Law School Report

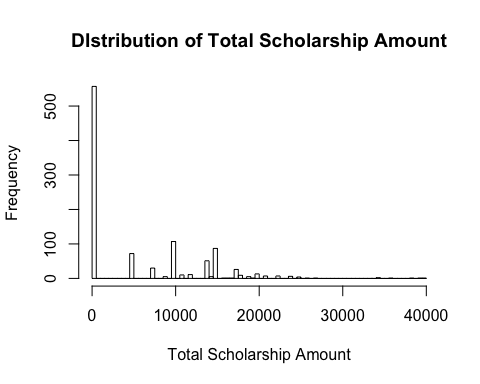
4/11/2019

## Read Data

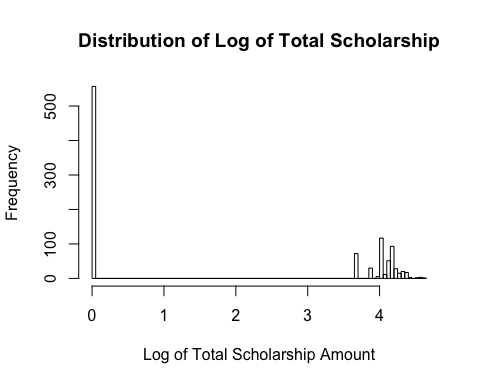
library(readxl)  
library(ISLR)  
library(devtools)  
  
# Read the data file  
ls\_data <- read\_excel("~/Desktop/Final v2 for Stats.xlsx")  
# Encode for Emplyment column: 0 for 'N', and 1 for 'Y'  
ls\_data$Employed = factor( ls\_data$Employed,  
 levels = c('N', 'Y'),  
 labels = c('0', '1'))  
# For Bar Pass, 'Passed' was transformed to 1, all the rest ('FAILED', 'DID NOT TAKE', 'UNKNOWN' and '0') were transformed to 0  
ls\_data$Bar\_Status = factor(ls\_data$Bar\_Status,  
 levels = c('PASSED', 'FAILED', 'DID NOT TAKE', 'UNKNOWN', '0'),  
 labels = c('1', '0', '0', '0', '0'))

Plot the histgram for distribution of scholarship amount

Total\_Scholarship\_Amt = ls\_data$Total\_Scholarship\_Amt  
Log\_Total\_Scholarship\_Amt = log10(Total\_Scholarship\_Amt + 1)  
hist(Total\_Scholarship\_Amt, breaks = 100, xlab = "Total Scholarship Amount", main = "DIstribution of Total Scholarship Amount")



hist(Log\_Total\_Scholarship\_Amt, breaks = 100, xlab = "Log of Total Scholarship Amount", main = "Distribution of Log of Total Scholarship")



the statistics of each column of the preprocessed dataset:

# ls\_data$Total\_Scholarship\_Amt = (ls\_data$Total\_Scholarship\_Amt)  
summary(ls\_data)

## Matriculant\_Year ID Employed Bar\_Status Fall\_1L   
## Min. :2011 Min. : 1.0 0:194 1:814 Min. :1.527   
## 1st Qu.:2012 1st Qu.: 300.8 1:830 0:210 1st Qu.:3.060   
## Median :2013 Median : 598.5 Median :3.327   
## Mean :2013 Mean : 598.4 Mean :3.308   
## 3rd Qu.:2014 3rd Qu.: 891.2 3rd Qu.:3.593   
## Max. :2015 Max. :1182.0 Max. :4.240   
##   
## Spring\_1L Fall\_2L Spring\_2L Fall\_3L   
## Min. :2.069 Min. :2.115 Min. :2.152 Min. :2.268   
## 1st Qu.:3.068 1st Qu.:3.103 1st Qu.:3.131 1st Qu.:3.149   
## Median :3.327 Median :3.353 Median :3.370 Median :3.375   
## Mean :3.308 Mean :3.328 Mean :3.344 Mean :3.357   
## 3rd Qu.:3.562 3rd Qu.:3.570 3rd Qu.:3.574 3rd Qu.:3.575   
## Max. :4.160 Max. :4.130 Max. :4.100 Max. :4.103   
##   
## Spring\_3L Total\_Scholarship\_Amt Residency\_Long   
## Min. :2.263 Min. : 0 Length:1024   
## 1st Qu.:3.167 1st Qu.: 0 Class :character   
## Median :3.389 Median : 0 Mode :character   
## Mean :3.368 Mean : 5759   
## 3rd Qu.:3.586 3rd Qu.:10000   
## Max. :4.076 Max. :40000   
##   
## Current\_age\_of\_applicant Gender\_Applicant High\_LSAT\_Score  
## Min. :24.00 Length:1024 Min. :146.0   
## 1st Qu.:27.00 Class :character 1st Qu.:159.0   
## Median :29.00 Mode :character Median :162.0   
## Mean :29.48 Mean :161.3   
## 3rd Qu.:31.00 3rd Qu.:164.0   
## Max. :62.00 Max. :177.0   
##   
## Cumulative\_GPA  
## Min. :2.09   
## 1st Qu.:3.32   
## Median :3.51   
## Mean :3.49   
## 3rd Qu.:3.68   
## Max. :4.22   
## NA's :7

