Chance of Admission - Final Report

Rabab Mohamed, Yordanos Alemu, Kovenda Mbuale

January 8 2021

# Introduction:

This report examines data on Graduate Admissions on 2019. The primary goal is to determine what factors are most predictive of Chance of Admit ( ranging from 0 to 1 ). The predictor variables include GRE Scores ( out of 340 ), TOEFL Scores ( out of 120 ), University Rating ( out of 5 ), Statement of Purpose and Letter of Recommendation Strength ( out of 5 ), Undergraduate GPA ( out of 10 ), Research Experience ( either 0 or 1 ). There are 400 observations in the data set.

Response variable(s) – Need at least 1. All must be continuous variables: 1. Chance of Admission

Predictor variables – Need at least 5. Can be a mixture of continuous and categorical 1. GRE Scores ( out of 340 ) 2. TOEFL Scores ( out of 120 ) 3. University Rating ( out of 5 ) 4. Statement of Purpose and Letter of Recommendation Strength ( out of 5 ) 5. Undergraduate GPA ( out of 10 ) 6. Research Experience ( either 0 or 1 ) Column Names: 1. Research 2. CGPA 3. LOR 4. SOP 5. University Rating 6. TOEFL Score 7. GRE Score First, we will load the data. These data have variables about students at a small liberal arts college.

# Activate the Stat2Data package (must be installed first - one-time step)  
library(readr)  
#x <- na.omit(airquality)  
Admission\_Predict\_withMissingValues <- read\_csv("Admission\_Predict.csv")

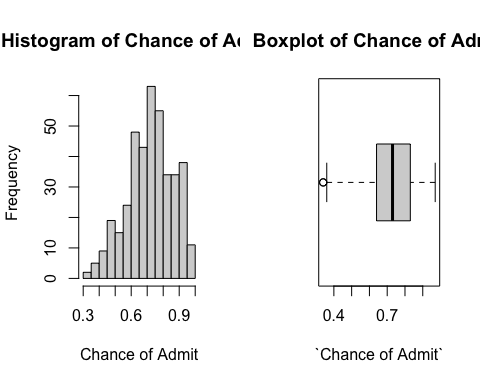
##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## `Serial No.` = col\_double(),  
## `GRE Score` = col\_double(),  
## `TOEFL Score` = col\_double(),  
## `University Rating` = col\_double(),  
## SOP = col\_double(),  
## LOR = col\_double(),  
## CGPA = col\_double(),  
## Research = col\_double(),  
## `Chance of Admit` = col\_double()  
## )

Admission\_Predict <- na.omit(Admission\_Predict\_withMissingValues)

## Introduction

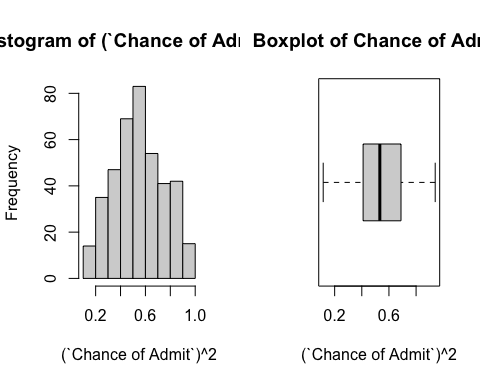
## Exploratory Analysis

# Check distribution of repsonse variable   
attach (Admission\_Predict)  
  
par (mfrow = c(1, 2))  
hist (`Chance of Admit`)  
boxplot (`Chance of Admit`, horizontal = T, xlab="`Chance of Admit`", main="Boxplot of Chance of Admit")

 The distribution of Chance of Admit is left skewed. We will use residual analysis to guide the need for transformation later on in the report.

From our exploratory analysis we saw that the variable chance of Admit is left skewed. Therefore we will try different transformations.

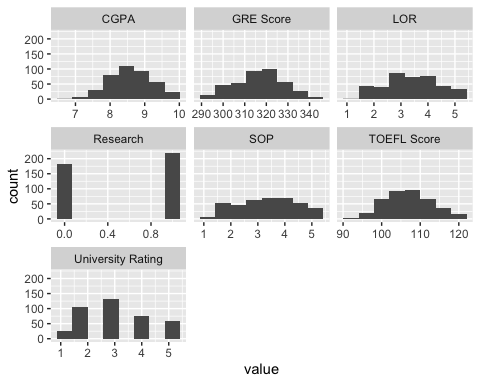
admission\_chance=(`Chance of Admit`)^2  
par (mfrow = c(1, 2))  
hist ((`Chance of Admit`)^2)  
boxplot (admission\_chance, horizontal = T, xlab="(`Chance of Admit`)^2", main="Boxplot of Chance of Admit")



The distribution of (Chance of Admit)^2 is more symmetric. We will start by modeling Chance of Admit and then use the Box-Cox analysis to determine the most appropriate transformation.

