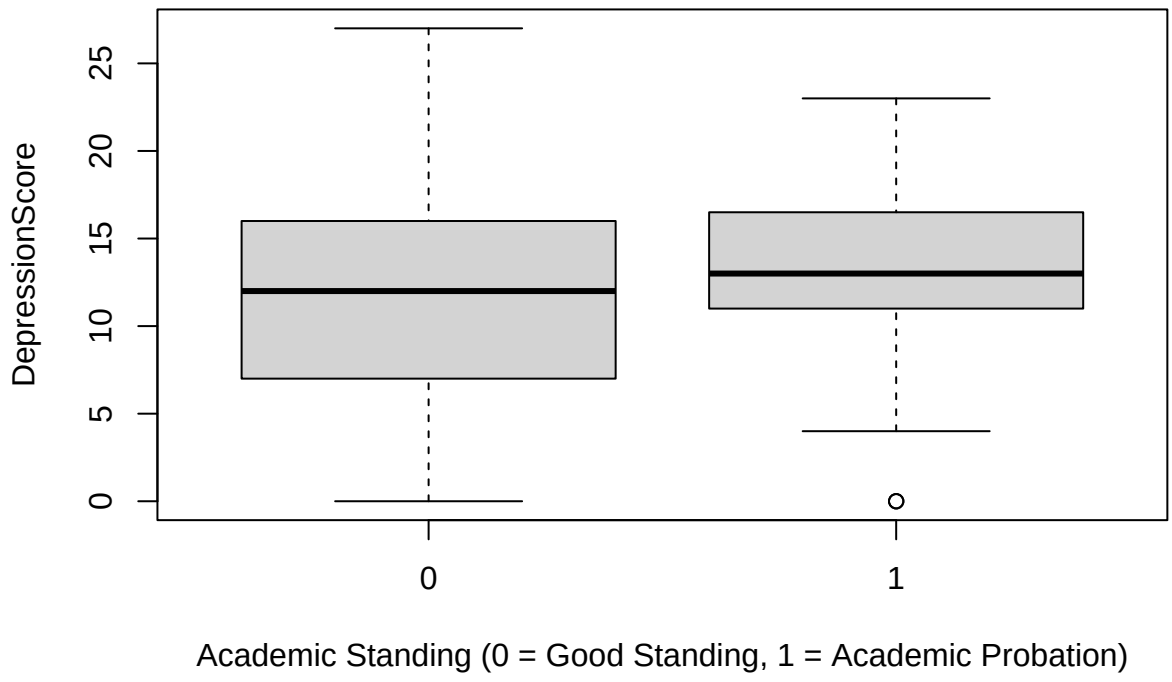
```
hist(Acad.new$DepressionScore, main = "Histogram of Depression Score Index",
     xlab = "Depression Score Index")
```



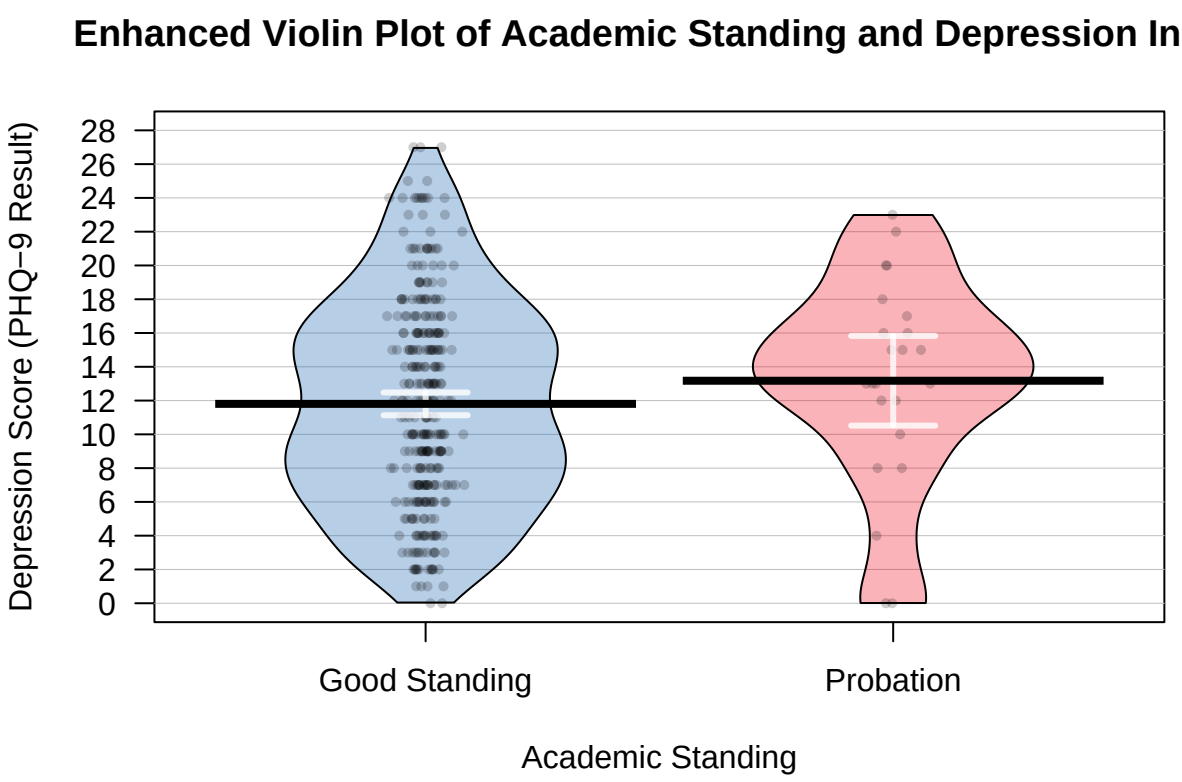

```
plot(Stand.Bin ~ DepressionScore, xlab = "Overall Depression Index",
     ylab = "Academic Standing",
     main = "Scatterplot of Depression Index and Academic Standing",
     pch = 16, data = Acad.new)
```




```
boxplot(DepressionScore ~ Stand.Bin, data = Acad.new,
        xlab = "Academic Standing (0 = Good Standing, 1 = Academic Probation)",
        main = "Boxplot to Compare Depression Index with Academic Standing")
```




```
# Enhanced Violin Plots

library(yarrr)

pirateplot(DepressionScore ~ Standing, data = Acad.new, inf.method = "ci",
           inf.disp = "line",
           main = "Enhanced Violin Plot of Academic Standing and Depression Index",
           xlab = "Academic Standing", ylab = "Depression Score (PHQ-9 Result)")
```