## PCA Analysis

PCA for 6 Semesters:

six\_semesters.pca <- prcomp(ls\_data[,c(5:10)], center = TRUE, scale. = TRUE)  
summary(six\_semesters.pca)

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6  
## Standard deviation 2.3865 0.46080 0.24234 0.14509 0.09110 0.06415  
## Proportion of Variance 0.9492 0.03539 0.00979 0.00351 0.00138 0.00069  
## Cumulative Proportion 0.9492 0.98463 0.99442 0.99793 0.99931 1.00000

str(six\_semesters.pca)

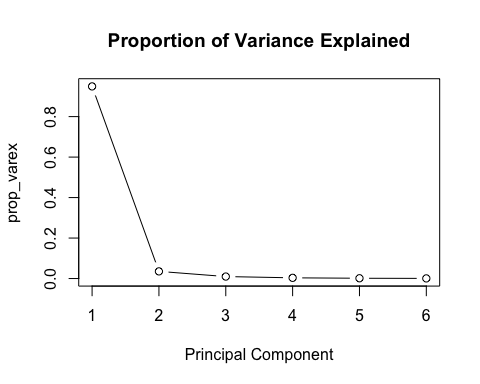
## List of 5  
## $ sdev : num [1:6] 2.3865 0.4608 0.2423 0.1451 0.0911 ...  
## $ rotation: num [1:6, 1:6] 0.388 0.409 0.415 0.415 0.413 ...  
## ..- attr(\*, "dimnames")=List of 2  
## .. ..$ : chr [1:6] "Fall\_1L" "Spring\_1L" "Fall\_2L" "Spring\_2L" ...  
## .. ..$ : chr [1:6] "PC1" "PC2" "PC3" "PC4" ...  
## $ center : Named num [1:6] 3.31 3.31 3.33 3.34 3.36 ...  
## ..- attr(\*, "names")= chr [1:6] "Fall\_1L" "Spring\_1L" "Fall\_2L" "Spring\_2L" ...  
## $ scale : Named num [1:6] 0.377 0.351 0.33 0.318 0.308 ...  
## ..- attr(\*, "names")= chr [1:6] "Fall\_1L" "Spring\_1L" "Fall\_2L" "Spring\_2L" ...  
## $ x : num [1:1024, 1:6] -3.113 -3.605 -2.244 -3.928 0.224 ...  
## ..- attr(\*, "dimnames")=List of 2  
## .. ..$ : NULL  
## .. ..$ : chr [1:6] "PC1" "PC2" "PC3" "PC4" ...  
## - attr(\*, "class")= chr "prcomp"

six\_semesters.pca$rotation

## PC1 PC2 PC3 PC4 PC5  
## Fall\_1L 0.3876737 -0.78424946 -0.47461062 -0.0961750 0.009824168  
## Spring\_1L 0.4089143 -0.26858331 0.66381558 0.5462685 -0.145877781  
## Fall\_2L 0.4152271 0.06333156 0.35995741 -0.5484640 0.620788703  
## Spring\_2L 0.4150650 0.22282538 0.03882784 -0.4347948 -0.648917647  
## Fall\_3L 0.4127223 0.33415693 -0.22737533 0.1041014 -0.198117481  
## Spring\_3L 0.4092260 0.38404804 -0.38899506 0.4377723 0.364554380  
## PC6  
## Fall\_1L -0.007776313  
## Spring\_1L 0.017646108  
## Fall\_2L -0.088401210  
## Spring\_2L 0.407948052  
## Fall\_3L -0.784991890  
## Spring\_3L 0.457361443