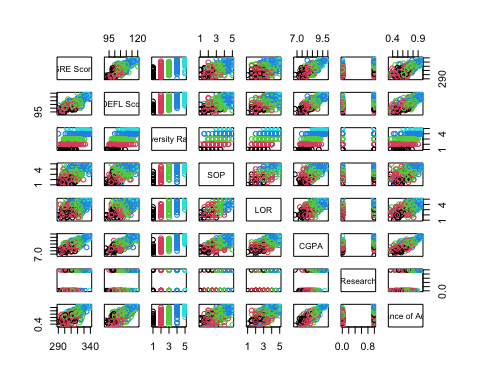
Distributions of the quantitative predictor variables:

library (ggplot2)  
library (tidyr)  
  
ggplot(gather(Admission\_Predict [, 2:8]), aes(value)) +   
 geom\_histogram(bins = 8) +   
 facet\_wrap(~key, scales = 'free\_x')



The predictor variables are not either extremely right or left skewed. We think the variables Research and University Rating are categorical variables but we will use box plot to see weather they have an influence on the response variable which is chance of admission.

# Scatterplot matrix of columns   
pairs (Admission\_Predict[,2:9], col=`University Rating`)



The Exploratory Analysis shows that all our variables are quantitative except University Rating and Research.

# Table of pairwise correlations  
cormat = cor (Admission\_Predict [,2:9], use = "complete.obs")   
round (cormat, 2)

## GRE Score TOEFL Score University Rating SOP LOR CGPA  
## GRE Score 1.00 0.84 0.67 0.61 0.56 0.83  
## TOEFL Score 0.84 1.00 0.70 0.66 0.57 0.83  
## University Rating 0.67 0.70 1.00 0.73 0.66 0.75  
## SOP 0.61 0.66 0.73 1.00 0.73 0.72  
## LOR 0.56 0.57 0.66 0.73 1.00 0.67  
## CGPA 0.83 0.83 0.75 0.72 0.67 1.00  
## Research 0.58 0.49 0.45 0.44 0.40 0.52  
## Chance of Admit 0.80 0.79 0.71 0.68 0.67 0.87  
## Research Chance of Admit  
## GRE Score 0.58 0.80  
## TOEFL Score 0.49 0.79  
## University Rating 0.45 0.71  
## SOP 0.44 0.68  
## LOR 0.40 0.67  
## CGPA 0.52 0.87  
## Research 1.00 0.55  
## Chance of Admit 0.55 1.00

The pairs plot shows all our predictor variables have a linear relationship with our response variable Chance of Admit except for Research. This will be examined further via residual analysis.

The three predictors GRE Score, TOEFL Score and CGPA are highly correlated with each other (r = 0.83 to 0.84). However, Research and Letter of Recommendation are not highly correlated with the other predictor variables.

Since CGPA has the highest correlation with chance of admission, both GRE Score and the TOEFL score has the same correlation to CGPA which means students may not need to take both tests since their correlation to CGPA is the same with r = 0.83. This statement is only an assumption we will confirm this and also recognize other issues of multicollinearity with correlated predictors later. #(self - ref ----See chapter 3)

## Simple Linear Regression

We start with a simple linear regression of GPA and admission\_chance ## Is there a relationship between GPA and admission\_chance?

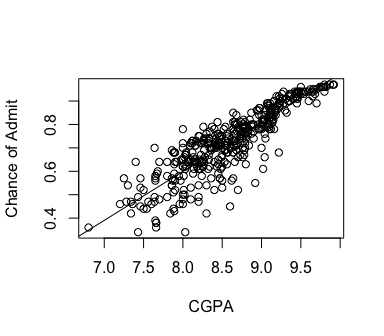
plot ( `Chance of Admit` ~ CGPA, data=Admission\_Predict)  
fit0 = lm (`Chance of Admit` ~ CGPA, data= Admission\_Predict)  
summary (fit0)

##   
## Call:  
## lm(formula = `Chance of Admit` ~ CGPA, data = Admission\_Predict)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.274575 -0.030084 0.009443 0.041954 0.180734   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.07151 0.05034 -21.29 <2e-16 \*\*\*  
## CGPA 0.20885 0.00584 35.76 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06957 on 398 degrees of freedom  
## Multiple R-squared: 0.7626, Adjusted R-squared: 0.762   
## F-statistic: 1279 on 1 and 398 DF, p-value: < 2.2e-16

confint (fit0)