Data Exploration

```
library(mosaic)
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2

##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.

##
## Attaching package: 'mosaic'

## The following objects are masked from 'package:psych':
##
##     logit, rescale

## The following objects are masked from 'package:dplyr':
##
##     count, do, tally
```

```
## The following object is masked from 'package:purrr':
##
##     cross

## The following object is masked from 'package:ggthemes':
##
##     theme_map

## The following object is masked from 'package:ggplot2':
##
##     stat

## The following object is masked from 'package:BayesFactor':
##
##     compare

## The following object is masked from 'package:Matrix':
##
##     mean

## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##     quantile, sd, t.test, var

## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```

```
favstats(DepressionScore ~ Standing, data = Acad.new)
```

```
##        Standing min Q1 median   Q3 max     mean       sd   n missing
## 1 Good Standing   0  7     12 16.0  27 11.80556 6.168117 324       0
## 2     Probation   0 11     13 16.5  23 13.17391 6.139679  23       0
```

```
tally(~ Standing + Sex, data = Acad.new)
```

```
##                Sex
## Standing        Female Male
##   Good Standing    286   38
##   Probation         23    0
```

```
tally(~ Standing + Age, data = Acad.new)
```

```
##                Age
## Standing        18less 19to24 25above
##   Good Standing     36    157     131
##   Probation          0     14       9
```

```
tally(~ Standing + DepressionLevel, data = Acad.new)
```

```
##                DepressionLevel
## Standing        Mild Moderate Moderately Severe Normal Severe
##   Good Standing   86       78                78     44     38
##   Probation        2        7                 7      3      4
```

```
tally(~ Standing + EduLevel, data = Acad.new)
```

```
##                EduLevel
## Standing        Bach  HS Mast
##   Good Standing  185 116   23
##   Probation       18   5    0
```

```
tally(~ EduLevel + Age, data = Acad.new)
```

```
##         Age
## EduLevel 18less 19to24 25above
##     Bach      3     98     102
##     HS       33     70      18
##     Mast      0      3      20
```

```
tally(~ EduLevel + Sex, data = Acad.new)
```

```
##         Sex
## EduLevel Female Male
##     Bach    183   20
##     HS      111   10
##     Mast     15    8
```

```
tally(~ Sex + Age, data = Acad.new)
```

```
##         Age
## Sex      18less 19to24 25above
##   Female     32    155     122
##   Male        4     16      18
```

Extra Data Wrangling

```
Acad.final <- Acad.new[!(Acad.new$Age == "18less"),]
Acad.final <- Acad.final[!(Acad.final$EduLevel == "Mast"),]
head(Acad.final)
```

```
## # A tibble: 6 x 22
##   Sex    Age    EduLevel Pleasure Depressed Sleep Tired Appetite Feeling
##   <chr>  <chr>  <chr>       <dbl>     <dbl> <dbl> <dbl>    <dbl>   <dbl>
## 1 Male   19to24 Bach            3         2     4     4        3       3
```

```
## 2 Male   19to24 Bach            2         1     2     1        1       1
## 3 Female 19to24 Bach            2         2     3     3        1       1
## 4 Female 19to24 Bach            3         1     4     3        1       2
## 5 Female 19to24 HS              3         4     2     2        4       3
## 6 Female 19to24 Bach            2         4     4     3        2       3
## # ... with 13 more variables: Concentrating <dbl>, Slow <dbl>, Thoughts <dbl>,
## #   Job <chr>, LivingSituation <chr>, Study <chr>, Electronics <chr>,
## #   SocialMedia <chr>, GPA <dbl>, Standing <chr>, DepressionScore <dbl>,
## #   DepressionLevel <chr>, Stand.Bin <dbl>
```

```
Acad.final$Job <- factor(Acad.final$Job, c("None","PartTime","FullTime"))
```

Modeling our Data

```
# Logistic Regression with All Interesting Predictors - FULL ADDITIVE

LR.Add.mod <- glm(Stand.Bin ~ DepressionScore + Age + EduLevel + Job +
                    LivingSituation + Study,
                  family = binomial(link = "logit"), data = Acad.final)
summary(LR.Add.mod)
```