#compute standard deviation of each principal component  
std\_dev <- six\_semesters.pca$sdev  
pr\_var <- std\_dev^2  
#proportion of variance explained  
prop\_varex <- pr\_var/sum(pr\_var)

plot(prop\_varex, xlab = "Principal Component",  
 main = "Proportion of Variance Explained",  
 type = "b")



# apply the first 2 factors of PCA for analysis for 6 semsters: six\_semesters.pca$x[,1] + six\_semesters.pca$x[,2]

## Logistic Regression Analysis

# 1. Without controvariables

1. Undergraduate GPA & first LSAT

glm\_emply\_1 = glm(Employed ~ Cumulative\_GPA + High\_LSAT\_Score, family = binomial, data = ls\_data)   
summary(glm\_emply\_1)

##   
## Call:  
## glm(formula = Employed ~ Cumulative\_GPA + High\_LSAT\_Score, family = binomial,   
## data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.1826 0.4879 0.5958 0.6810 1.0316   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -12.51208 3.30031 -3.791 0.000150 \*\*\*  
## Cumulative\_GPA 0.98335 0.28856 3.408 0.000655 \*\*\*  
## High\_LSAT\_Score 0.06560 0.01918 3.420 0.000626 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 985.40 on 1016 degrees of freedom  
## Residual deviance: 962.91 on 1014 degrees of freedom  
## (7 observations deleted due to missingness)  
## AIC: 968.91  
##   
## Number of Fisher Scoring iterations: 4

glm\_bp\_1 = glm(Bar\_Status ~ Cumulative\_GPA + High\_LSAT\_Score, family = binomial, data = ls\_data)  
summary(glm\_bp\_1)

##   
## Call:  
## glm(formula = Bar\_Status ~ Cumulative\_GPA + High\_LSAT\_Score,   
## family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1368 -0.7118 -0.6103 -0.4813 2.2499   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 15.17864 3.23082 4.698 2.63e-06 \*\*\*  
## Cumulative\_GPA -0.96508 0.28185 -3.424 0.000617 \*\*\*  
## High\_LSAT\_Score -0.08192 0.01875 -4.370 1.24e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1033.2 on 1016 degrees of freedom  
## Residual deviance: 1003.2 on 1014 degrees of freedom  
## (7 observations deleted due to missingness)  
## AIC: 1009.2  
##   
## Number of Fisher Scoring iterations: 4

1. Year 1 only

glm\_emply\_2 = glm(Employed ~ Fall\_1L + Spring\_1L, family = binomial, data = ls\_data)  
summary(glm\_emply\_2)

##   
## Call:  
## glm(formula = Employed ~ Fall\_1L + Spring\_1L, family = binomial,   
## data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5029 0.3146 0.4653 0.6577 1.7117   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.6174 0.8144 -8.125 4.47e-16 \*\*\*  
## Fall\_1L -1.2963 0.5817 -2.229 0.0258 \*   
## Spring\_1L 3.8054 0.6417 5.931 3.02e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 994.16 on 1023 degrees of freedom  
## Residual deviance: 877.56 on 1021 degrees of freedom  
## AIC: 883.56  
##   
## Number of Fisher Scoring iterations: 5

glm\_bp\_2 = glm(Bar\_Status ~ Fall\_1L + Spring\_1L, family = binomial, data = ls\_data)  
summary(glm\_bp\_2)