## 2.5 % 97.5 %  
## (Intercept) -1.1704792 -0.9725442  
## CGPA 0.1973654 0.2203290

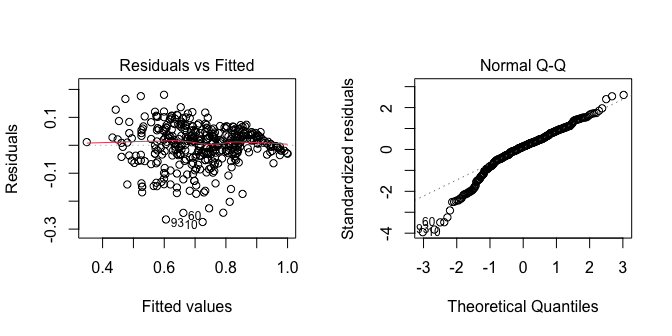
abline (fit0)



With 95% confidence, Chance of Admit increases between 19.7 and 22.0 % for every CGPA point.

Residual analysis:

par (mfrow = c(1,2))  
plot (fit0, which=1:2)



From the Residual Vs Fitted we can see that there is no problem with the linearity but there is non-constant variance. And from the Normal Q-Q we can see there is a problem with the normal distribution. We will try a square transformation for Chance of Admit and CGPA.

## Is there a relationship between CGPA and square(Chance of Admit)?

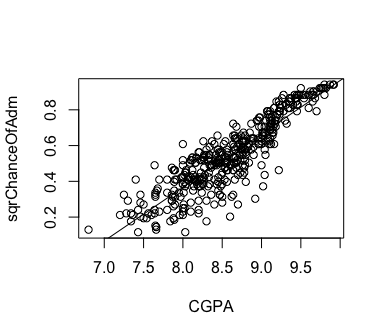
sqrChanceOfAdm = (`Chance of Admit`)^2  
plot ( sqrChanceOfAdm ~ CGPA, data=Admission\_Predict)  
fit1 = lm (sqrChanceOfAdm ~ CGPA, data= Admission\_Predict)  
summary (fit1)

##   
## Call:  
## lm(formula = sqrChanceOfAdm ~ CGPA, data = Admission\_Predict)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.34279 -0.04772 0.01103 0.06431 0.24254   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.026566 0.067137 -30.18 <2e-16 \*\*\*  
## CGPA 0.299053 0.007789 38.39 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09278 on 398 degrees of freedom  
## Multiple R-squared: 0.7874, Adjusted R-squared: 0.7869   
## F-statistic: 1474 on 1 and 398 DF, p-value: < 2.2e-16

confint (fit1)

## 2.5 % 97.5 %  
## (Intercept) -2.1585538 -1.8945773  
## CGPA 0.2837404 0.3143658

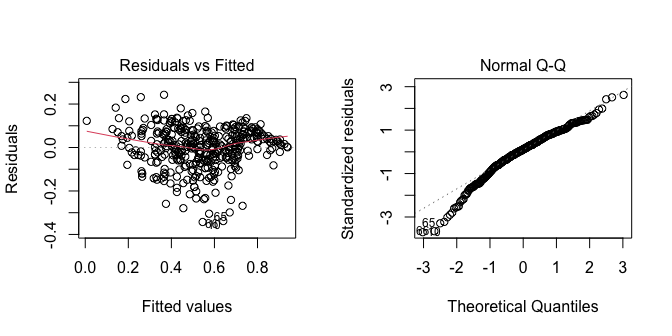
abline (fit1)



The plot of sqrChanceOfAdm vs CGPA shows a more linear relationship. square(CGPA) explains 78.74% of the total variation in sqrChanceOfAdm with a residual standard error of 0.09278 sqrChanceOfAdm.

Residual analysis:

par (mfrow = c(1,2))  
plot (fit1, which=1:2)



From the Residual Vs Fitted we can see that the linearity of the model has gotten worst then the previous untransformed model and the constant variance has not improved either. And from the Normal Q-Q we can see the problem with the normal distribution has been improved.

The transformed model deosnt show much improvement from the untransforemed model except for the improvement in normal distribution but if you look at the symmetry of the transformed Chance of Admit you can see the improvement much better. We will use the Box-Cox analysis to determine the most appropriate transformation for our model.