```
##
## Call:
## glm(formula = Stand.Bin ~ DepressionScore + Age + EduLevel +
##     Job + LivingSituation + Study, family = binomial(link = "logit"),
##     data = Acad.final)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5996  -0.4462  -0.3932  -0.3253   2.5256
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -15.60105 1455.39766  -0.011    0.991
## DepressionScore             0.03535    0.03608   0.980    0.327
## Age25above                 -0.31396    0.50868  -0.617    0.537
## EduLevelHS                 -0.65422    0.57136  -1.145    0.252
## JobPartTime                -0.03571    0.61592  -0.058    0.954
## JobFullTime                -0.09215    0.55719  -0.165    0.869
## LivingSituationoffCampus   13.19898 1455.39772   0.009    0.993
## LivingSituationwParents    13.06202 1455.39770   0.009    0.993
## Study2to4hrs                0.01834    0.49584   0.037    0.970
## Study4above                -1.13848    1.07513  -1.059    0.290
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 160.38  on 287  degrees of freedom
## Residual deviance: 156.47  on 278  degrees of freedom
```

```
## AIC: 176.47
##
## Number of Fisher Scoring iterations: 14
```

Large p-values indicate we have too many predictors in our model. It is possible that some of our predictors share information, so we want to continue with some model selection analysis.

Model Selection

```
mod.null <- glm(Stand.Bin ~ 1, family = binomial(link = "logit"), data = Acad.final)

step(mod.null, Stand.Bin ~ (DepressionScore + Age + EduLevel + Job +
                   LivingSituation + Study)^2, direction = "forward")
```

```
## Start:  AIC=162.38
## Stand.Bin ~ 1
##
##                   Df Deviance    AIC
## <none>                 160.38 162.38
## + EduLevel         1   159.40 163.40
## + DepressionScore  1   159.75 163.75
## + Age              1   160.31 164.31
## + Study            2   158.97 164.97
## + LivingSituation  2   160.21 166.21
## + Job              2   160.35 166.35

##
## Call:  glm(formula = Stand.Bin ~ 1, family = binomial(link = "logit"),
##     data = Acad.final)
##
## Coefficients:
## (Intercept)
##      -2.444
##
## Degrees of Freedom: 287 Total (i.e. Null);  287 Residual
## Null Deviance:       160.4
## Residual Deviance: 160.4     AIC: 162.4
```

```
# dredge to assess equivalent models
mod.new <- glm(Stand.Bin ~ (DepressionScore + Age + EduLevel + Job +
                   LivingSituation + Study)^2,
               family = binomial(link = "logit"), data = Acad.final)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(mod.new)
```

```
##
## Call:
## glm(formula = Stand.Bin ~ (DepressionScore + Age + EduLevel +
```

```
##     Job + LivingSituation + Study)^2, family = binomial(link = "logit"),
##     data = Acad.final)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1309  -0.4308  -0.2663  -0.1296   2.9800
##
## Coefficients: (7 not defined because of singularities)
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             -2.387e+01  1.773e+04  -0.001   0.9989
## DepressionScore                          1.673e-01  1.209e-01   1.384   0.1663
## Age25above                               4.259e-01  2.226e+00   0.191   0.8482
## EduLevelHS                               7.589e-01  2.333e+00   0.325   0.7450
## JobPartTime                              3.442e+00  2.259e+00   1.523   0.1276
## JobFullTime                             -1.040e+00  2.029e+00  -0.513   0.6081
## LivingSituationoffCampus                 2.056e+01  1.773e+04   0.001   0.9991
## LivingSituationwParents                  1.963e+01  1.773e+04   0.001   0.9991
## Study2to4hrs                            -9.581e-01  2.136e+00  -0.449   0.6538
## Study4above                              1.263e+01  1.384e+01   0.913   0.3615
## DepressionScore:Age25above              -9.322e-02  1.128e-01  -0.826   0.4086
## DepressionScore:EduLevelHS              -3.896e-02  1.065e-01  -0.366   0.7146
## DepressionScore:JobPartTime             -2.354e-01  1.352e-01  -1.742   0.0815
## DepressionScore:JobFullTime              8.546e-02  1.138e-01   0.751   0.4528
## DepressionScore:LivingSituationoffCampus -3.150e-02  1.108e-01  -0.284   0.7761
## DepressionScore:LivingSituationwParents          NA         NA      NA       NA
## DepressionScore:Study2to4hrs            -1.082e-01  9.821e-02  -1.102   0.2705
## DepressionScore:Study4above             -7.746e-01  9.011e-01  -0.860   0.3900
## Age25above:EduLevelHS                    2.110e-01  1.837e+00   0.115   0.9085
## Age25above:JobPartTime                  -2.464e+00  2.021e+00  -1.219   0.2227
## Age25above:JobFullTime                   6.019e-01  1.731e+00   0.348   0.7281
## Age25above:LivingSituationoffCampus      2.789e-01  1.612e+00   0.173   0.8627
## Age25above:LivingSituationwParents               NA         NA      NA       NA
## Age25above:Study2to4hrs                  1.138e+00  1.438e+00   0.792   0.4285
## Age25above:Study4above                  -1.504e+01  3.696e+03  -0.004   0.9968
## EduLevelHS:JobPartTime                  -1.733e+00  1.960e+00  -0.884   0.3765
## EduLevelHS:JobFullTime                   2.080e-01  1.723e+00   0.121   0.9039
## EduLevelHS:LivingSituationoffCampus     -1.596e+00  1.453e+00  -1.098   0.2720
## EduLevelHS:LivingSituationwParents               NA         NA      NA       NA
## EduLevelHS:Study2to4hrs                  2.204e+00  1.415e+00   1.557   0.1195
## EduLevelHS:Study4above                  -2.533e+01  2.730e+03  -0.009   0.9926
## JobPartTime:LivingSituationoffCampus     2.389e-01  1.620e+00   0.148   0.8827
## JobFullTime:LivingSituationoffCampus    -1.176e+00  1.567e+00  -0.750   0.4531
## JobPartTime:LivingSituationwParents              NA         NA      NA       NA
## JobFullTime:LivingSituationwParents              NA         NA      NA       NA
## JobPartTime:Study2to4hrs                 8.411e-01  1.845e+00   0.456   0.6485
## JobFullTime:Study2to4hrs                 1.217e+00  1.775e+00   0.685   0.4931
## JobPartTime:Study4above                 -2.096e+01  3.931e+03  -0.005   0.9957
## JobFullTime:Study4above                 -1.887e+01  4.410e+03  -0.004   0.9966
## LivingSituationoffCampus:Study2to4hrs    5.591e-01  1.348e+00   0.415   0.6784
## LivingSituationwParents:Study2to4hrs             NA         NA      NA       NA
## LivingSituationoffCampus:Study4above    -2.052e+01  2.901e+03  -0.007   0.9944
## LivingSituationwParents:Study4above              NA         NA      NA       NA
##
## (Intercept)
```