##   
## Call:  
## glm(formula = Bar\_Status ~ Fall\_1L + Spring\_1L, family = binomial,   
## data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2743 -0.6239 -0.4021 -0.2157 2.6420   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 10.2329 0.9282 11.025 < 2e-16 \*\*\*  
## Fall\_1L 0.4280 0.5892 0.726 0.468   
## Spring\_1L -4.0496 0.6582 -6.153 7.62e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1039.1 on 1023 degrees of freedom  
## Residual deviance: 829.5 on 1021 degrees of freedom  
## AIC: 835.5  
##   
## Number of Fisher Scoring iterations: 5

1. Year 2 only

glm\_emply\_3 = glm(Employed ~ Fall\_2L + Spring\_2L, family = binomial, data = ls\_data)  
summary(glm\_emply\_3)

##   
## Call:  
## glm(formula = Employed ~ Fall\_2L + Spring\_2L, family = binomial,   
## data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4905 0.2949 0.4644 0.6570 1.6959   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -8.3181 0.9512 -8.744 <2e-16 \*\*\*  
## Fall\_2L -0.4663 1.5227 -0.306 0.7594   
## Spring\_2L 3.4600 1.5888 2.178 0.0294 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 994.16 on 1023 degrees of freedom  
## Residual deviance: 868.50 on 1021 degrees of freedom  
## AIC: 874.5  
##   
## Number of Fisher Scoring iterations: 5

glm\_bp\_3 = glm(Bar\_Status ~ Fall\_2L + Spring\_2L, family = binomial, data = ls\_data)  
summary(glm\_bp\_3)

##   
## Call:  
## glm(formula = Bar\_Status ~ Fall\_2L + Spring\_2L, family = binomial,   
## data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.1801 -0.6150 -0.3627 -0.1715 2.8513   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 13.946 1.169 11.935 < 2e-16 \*\*\*  
## Fall\_2L 2.424 1.589 1.525 0.127   
## Spring\_2L -7.137 1.688 -4.227 2.37e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1039.08 on 1023 degrees of freedom  
## Residual deviance: 783.39 on 1021 degrees of freedom  
## AIC: 789.39  
##   
## Number of Fisher Scoring iterations: 5

1. Undergraduate GPA & first LSAT & Year one

glm\_emply\_4 = glm(Employed ~ Cumulative\_GPA + High\_LSAT\_Score + Fall\_1L + Spring\_1L, family = binomial, data = ls\_data)   
summary(glm\_emply\_4)

##   
## Call:  
## glm(formula = Employed ~ Cumulative\_GPA + High\_LSAT\_Score + Fall\_1L +   
## Spring\_1L, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5015 0.3081 0.4639 0.6626 1.7219   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.98464 3.50555 -1.422 0.1550   
## Cumulative\_GPA 0.33909 0.32025 1.059 0.2897   
## High\_LSAT\_Score -0.01803 0.02173 -0.830 0.4067   
## Fall\_1L -1.35110 0.58504 -2.309 0.0209 \*   
## Spring\_1L 3.89137 0.65395 5.951 2.67e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 985.40 on 1016 degrees of freedom  
## Residual deviance: 868.32 on 1012 degrees of freedom  
## (7 observations deleted due to missingness)  
## AIC: 878.32  
##   
## Number of Fisher Scoring iterations: 5

glm\_bp\_4 = glm(Bar\_Status ~ Cumulative\_GPA + High\_LSAT\_Score + Fall\_1L + Spring\_1L, family = binomial, data = ls\_data)  
summary(glm\_bp\_4)

##   
## Call:  
## glm(formula = Bar\_Status ~ Cumulative\_GPA + High\_LSAT\_Score +   
## Fall\_1L + Spring\_1L, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.3869 -0.6239 -0.3978 -0.2117 2.6442   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 5.18729 3.59752 1.442 0.1493   
## Cumulative\_GPA -0.05463 0.33000 -0.166 0.8685   
## High\_LSAT\_Score 0.03679 0.02229 1.651 0.0988 .   
## Fall\_1L 0.31263 0.59294 0.527 0.5980   
## Spring\_1L -4.14793 0.67110 -6.181 6.38e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1033.15 on 1016 degrees of freedom  
## Residual deviance: 820.49 on 1012 degrees of freedom  
## (7 observations deleted due to missingness)  
## AIC: 830.49  
##   
## Number of Fisher Scoring iterations: 5