### First Order Model

Next, we fit a first-order linear model will all seven predictors. ## fit 1

order1\_fit1 = lm (`Chance of Admit` ~ `GRE Score`+`TOEFL Score`+ `University Rating`+ SOP + LOR + CGPA + Research)  
   
summary (order1\_fit1)

##   
## Call:  
## lm(formula = `Chance of Admit` ~ `GRE Score` + `TOEFL Score` +   
## `University Rating` + SOP + LOR + CGPA + Research)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.26259 -0.02103 0.01005 0.03628 0.15928   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.2594325 0.1247307 -10.097 < 2e-16 \*\*\*  
## `GRE Score` 0.0017374 0.0005979 2.906 0.00387 \*\*   
## `TOEFL Score` 0.0029196 0.0010895 2.680 0.00768 \*\*   
## `University Rating` 0.0057167 0.0047704 1.198 0.23150   
## SOP -0.0033052 0.0055616 -0.594 0.55267   
## LOR 0.0223531 0.0055415 4.034 6.6e-05 \*\*\*  
## CGPA 0.1189395 0.0122194 9.734 < 2e-16 \*\*\*  
## Research 0.0245251 0.0079598 3.081 0.00221 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06378 on 392 degrees of freedom  
## Multiple R-squared: 0.8035, Adjusted R-squared: 0.8   
## F-statistic: 228.9 on 7 and 392 DF, p-value: < 2.2e-16

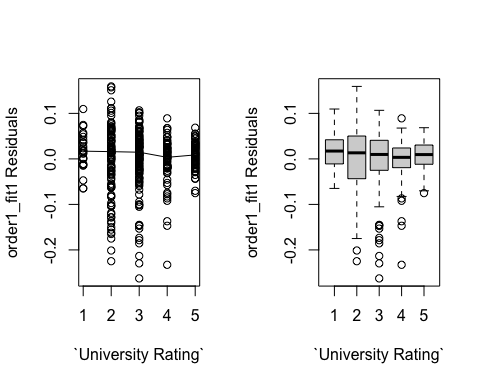
anova (order1\_fit1)

## Analysis of Variance Table  
##   
## Response: Chance of Admit  
## Df Sum Sq Mean Sq F value Pr(>F)   
## `GRE Score` 1 5.2273 5.2273 1284.9016 < 2.2e-16 \*\*\*  
## `TOEFL Score` 1 0.3921 0.3921 96.3830 < 2.2e-16 \*\*\*  
## `University Rating` 1 0.2370 0.2370 58.2444 1.779e-13 \*\*\*  
## SOP 1 0.0757 0.0757 18.6143 2.028e-05 \*\*\*  
## LOR 1 0.1574 0.1574 38.6845 1.278e-09 \*\*\*  
## CGPA 1 0.3918 0.3918 96.2991 < 2.2e-16 \*\*\*  
## Research 1 0.0386 0.0386 9.4934 0.002208 \*\*   
## Residuals 392 1.5948 0.0041   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The analysis of variance table suggests that all the predictors are significant because they each have a p-value less than 0.05. The coefficient tests suggest that all of the predictors are significant except for Statement of Purpose (SOP) and University Rating. The R-squared is 0.8035, with adjusted R-squared = 0.8, which indicate that most of the variability in Chance of Admission (Chance of Admit) is being explained by this model. The residual standard error is 0.06378.

# Trying to Fix the first insignificant variable University Rating

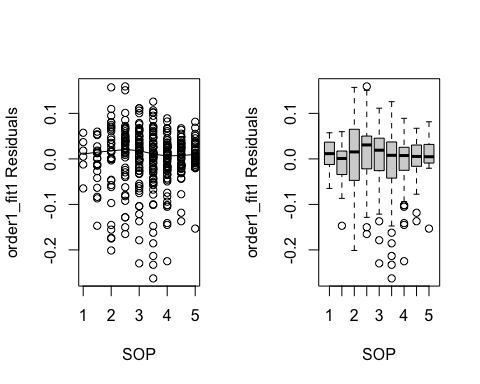
# New code chunk with residuals vs. `University Rating`  
par (mfrow = c(1, 2))  
plot (`University Rating`[!is.na(`Chance of Admit`)], order1\_fit1$residuals, ylab="order1\_fit1 Residuals", xlab="`University Rating`")  
lines (lowess (`University Rating`[!is.na(`Chance of Admit`)], order1\_fit1$residuals))  
boxplot (order1\_fit1$residuals ~ `University Rating`, ylab="order1\_fit1 Residuals",  
xlab="`University Rating`")



Since the variable University Rating appears to have a linear relationship with Chance of Admission, changing University Rating will not improve the first-order model.