```
## DepressionScore
## Age25above
## EduLevelHS
## JobPartTime
## JobFullTime
## LivingSituationoffCampus
## LivingSituationwParents
## Study2to4hrs
## Study4above
## DepressionScore:Age25above
## DepressionScore:EduLevelHS
## DepressionScore:JobPartTime              .
## DepressionScore:JobFullTime
## DepressionScore:LivingSituationoffCampus
## DepressionScore:LivingSituationwParents
## DepressionScore:Study2to4hrs
## DepressionScore:Study4above
## Age25above:EduLevelHS
## Age25above:JobPartTime
## Age25above:JobFullTime
## Age25above:LivingSituationoffCampus
## Age25above:LivingSituationwParents
## Age25above:Study2to4hrs
## Age25above:Study4above
## EduLevelHS:JobPartTime
## EduLevelHS:JobFullTime
## EduLevelHS:LivingSituationoffCampus
## EduLevelHS:LivingSituationwParents
## EduLevelHS:Study2to4hrs
## EduLevelHS:Study4above
## JobPartTime:LivingSituationoffCampus
## JobFullTime:LivingSituationoffCampus
## JobPartTime:LivingSituationwParents
## JobFullTime:LivingSituationwParents
## JobPartTime:Study2to4hrs
## JobFullTime:Study2to4hrs
## JobPartTime:Study4above
## JobFullTime:Study4above
## LivingSituationoffCampus:Study2to4hrs
## LivingSituationwParents:Study2to4hrs
## LivingSituationoffCampus:Study4above
## LivingSituationwParents:Study4above
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 160.38  on 287  degrees of freedom
## Residual deviance: 129.42  on 252  degrees of freedom
## AIC: 201.42
##
## Number of Fisher Scoring iterations: 19
```

```
mod.new2 <- glm(Stand.Bin ~ DepressionScore * Age + DepressionScore * Job +
                  Age * Job + EduLevel * LivingSituation + EduLevel * Study +
                Job * Study + DepressionScore + Age + EduLevel + Job +
                  LivingSituation + Study,
              family = binomial(link = "logit"), data = Acad.final)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
library(MuMIn)
options(na.action = "na.fail")
suppressWarnings(dredgeoutput <- dredge(mod.new2, rank = "AIC", extra = "R^2",
                                        fixed = "DepressionScore"))
```

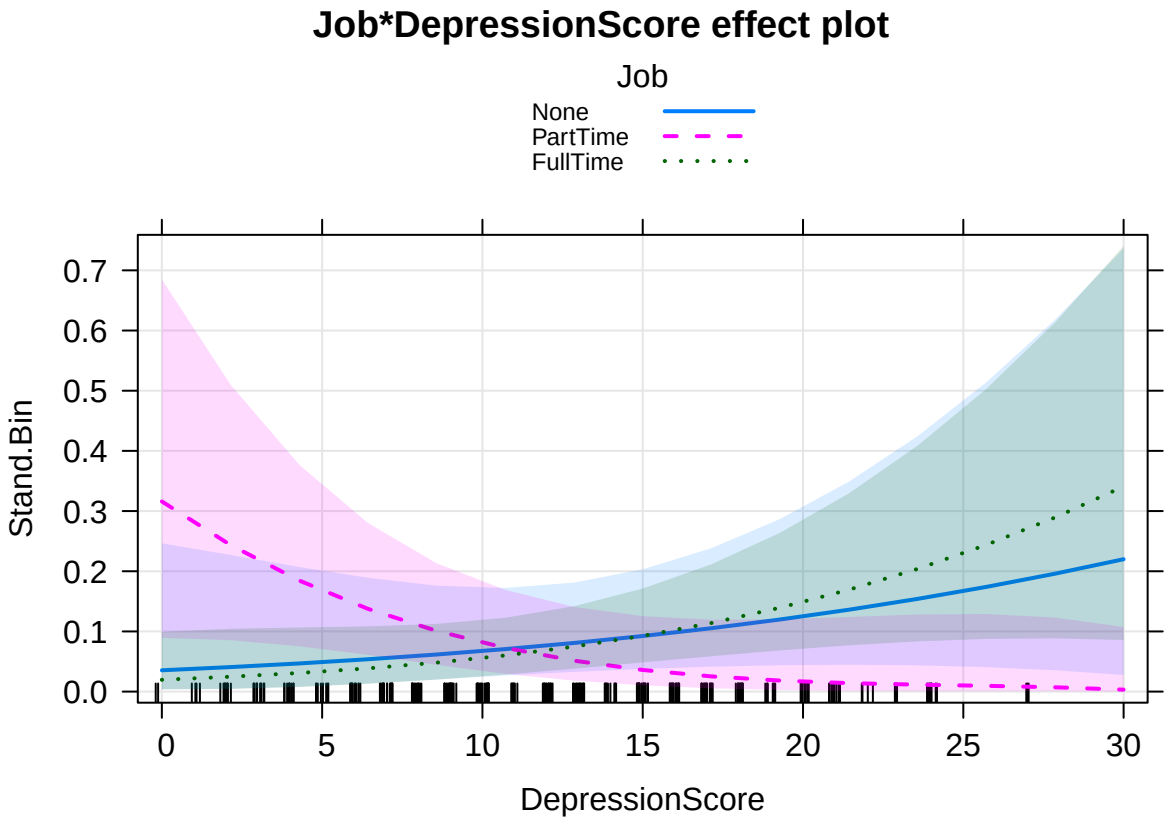
```
## Fixed terms are "DepressionScore" and "(Intercept)"
```