1. All 6 semesters

glm\_emply\_5 = glm(Employed ~ six\_semesters.pca$x[,1] + six\_semesters.pca$x[,2], family = binomial, data = ls\_data)  
summary(glm\_emply\_5)

##   
## Call:  
## glm(formula = Employed ~ six\_semesters.pca$x[, 1] + six\_semesters.pca$x[,   
## 2], family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5362 0.2919 0.4613 0.6593 1.6952   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.71896 0.09724 17.678 <2e-16 \*\*\*  
## six\_semesters.pca$x[, 1] 0.39696 0.03919 10.130 <2e-16 \*\*\*  
## six\_semesters.pca$x[, 2] 0.57001 0.17846 3.194 0.0014 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 994.16 on 1023 degrees of freedom  
## Residual deviance: 861.07 on 1021 degrees of freedom  
## AIC: 867.07  
##   
## Number of Fisher Scoring iterations: 5

glm\_bp\_5 = glm(Bar\_Status ~ six\_semesters.pca$x[,1] + six\_semesters.pca$x[,2], family = binomial, data = ls\_data)  
summary(glm\_bp\_5)

##   
## Call:  
## glm(formula = Bar\_Status ~ six\_semesters.pca$x[, 1] + six\_semesters.pca$x[,   
## 2], family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.1355 -0.6017 -0.3551 -0.1614 2.8910   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.88967 0.11291 -16.737 < 2e-16 \*\*\*  
## six\_semesters.pca$x[, 1] -0.62438 0.04864 -12.837 < 2e-16 \*\*\*  
## six\_semesters.pca$x[, 2] -0.82733 0.19123 -4.326 1.52e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1039.08 on 1023 degrees of freedom  
## Residual deviance: 771.33 on 1021 degrees of freedom  
## AIC: 777.33  
##   
## Number of Fisher Scoring iterations: 5

# 2. With control variables: Current\_age\_of\_applicant, Matriculant\_Year, Total\_Scholarship\_Amt, Gender\_Applicant, Residency\_Long

1. Undergraduate GPA & first LSAT

glm\_emply\_cv\_1 = glm(Employed ~ Cumulative\_GPA + High\_LSAT\_Score + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)   
summary(glm\_emply\_cv\_1)

##   
## Call:  
## glm(formula = Employed ~ Cumulative\_GPA + High\_LSAT\_Score + Current\_age\_of\_applicant +   
## Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant +   
## Residency\_Long, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2088 0.4854 0.5845 0.6726 1.4420   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.121e+02 1.458e+02 -0.769 0.441762   
## Cumulative\_GPA 9.619e-01 3.089e-01 3.114 0.001845 \*\*   
## High\_LSAT\_Score 7.790e-02 2.193e-02 3.551 0.000383 \*\*\*  
## Current\_age\_of\_applicant -3.911e-02 2.360e-02 -1.658 0.097399 .   
## Matriculant\_Year 4.924e-02 7.167e-02 0.687 0.492080   
## Total\_Scholarship\_Amt -1.522e-05 1.450e-05 -1.050 0.293918   
## Gender\_ApplicantM -4.125e-02 1.675e-01 -0.246 0.805435   
## Residency\_Longresident -1.820e-01 1.878e-01 -0.969 0.332564   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 984.98 on 1015 degrees of freedom  
## Residual deviance: 956.21 on 1008 degrees of freedom  
## (8 observations deleted due to missingness)  
## AIC: 972.21  
##   
## Number of Fisher Scoring iterations: 4

glm\_bp\_cv\_1 = glm(Bar\_Status ~ Cumulative\_GPA + High\_LSAT\_Score + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
summary(glm\_bp\_cv\_1)