# Trying to Fix our second insignificant variable SOP

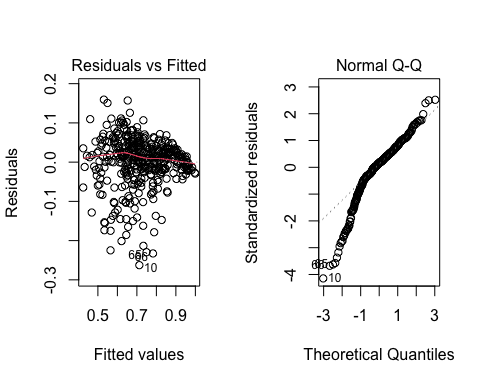
# New code chunk with residuals vs. SOP  
par (mfrow = c(1, 2))  
plot (SOP[!is.na(`Chance of Admit`)], order1\_fit1$residuals, ylab="order1\_fit1 Residuals", xlab="SOP")  
lines (lowess (SOP[!is.na(`Chance of Admit`)], order1\_fit1$residuals))  
boxplot (order1\_fit1$residuals ~ SOP, ylab="order1\_fit1 Residuals",  
xlab="SOP")



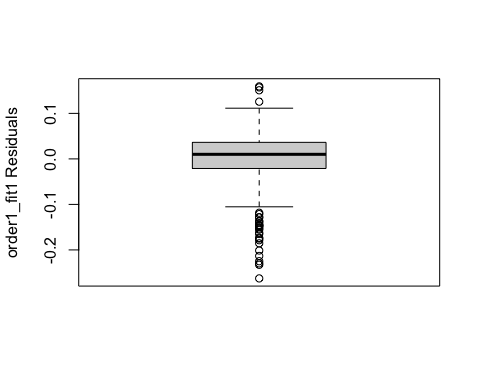
Since the variable Statement of Purpose (SOP) appears to have a linear relationship with Chance of Admission, changing Statement of Purpose (SOP) will not improve the first-order model.

## Residual Analysis of the first order model

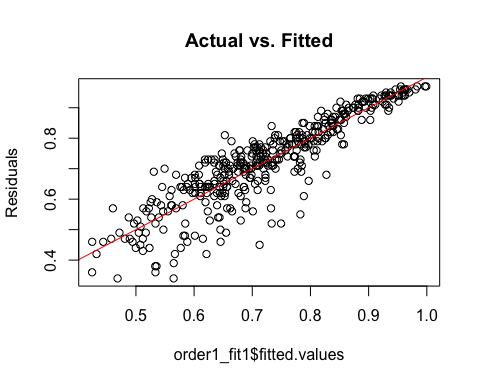
par (mfrow = c(1, 2))  
plot (order1\_fit1, which = c(1, 2))



boxplot (order1\_fit1$residuals, ylab="order1\_fit1 Residuals")



plot (order1\_fit1$fitted.values,`Chance of Admit`, main="Actual vs. Fitted", ylab="Residuals")  
abline (0, 1, col="red")

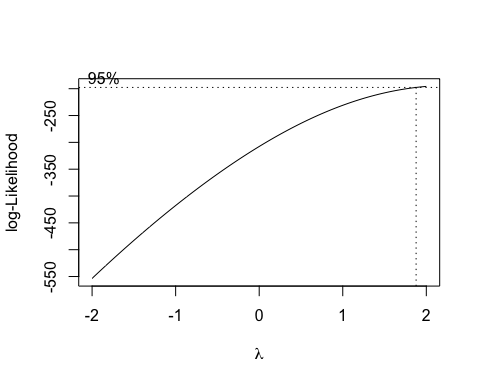


Residual analysis shows linearity but non-constant variance, along with some left skewness in the residual distribution. This could be corrected by transforming the response variable.

## Box-Cox Analysis

Box-Cox analysis is one way to choose a response variable transformation from the set of power transformations. It can also just help choose between log and square root.

library ("MASS")  
boxcox (order1\_fit1)



The Box-Cox analysis suggests a power of 2 transformation, The value, 2, is just inside the 95% confidence interval, so we will use a square transformation.

squareChance\_Admit = (`Chance of Admit`)^2  
order2\_fit1 = lm (squareChance\_Admit ~ `GRE Score`+`TOEFL Score`+ `University Rating`+ SOP + LOR + CGPA + Research)  
summary (order2\_fit1)