```
head(dredgeoutput)
```

```
## Global model call: glm(formula = Stand.Bin ~ DepressionScore * Age + DepressionScore *
##     Job + Age * Job + EduLevel * LivingSituation + EduLevel *
##     Study + Job * Study + DepressionScore + Age + EduLevel +
##     Job + LivingSituation + Study, family = binomial(link = "logit"),
##     data = Acad.final)
## ---
## Model selection table
##      (Int) Age     DpS EdL Job DpS:Job      R^2 df  logLik   AIC delta weight
## 133 -3.305     0.06797       +       + 0.035010  6 -75.057 162.1  0.00  0.340
## 135 -3.236     0.07056   +   +       + 0.037510  7 -74.683 163.4  1.25  0.182
## 1   -2.795     0.02772                 0.002159  2 -79.877 163.8  1.64  0.150
## 134 -3.233   + 0.06693       +       + 0.035830  7 -74.934 163.9  1.75  0.141
## 3   -2.703     0.03173   +             0.006125  3 -79.303 164.6  2.49  0.098
## 136 -3.102   + 0.06882   +   +       + 0.039400  8 -74.399 164.8  2.69  0.089
## Models ranked by AIC(x)
```

```
# Effects Plots

## Fitting Interaction Models
int.mod1 <- glm(Stand.Bin ~ Job * DepressionScore, family = binomial,
              data = Acad.final)

plot(allEffects(int.mod1, resid = T), grid = T, type = "response")
```


```
plot(allEffects(int.mod1), grid = T, multiline=T, ci.style="bands", lty=c(1:3),
     type = "response")
```



Final Model

```
# Need to include DepressionScore variable in Final Model because
# it pertains to our primary research question

mod.final <- glm(Stand.Bin ~ DepressionScore * Job,
                 family = binomial, data = Acad.final)
summary(mod.final)
```

```
##
## Call:
## glm(formula = Stand.Bin ~ DepressionScore * Job, family = binomial,
##     data = Acad.final)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.8129  -0.4568  -0.3510  -0.2586   2.5651
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -3.30492    1.11604  -2.961  0.00306 **
## DepressionScore             0.06797    0.07059   0.963  0.33562
## JobPartTime                 2.53250    1.36754   1.852  0.06405 .
## JobFullTime                -0.60706    1.41569  -0.429  0.66806
```

```
## DepressionScore:JobPartTime -0.23327    0.10783  -2.163  0.03052 *
## DepressionScore:JobFullTime  0.04044    0.08890   0.455  0.64918
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 160.38  on 287  degrees of freedom
## Residual deviance: 150.11  on 282  degrees of freedom
## AIC: 162.11
##
## Number of Fisher Scoring iterations: 6
```

Final Fitted Model:

$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = -3.30492 + 0.06797x_D + 2.53250 * I_{Job=PartTime} - 0.60706 * I_{Job=FullTime} - 0.23327(x_D * I_{Job=PartTime}) + 0.04044(x_D * I_{Job=FullTime})$

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 questionnaire which is recorded as an individual's Depression Index Score in points * $I_{Job=PartTime}$ is an indicator variable that is 1 when the individual is working a part time job and 0 otherwise * $I_{Job=FullTime}$ is an indicator variable that is 1 when the individual is working a full time job and 0 otherwise

```
# Exponentiate the Coefficients to Interpret

exp(mod.final$coefficients)
```

```
##                 (Intercept)             DepressionScore
##                  0.03670216                  1.07033487
##                 JobPartTime                 JobFullTime
##                 12.58493415                  0.54495135
## DepressionScore:JobPartTime DepressionScore:JobFullTime
##                  0.79193628                  1.04127021
```