##   
## Call:  
## glm(formula = Bar\_Status ~ Cumulative\_GPA + High\_LSAT\_Score +   
## Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt +   
## Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.6226 -0.6969 -0.5935 -0.4714 2.2287   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.818e+02 1.419e+02 1.985 0.04710 \*   
## Cumulative\_GPA -9.564e-01 3.037e-01 -3.149 0.00164 \*\*   
## High\_LSAT\_Score -9.511e-02 2.138e-02 -4.449 8.64e-06 \*\*\*  
## Current\_age\_of\_applicant 3.979e-02 2.348e-02 1.694 0.09022 .   
## Matriculant\_Year -1.320e-01 6.978e-02 -1.892 0.05848 .   
## Total\_Scholarship\_Amt 1.528e-05 1.397e-05 1.093 0.27426   
## Gender\_ApplicantM -2.199e-02 1.632e-01 -0.135 0.89278   
## Residency\_Longresident -5.983e-02 1.786e-01 -0.335 0.73757   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1032.69 on 1015 degrees of freedom  
## Residual deviance: 991.52 on 1008 degrees of freedom  
## (8 observations deleted due to missingness)  
## AIC: 1007.5  
##   
## Number of Fisher Scoring iterations: 4

1. Year 1 only

glm\_emply\_cv\_2 = glm(Employed ~ Fall\_1L + Spring\_1L + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
summary(glm\_emply\_cv\_2)

##   
## Call:  
## glm(formula = Employed ~ Fall\_1L + Spring\_1L + Current\_age\_of\_applicant +   
## Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant +   
## Residency\_Long, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4470 0.3045 0.4601 0.6519 1.7558   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 6.825e+01 1.410e+02 0.484 0.6284   
## Fall\_1L -1.261e+00 5.933e-01 -2.125 0.0336 \*   
## Spring\_1L 3.798e+00 6.530e-01 5.816 6.02e-09 \*\*\*  
## Current\_age\_of\_applicant -5.421e-02 2.307e-02 -2.350 0.0188 \*   
## Matriculant\_Year -3.635e-02 6.991e-02 -0.520 0.6031   
## Total\_Scholarship\_Amt -5.070e-06 1.361e-05 -0.373 0.7095   
## Gender\_ApplicantM -1.583e-01 1.728e-01 -0.916 0.3597   
## Residency\_Longresident -1.112e-01 1.932e-01 -0.576 0.5650   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 993.74 on 1022 degrees of freedom  
## Residual deviance: 869.81 on 1015 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 885.81  
##   
## Number of Fisher Scoring iterations: 5

glm\_bp\_cv\_2 = glm(Bar\_Status ~ Fall\_1L + Spring\_1L + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
summary(glm\_bp\_cv\_2)

##   
## Call:  
## glm(formula = Bar\_Status ~ Fall\_1L + Spring\_1L + Current\_age\_of\_applicant +   
## Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant +   
## Residency\_Long, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.3994 -0.6167 -0.3991 -0.2116 2.6906   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.404e+02 1.448e+02 0.969 0.3324   
## Fall\_1L 3.647e-01 6.011e-01 0.607 0.5441   
## Spring\_1L -4.021e+00 6.689e-01 -6.010 1.85e-09 \*\*\*  
## Current\_age\_of\_applicant 5.108e-02 2.417e-02 2.113 0.0346 \*   
## Matriculant\_Year -6.537e-02 7.180e-02 -0.911 0.3625   
## Total\_Scholarship\_Amt 1.200e-05 1.399e-05 0.858 0.3910   
## Gender\_ApplicantM 7.032e-02 1.775e-01 0.396 0.6920   
## Residency\_Longresident -7.503e-02 1.954e-01 -0.384 0.7010   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1038.62 on 1022 degrees of freedom  
## Residual deviance: 820.63 on 1015 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 836.63  
##   
## Number of Fisher Scoring iterations: 5

1. Year 2 only

glm\_emply\_cv\_3 = glm(Employed ~ Fall\_2L + Spring\_2L + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
summary(glm\_emply\_cv\_3)