##   
## Call:  
## lm(formula = squareChance\_Admit ~ `GRE Score` + `TOEFL Score` +   
## `University Rating` + SOP + LOR + CGPA + Research)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.32801 -0.03855 0.01350 0.05363 0.21244   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.2647537 0.1619522 -13.984 < 2e-16 \*\*\*  
## `GRE Score` 0.0025482 0.0007763 3.282 0.001122 \*\*   
## `TOEFL Score` 0.0042825 0.0014147 3.027 0.002631 \*\*   
## `University Rating` 0.0134879 0.0061940 2.178 0.030033 \*   
## SOP -0.0026650 0.0072213 -0.369 0.712289   
## LOR 0.0271785 0.0071951 3.777 0.000183 \*\*\*  
## CGPA 0.1623005 0.0158659 10.230 < 2e-16 \*\*\*  
## Research 0.0373290 0.0103351 3.612 0.000343 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08282 on 392 degrees of freedom  
## Multiple R-squared: 0.8332, Adjusted R-squared: 0.8302   
## F-statistic: 279.7 on 7 and 392 DF, p-value: < 2.2e-16

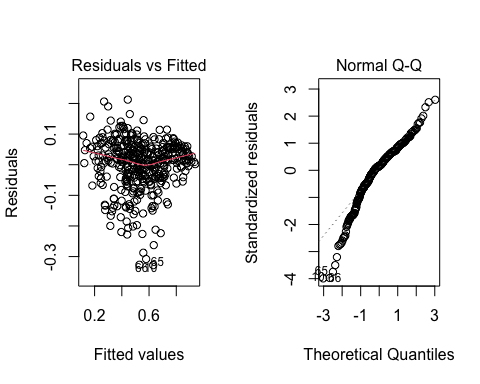
anova (order2\_fit1)

## Analysis of Variance Table  
##   
## Response: squareChance\_Admit  
## Df Sum Sq Mean Sq F value Pr(>F)   
## `GRE Score` 1 10.8206 10.8206 1577.674 < 2.2e-16 \*\*\*  
## `TOEFL Score` 1 0.8204 0.8204 119.614 < 2.2e-16 \*\*\*  
## `University Rating` 1 0.5649 0.5649 82.359 < 2.2e-16 \*\*\*  
## SOP 1 0.1469 0.1469 21.419 5.026e-06 \*\*\*  
## LOR 1 0.2532 0.2532 36.912 2.928e-09 \*\*\*  
## CGPA 1 0.7308 0.7308 106.553 < 2.2e-16 \*\*\*  
## Research 1 0.0895 0.0895 13.046 0.0003435 \*\*\*  
## Residuals 392 2.6886 0.0069   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

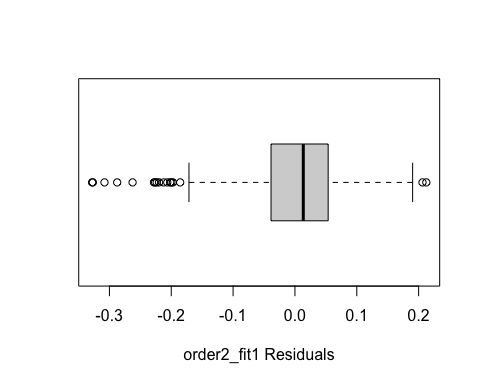
University Rating is now more significant to the model the p value went from 0.23150 to 0.030033, however Statement of Purpose is still insignificant to the model and we might eliminate it to improve the model. We cant compare the Residual standard error because we changed scales. The Multiple R-squared increased by 0.03.

## Residual analysis of the square-transformed first-order model

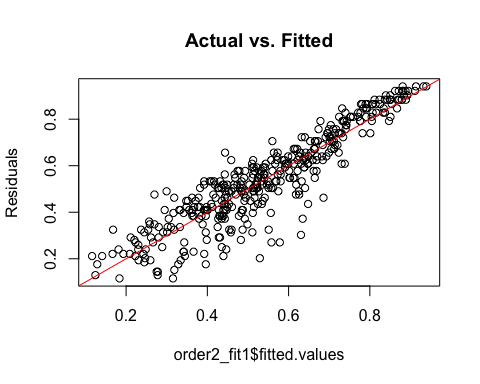
par (mfrow = c(1, 2))  
plot (order2\_fit1, which = c(1, 2))



boxplot (order2\_fit1$residuals, horizontal = T, xlab="order2\_fit1 Residuals")



plot (order2\_fit1$fitted.values,squareChance\_Admit, main="Actual vs. Fitted", ylab="Residuals")  
abline (0, 1, col="red")



The residual analysis of the square-transformed model looks fairly like a fairly good improvement from the first order model. There are some residuals at both ends of the scale that are somewhat more spread out compared to a normal distribution. The plot of observed vs fitted squareChance\_Admit looks good.