```
# 95% Confidence Interval for these Coefficients

confint(mod.final)
```

```
## Waiting for profiling to be done...
```

```
##                                   2.5 %      97.5 %
## (Intercept)                 -5.84110531 -1.37406321
## DepressionScore             -0.07050762  0.21331761
## JobPartTime                 -0.03716251  5.44322707
## JobFullTime                 -3.34602096  2.34900743
## DepressionScore:JobPartTime -0.45977168 -0.03022012
## DepressionScore:JobFullTime -0.13653375  0.21671969
```

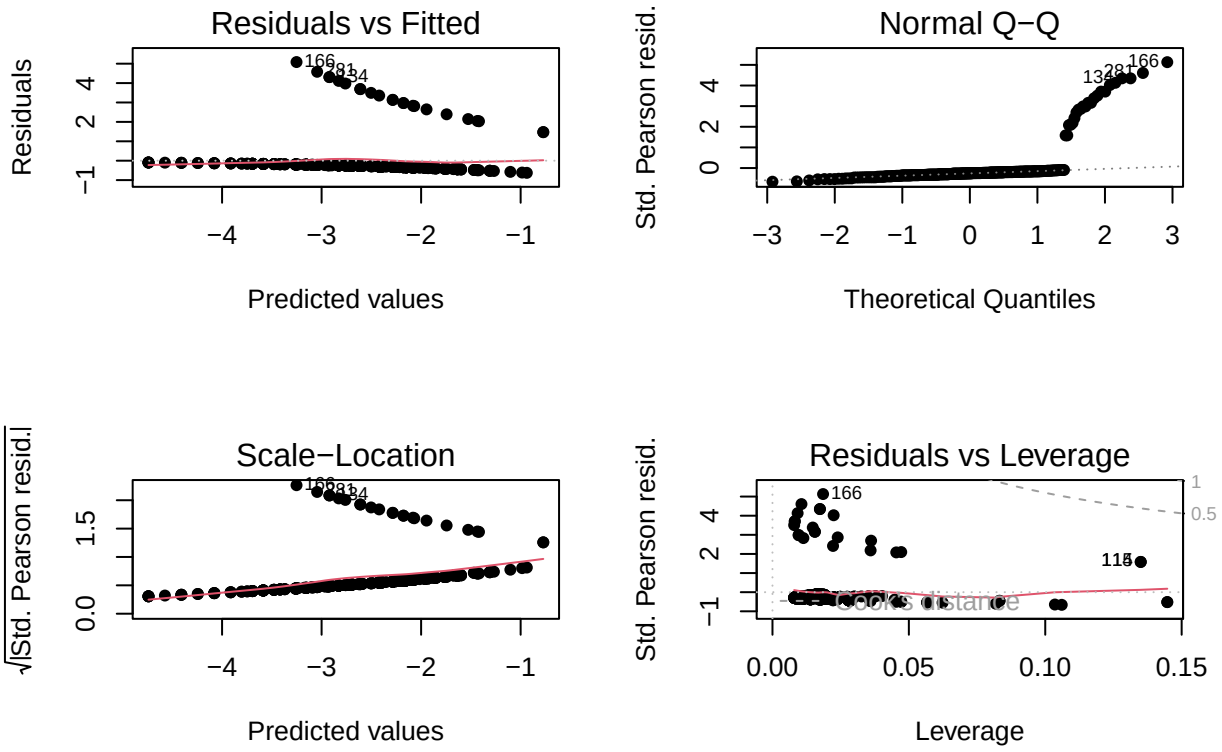
```
exp(confint(mod.final))
```

```
## Waiting for profiling to be done...
```

```
##                                    2.5 %      97.5 %
## (Intercept)                   0.002905629   0.2530766
## DepressionScore               0.931920639   1.2377777
## JobPartTime                   0.963519540 231.1870373
## JobFullTime                   0.035224234  10.4751672
## DepressionScore:JobPartTime   0.631427794   0.9702319
## DepressionScore:JobFullTime   0.872376878   1.2419959
```

Model Diagnostics Array

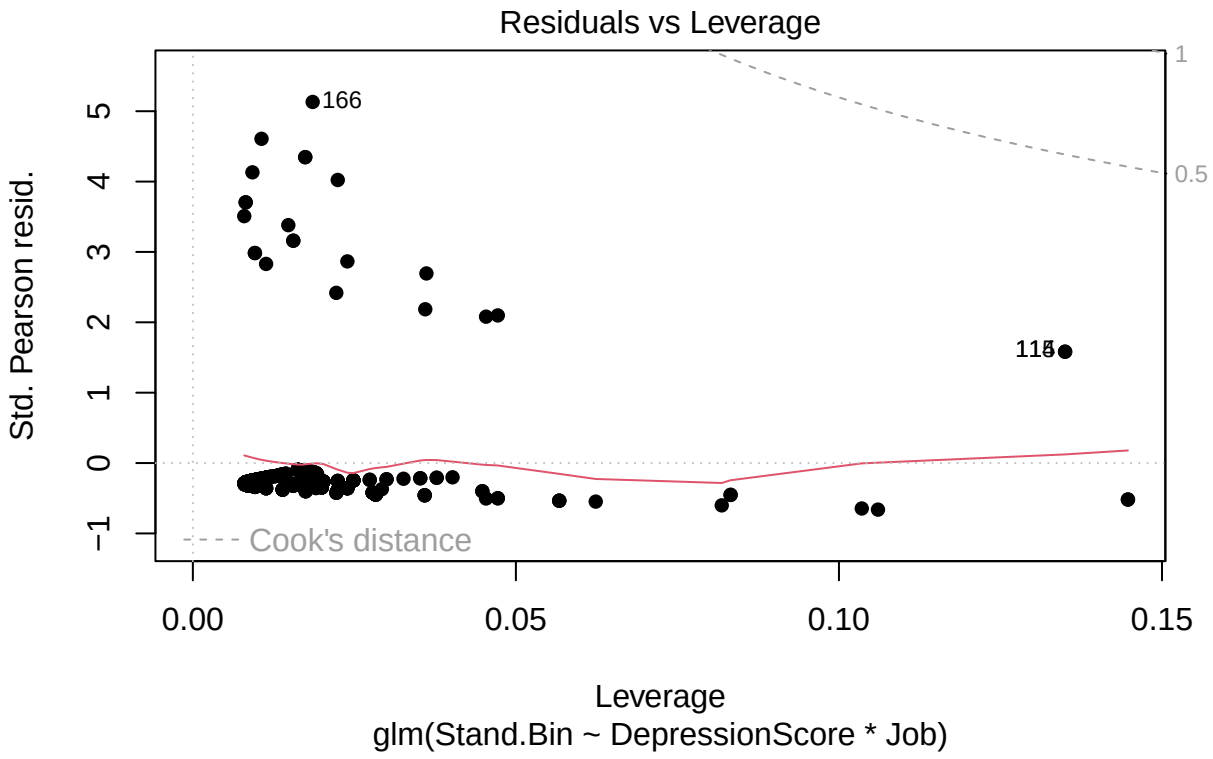
```
# Standard Model Diagnostics Array

par(mfrow=c(2,2))
plot(mod.final, pch=16)
```



Cook's Distance is a measure of how much the model changes if you took the point

all diagnostic plots in appendix