##   
## Call:  
## glm(formula = Employed ~ Fall\_2L + Spring\_2L + Current\_age\_of\_applicant +   
## Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant +   
## Residency\_Long, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5075 0.2857 0.4596 0.6516 1.7661   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 7.500e+01 1.434e+02 0.523 0.6009   
## Fall\_2L -4.400e-01 1.537e+00 -0.286 0.7746   
## Spring\_2L 3.511e+00 1.606e+00 2.186 0.0288 \*   
## Current\_age\_of\_applicant -6.659e-02 2.338e-02 -2.848 0.0044 \*\*  
## Matriculant\_Year -4.042e-02 7.111e-02 -0.568 0.5697   
## Total\_Scholarship\_Amt -1.074e-05 1.381e-05 -0.778 0.4368   
## Gender\_ApplicantM -1.269e-01 1.736e-01 -0.731 0.4646   
## Residency\_Longresident -1.573e-01 1.954e-01 -0.805 0.4206   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 993.74 on 1022 degrees of freedom  
## Residual deviance: 858.33 on 1015 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 874.33  
##   
## Number of Fisher Scoring iterations: 5

glm\_bp\_cv\_3 = glm(Bar\_Status ~ Fall\_2L + Spring\_2L + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
summary(glm\_bp\_cv\_3)

##   
## Call:  
## glm(formula = Bar\_Status ~ Fall\_2L + Spring\_2L + Current\_age\_of\_applicant +   
## Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant +   
## Residency\_Long, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2079 -0.5971 -0.3550 -0.1629 2.8695   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.383e+02 1.519e+02 0.911 0.36251   
## Fall\_2L 2.586e+00 1.616e+00 1.601 0.10944   
## Spring\_2L -7.436e+00 1.723e+00 -4.316 1.59e-05 \*\*\*  
## Current\_age\_of\_applicant 7.139e-02 2.530e-02 2.822 0.00478 \*\*   
## Matriculant\_Year -6.266e-02 7.530e-02 -0.832 0.40533   
## Total\_Scholarship\_Amt 2.190e-05 1.461e-05 1.498 0.13401   
## Gender\_ApplicantM 4.924e-02 1.831e-01 0.269 0.78803   
## Residency\_Longresident -7.763e-02 2.028e-01 -0.383 0.70190   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1038.62 on 1022 degrees of freedom  
## Residual deviance: 770.27 on 1015 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 786.27  
##   
## Number of Fisher Scoring iterations: 5

1. Undergraduate GPA & first LSAT & Year one

glm\_emply\_cv\_4 = glm(Employed ~ Cumulative\_GPA + High\_LSAT\_Score + Fall\_1L + Spring\_1L + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)   
summary(glm\_emply\_cv\_4)

##   
## Call:  
## glm(formula = Employed ~ Cumulative\_GPA + High\_LSAT\_Score + Fall\_1L +   
## Spring\_1L + Current\_age\_of\_applicant + Matriculant\_Year +   
## Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long,   
## family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4862 0.3048 0.4632 0.6506 1.7504   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 6.933e+01 1.534e+02 0.452 0.6513   
## Cumulative\_GPA 1.932e-01 3.417e-01 0.566 0.5717   
## High\_LSAT\_Score -1.686e-02 2.509e-02 -0.672 0.5016   
## Fall\_1L -1.317e+00 5.951e-01 -2.213 0.0269 \*   
## Spring\_1L 3.884e+00 6.661e-01 5.830 5.54e-09 \*\*\*  
## Current\_age\_of\_applicant -4.819e-02 2.468e-02 -1.953 0.0509 .   
## Matriculant\_Year -3.602e-02 7.535e-02 -0.478 0.6326   
## Total\_Scholarship\_Amt -2.636e-06 1.557e-05 -0.169 0.8656   
## Gender\_ApplicantM -1.068e-01 1.773e-01 -0.603 0.5468   
## Residency\_Longresident -1.258e-01 1.978e-01 -0.636 0.5247   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 984.98 on 1015 degrees of freedom  
## Residual deviance: 862.64 on 1006 degrees of freedom  
## (8 observations deleted due to missingness)  
## AIC: 882.64  
##   
## Number of Fisher Scoring iterations: 5

glm\_bp\_cv\_4 = glm(Bar\_Status ~ Cumulative\_GPA + High\_LSAT\_Score + Fall\_1L + Spring\_1L + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
summary(glm\_bp\_cv\_4)