## Backward elimination method - Manual

We will remove Statement of Purpose (SOP) because its the only one that is insignificant to the square-transformed model.

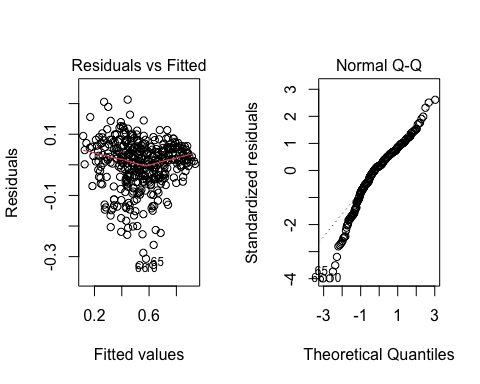
order2\_fit2 = lm (squareChance\_Admit ~ `GRE Score`+`TOEFL Score`+ `University Rating` + LOR + CGPA + Research)  
summary (order2\_fit2)

##   
## Call:  
## lm(formula = squareChance\_Admit ~ `GRE Score` + `TOEFL Score` +   
## `University Rating` + LOR + CGPA + Research)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.32784 -0.03944 0.01340 0.05341 0.21292   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.2606229 0.1613872 -14.007 < 2e-16 \*\*\*  
## `GRE Score` 0.0025701 0.0007732 3.324 0.000971 \*\*\*  
## `TOEFL Score` 0.0042174 0.0014021 3.008 0.002799 \*\*   
## `University Rating` 0.0127923 0.0058937 2.170 0.030568 \*   
## LOR 0.0261139 0.0065842 3.966 8.68e-05 \*\*\*  
## CGPA 0.1614677 0.0156873 10.293 < 2e-16 \*\*\*  
## Research 0.0370299 0.0102919 3.598 0.000362 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08273 on 393 degrees of freedom  
## Multiple R-squared: 0.8331, Adjusted R-squared: 0.8306   
## F-statistic: 327 on 6 and 393 DF, p-value: < 2.2e-16

anova (order2\_fit2)

## Analysis of Variance Table  
##   
## Response: squareChance\_Admit  
## Df Sum Sq Mean Sq F value Pr(>F)   
## `GRE Score` 1 10.8206 10.8206 1581.149 < 2.2e-16 \*\*\*  
## `TOEFL Score` 1 0.8204 0.8204 119.877 < 2.2e-16 \*\*\*  
## `University Rating` 1 0.5649 0.5649 82.541 < 2.2e-16 \*\*\*  
## LOR 1 0.3862 0.3862 56.432 3.957e-13 \*\*\*  
## CGPA 1 0.7446 0.7446 108.808 < 2.2e-16 \*\*\*  
## Research 1 0.0886 0.0886 12.945 0.0003617 \*\*\*  
## Residuals 393 2.6895 0.0068   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

par (mfrow = c(1, 2))  
plot (order2\_fit2, which = c(1, 2))



We do not see any obvious change in the model after dropping Statement of Purpose (SOP). So we will use ANOVA between the two fits:

anova (order2\_fit1, order2\_fit2)

## Analysis of Variance Table  
##   
## Model 1: squareChance\_Admit ~ `GRE Score` + `TOEFL Score` + `University Rating` +   
## SOP + LOR + CGPA + Research  
## Model 2: squareChance\_Admit ~ `GRE Score` + `TOEFL Score` + `University Rating` +   
## LOR + CGPA + Research  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 392 2.6886   
## 2 393 2.6895 -1 -0.00093413 0.1362 0.7123

The p-value for the hypothesis that the two fits are the same is 0.7123. Thus, we conclude that the second fit is not significantly different from the first fit. Therefore removing the Statement of Purpose from the model doesn't improve the model, however, removing it gives us a simpler model so we will do that.

## Changing the order of the predictor variables in the model

We noticed from our earlier correlation matrix that TOEFL Score and GRE Score are highly correlated to CGPA, we know that this means that the order in we put them into the model affects their Sum of Squares.

## Determining Final Order of variables in the model

|  |  |  |
| --- | --- | --- |
| Predictors | Individual SS | Individual R Squared |
| CGPA | 12.6889 | 0.7874 |
| GRE Score | 10.8206 | 0.6715 |
| TOEFL Score | 10.543 | 0.6542 |
| Research | 5.2035 | 0.3229 |
| University Rating | 8.7150 | 0.5408 |
| LOR | 7.3745 | 0.4576 |

We take the variable CGPA with the highest R-squared value to be first in the model. However to determine which variable comes next in the model, we need to not only to consider the high R-Square values that follow the first but also consider how much of their individual Sum of Squares is lost to CGPA being first in the model since we that CGPA is going to take 12.6889 Sum of Squares of the 13.4253 Total Sum of Squares.