```
plot(mod.final, 5, pch=16)
```



Assessing Model Fit

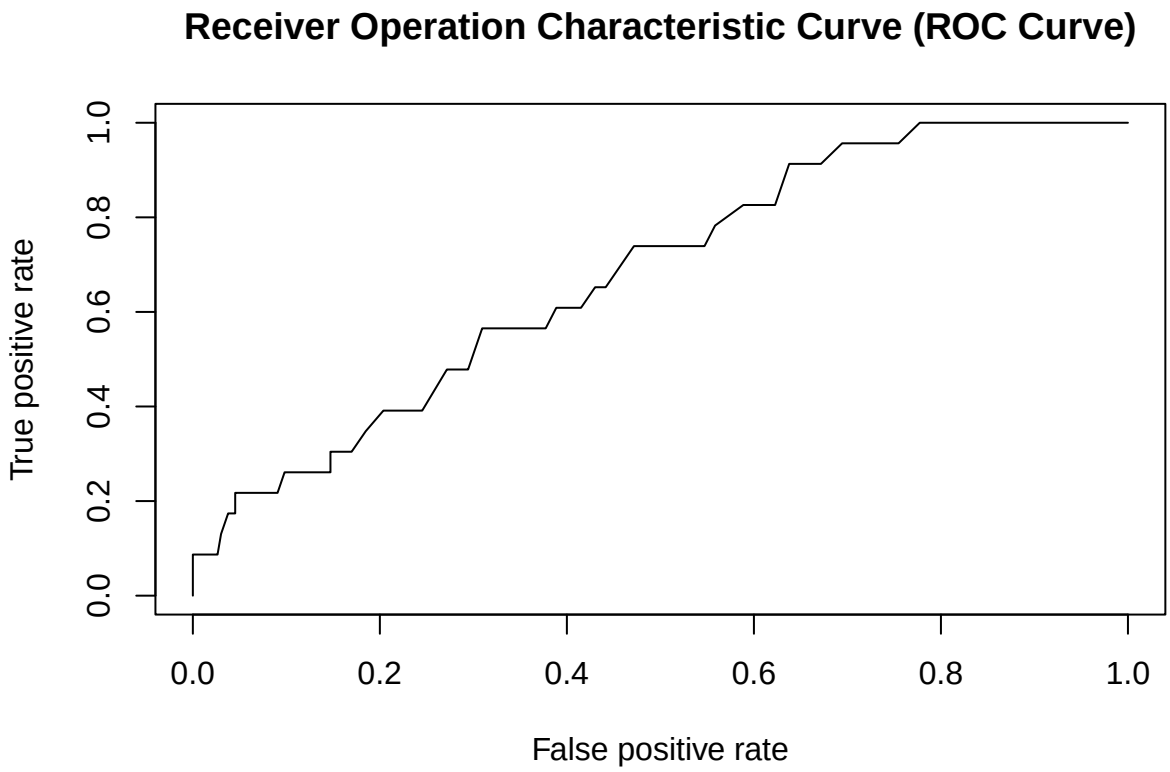
```
# Goodness of Fit for binary data: Hosmer-Lemeshow Test

obs <- Acad.final$Stand.Bin
expected <- fitted(mod.final)
hoslem.test(obs, expected, g = 10) # g = # of groups
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  obs, expected
## X-squared = 5.135, df = 8, p-value = 0.7431
```

Prediction Ability

```
# Generating an ROC Curve

probs <- fitted(mod.final)
```

```
pred.obj <- prediction(probs, Acad.final$Stand.Bin)

plot(performance(pred.obj, "tpr", "fpr"),
     main = "Receiver Operation Characteristic Curve (ROC Curve)")
```




```
# Area under the ROC Curve

area <- performance(pred.obj, "auc")
auc <- as.numeric(area@y.values)
auc
```

```
## [1] 0.6784249
```

Citations

```
citation("yarrr")
```

```
##
## To cite package 'yarrr' in publications use:
```

```
##
##   Phillips N (2017). _yarrr: A Companion to the e-Book "YaRrr!: The
##   Pirate's Guide to R"_. R package version 0.1.5,
##   <https://CRAN.R-project.org/package=yarrr>.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {yarrr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R"},
##     author = {Nathaniel Phillips},
##     year = {2017},
##     note = {R package version 0.1.5},
##     url = {https://CRAN.R-project.org/package=yarrr},
##   }
```

```
citation("ggplot2")
```

```
##
## To cite ggplot2 in publications, please use:
##
##   H. Wickham. ggplot2: Elegant Graphics for Data Analysis.
##   Springer-Verlag New York, 2016.
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     author = {Hadley Wickham},
##     title = {ggplot2: Elegant Graphics for Data Analysis},
##     publisher = {Springer-Verlag New York},
##     year = {2016},
##     isbn = {978-3-319-24277-4},
##     url = {https://ggplot2.tidyverse.org},
##   }
```