##   
## Call:  
## glm(formula = Bar\_Status ~ Cumulative\_GPA + High\_LSAT\_Score +   
## Fall\_1L + Spring\_1L + Current\_age\_of\_applicant + Matriculant\_Year +   
## Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long,   
## family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4926 -0.6117 -0.3957 -0.2070 2.6608   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 6.012e+01 1.574e+02 0.382 0.7025   
## Cumulative\_GPA 1.290e-01 3.544e-01 0.364 0.7159   
## High\_LSAT\_Score 3.882e-02 2.578e-02 1.506 0.1321   
## Fall\_1L 2.775e-01 6.038e-01 0.460 0.6459   
## Spring\_1L -4.141e+00 6.838e-01 -6.056 1.39e-09 \*\*\*  
## Current\_age\_of\_applicant 5.858e-02 2.571e-02 2.278 0.0227 \*   
## Matriculant\_Year -2.855e-02 7.731e-02 -0.369 0.7119   
## Total\_Scholarship\_Amt -1.217e-06 1.597e-05 -0.076 0.9393   
## Gender\_ApplicantM 4.926e-02 1.826e-01 0.270 0.7873   
## Residency\_Longresident -1.366e-01 1.995e-01 -0.685 0.4936   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1032.69 on 1015 degrees of freedom  
## Residual deviance: 811.75 on 1006 degrees of freedom  
## (8 observations deleted due to missingness)  
## AIC: 831.75  
##   
## Number of Fisher Scoring iterations: 5

1. All 6 semesters

glm\_emply\_cv\_5 = glm(Employed ~ six\_semesters.pca$x[,1] + six\_semesters.pca$x[,2] + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
summary(glm\_emply\_cv\_5)

##   
## Call:  
## glm(formula = Employed ~ six\_semesters.pca$x[, 1] + six\_semesters.pca$x[,   
## 2] + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt +   
## Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5541 0.2795 0.4518 0.6463 1.7466   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 8.453e+01 1.443e+02 0.586 0.5580   
## six\_semesters.pca$x[, 1] 4.055e-01 4.003e-02 10.129 <2e-16 \*\*\*  
## six\_semesters.pca$x[, 2] 5.809e-01 1.793e-01 3.239 0.0012 \*\*   
## Current\_age\_of\_applicant -6.783e-02 2.343e-02 -2.895 0.0038 \*\*   
## Matriculant\_Year -4.003e-02 7.156e-02 -0.559 0.5758   
## Total\_Scholarship\_Amt -9.392e-06 1.385e-05 -0.678 0.4975   
## Gender\_ApplicantM -1.102e-01 1.743e-01 -0.632 0.5272   
## Residency\_Longresident -1.353e-01 1.957e-01 -0.692 0.4892   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 993.74 on 1022 degrees of freedom  
## Residual deviance: 850.96 on 1015 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 866.96  
##   
## Number of Fisher Scoring iterations: 5

glm\_bp\_cv\_5 = glm(Bar\_Status ~ six\_semesters.pca$x[,1] + six\_semesters.pca$x[,2] + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt + Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
summary(glm\_bp\_cv\_5)

##   
## Call:  
## glm(formula = Bar\_Status ~ six\_semesters.pca$x[, 1] + six\_semesters.pca$x[,   
## 2] + Current\_age\_of\_applicant + Matriculant\_Year + Total\_Scholarship\_Amt +   
## Gender\_Applicant + Residency\_Long, family = binomial, data = ls\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.1537 -0.5847 -0.3396 -0.1501 2.9043   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.342e+02 1.534e+02 0.874 0.38186   
## six\_semesters.pca$x[, 1] -6.387e-01 4.992e-02 -12.794 < 2e-16 \*\*\*  
## six\_semesters.pca$x[, 2] -8.624e-01 1.942e-01 -4.441 8.97e-06 \*\*\*  
## Current\_age\_of\_applicant 7.156e-02 2.540e-02 2.817 0.00485 \*\*   
## Matriculant\_Year -6.868e-02 7.609e-02 -0.903 0.36675   
## Total\_Scholarship\_Amt 1.957e-05 1.468e-05 1.333 0.18244   
## Gender\_ApplicantM 1.526e-02 1.846e-01 0.083 0.93411   
## Residency\_Longresident -9.055e-02 2.037e-01 -0.444 0.65672   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1038.62 on 1022 degrees of freedom  
## Residual deviance: 758.11 on 1015 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 774.11  
##   
## Number of Fisher Scoring iterations: 5