```
citation("multcomp")
```

```
##
## Please cite the multcomp package by the following reference:
##
##   Torsten Hothorn, Frank Bretz and Peter Westfall (2008). Simultaneous
##   Inference in General Parametric Models. Biometrical Journal 50(3),
##   346--363.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {Simultaneous Inference in General Parametric Models},
##     author = {Torsten Hothorn and Frank Bretz and Peter Westfall},
##     journal = {Biometrical Journal},
##     year = {2008},
##     volume = {50},
##     number = {3},
##     pages = {346--363},
##   }
```

```r
citation("effects")
```

```
##
## To cite effects in publications use:
##
##   John Fox and Sanford Weisberg (2019). An R Companion to Applied
##   Regression, 3rd Edition. Thousand Oaks, CA
##   <https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html>
##
## For predictor effects or partial residuals also cite:
##
##   John Fox, Sanford Weisberg (2018). Visualizing Fit and Lack of Fit in
##   Complex Regression Models with Predictor Effect Plots and Partial
##   Residuals. Journal of Statistical Software, 87(9), 1-27. doi
##   10.18637/jss.v087.i09
##
## For generalized linear models also cite:
##
##   John Fox (2003). Effect Displays in R for Generalised Linear Models.
##   Journal of Statistical Software, 8(15), 1-27. doi
##   10.18637/jss.v008.i15
##
## For usage in multinomial and proportional-odds logit models also cite:
##
##   John Fox, Jangman Hong (2009). Effect Displays in R for Multinomial
##   and Proportional-Odds Logit Models: Extensions to the effects
##   Package. Journal of Statistical Software, 32(1), 1-24. doi
##   10.18637/jss.v032.i01
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```

```r
citation("gtsummary")
```

```
##
## To cite gtsummary in publications use:
##
##   Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. Reproducible
##   summary tables with the gtsummary package. The R Journal
##   2021;13:570-80. https://doi.org/10.32614/RJ-2021-053.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{gtsummary,
##     author = {Daniel D. Sjoberg and Karissa Whiting and Michael Curry and Jessica A. Lavery and Josep
##     title = {Reproducible Summary Tables with the gtsummary Package},
##     journal = {{The R Journal}},
##     year = {2021},
##     url = {https://doi.org/10.32614/RJ-2021-053},
##     doi = {10.32614/RJ-2021-053},
##     volume = {13},
##     issue = {1},
```

```
##     pages = {570-580},
##   }
```

```
citation("psych")
```

```
##
## To cite the psych package in publications use:
##
##   Revelle, W. (2022) psych: Procedures for Personality and
##   Psychological Research, Northwestern University, Evanston, Illinois,
##   USA, https://CRAN.R-project.org/package=psych Version = 2.2.3,.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {psych: Procedures for Psychological, Psychometric, and Personality Research},
##     author = {William Revelle},
##     organization = { Northwestern University},
##     address = { Evanston, Illinois},
##     year = {2022},
##     note = {R package version 2.2.3},
##     url = {https://CRAN.R-project.org/package=psych},
##   }
```

```
citation("ROCR")
```

```
##
## To cite ROCR in publications use:
##
##   Sing T, Sander O, Beerenwinkel N, Lengauer T (2005). "ROCR:
##   visualizing classifier performance in R." _Bioinformatics_, *21*(20),
##   7881. <http://rocr.bioinf.mpi-sb.mpg.de>.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     entry = {article},
##     title = {ROCR: visualizing classifier performance in R},
##     author = {T. Sing and O. Sander and N. Beerenwinkel and T. Lengauer},
##     year = {2005},
##     journal = {Bioinformatics},
##     volume = {21},
##     number = {20},
##     pages = {7881},
##     url = {http://rocr.bioinf.mpi-sb.mpg.de},
##   }
##
## We have invested a lot of time and effort in creating ROCR, please cite
## it when using it for data analysis.
```

```
citation("mosaic")
```

```
##
## To cite mosaic in publications, please use:
##
##   R. Pruim, D. T. Kaplan and N. J. Horton. The mosaic Package: Helping
##   Students to 'Think with Data' Using R (2017). The R Journal,
##   9(1):77-102.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     author = {Randall Pruim and Daniel T Kaplan and Nicholas J Horton},
##     title = {The mosaic Package: Helping Students to 'Think with Data' Using R},
##     journal = {The R Journal},
##     volume = {9},
##     number = {1},
##     pages = {77--102},
##     year = {2017},
##     url = {https://journal.r-project.org/archive/2017/RJ-2017-024/index.html},
##   }
```

```
citation("ResourceSelection")  # For hoslem.test function
```

```
##
## To cite package 'ResourceSelection' in publications use:
##
##   Lele SR, Keim JL, Solymos P (2019). _ResourceSelection: Resource
##   Selection (Probability) Functions for Use-Availability Data_. R
##   package version 0.3-5,
##   <https://CRAN.R-project.org/package=ResourceSelection>.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {ResourceSelection: Resource Selection (Probability) Functions for Use-Availability
## Data},
##     author = {Subhash R. Lele and Jonah L. Keim and Peter Solymos},
##     year = {2019},
##     note = {R package version 0.3-5},
##     url = {https://CRAN.R-project.org/package=ResourceSelection},
##   }
```

```
citation("MuMIn")
```

```
##
## To cite package 'MuMIn' in publications use:
##
##   Bartoń K (2022). _MuMIn: Multi-Model Inference_. R package version
##   1.46.0, <https://CRAN.R-project.org/package=MuMIn>.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {MuMIn: Multi-Model Inference},
```

```
##     author = {Kamil Bartoń},
##     year = {2022},
##     note = {R package version 1.46.0},
##     url = {https://CRAN.R-project.org/package=MuMIn},
##   }
##
## ATTENTION: This citation information has been auto-generated from the
## package DESCRIPTION file and may need manual editing, see
## 'help("citation")'.
```