If we re-arrange the table in terms of the set criteria it looks as follows (The number corresponds to the variable's determined position in the coming re-arranged model).

Individual SS And Individual R Squared 1. 12.6889 CGPA 0.7874 2. 10.8206 GRE Score 0.6715 3. 10.543 TOEFL Score 0.6542 4. 8.7150 University Rating 0.5408 5. 7.3745 LOR 0.4576 6. 5.2035 Research 0.3229

This leads us to create a an order of the variables in the following order: 1. CGPA 2. GRE Score 3. TOEFL Score 4. University Rating 5. LOR 6. Research.

order2\_fit3 = lm (squareChance\_Admit ~ CGPA + `GRE Score` + `TOEFL Score` + `University Rating` + LOR + Research )  
summary (order2\_fit3)

##   
## Call:  
## lm(formula = squareChance\_Admit ~ CGPA + `GRE Score` + `TOEFL Score` +   
## `University Rating` + LOR + Research)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.32784 -0.03944 0.01340 0.05341 0.21292   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.2606229 0.1613872 -14.007 < 2e-16 \*\*\*  
## CGPA 0.1614677 0.0156873 10.293 < 2e-16 \*\*\*  
## `GRE Score` 0.0025701 0.0007732 3.324 0.000971 \*\*\*  
## `TOEFL Score` 0.0042174 0.0014021 3.008 0.002799 \*\*   
## `University Rating` 0.0127923 0.0058937 2.170 0.030568 \*   
## LOR 0.0261139 0.0065842 3.966 8.68e-05 \*\*\*  
## Research 0.0370299 0.0102919 3.598 0.000362 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08273 on 393 degrees of freedom  
## Multiple R-squared: 0.8331, Adjusted R-squared: 0.8306   
## F-statistic: 327 on 6 and 393 DF, p-value: < 2.2e-16

anova (order2\_fit3)

## Analysis of Variance Table  
##   
## Response: squareChance\_Admit  
## Df Sum Sq Mean Sq F value Pr(>F)   
## CGPA 1 12.6889 12.6889 1854.149 < 2.2e-16 \*\*\*  
## `GRE Score` 1 0.3388 0.3388 49.506 8.846e-12 \*\*\*  
## `TOEFL Score` 1 0.0832 0.0832 12.159 0.0005437 \*\*\*  
## `University Rating` 1 0.1057 0.1057 15.447 0.0001003 \*\*\*  
## LOR 1 0.1201 0.1201 17.546 3.466e-05 \*\*\*  
## Research 1 0.0886 0.0886 12.945 0.0003617 \*\*\*  
## Residuals 393 2.6895 0.0068   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Initial Interpretation of the Model

The predictor variables explain 83.06 % of the change in the square(chance of admission). We think that the intercept does not provide significant information so we won't comment on it. For an increase of 1 CGPA point your chance of admission increases by 16% and coming from a University that is rated 1 point higher than a fellow applicant gives you 1.28% higher chance of admission. It is interesting to see that an applicant's Letter of Recommendation (LOR) takes precedence over both of the required Standardized test scores, since an increase of 1 point in the score of an applicant's Letter of Recommendation (LOR) increases their chance of admission by 2.6% while a 1 point increase in the standardized test scores only gives an applicant 0.26% for GRE Score and 0.42% for TOEFL Score. Doing research happens to be the second most important contributor to an applicant's chance of admission after CGPA.

## Apply a model selection to the first order model

# First order model

part2\_fit1 = lm (`Chance of Admit` ~ `GRE Score`+`TOEFL Score`+ `University Rating`+ SOP + LOR + CGPA + Research)  
summary (part2\_fit1)

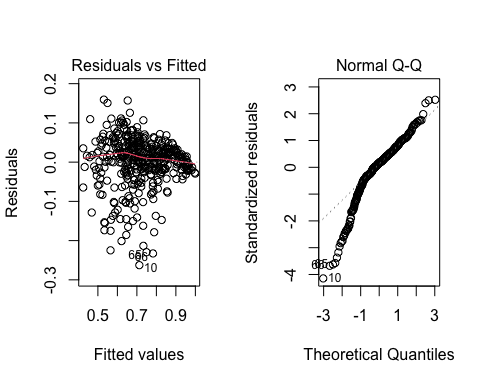
##   
## Call:  
## lm(formula = `Chance of Admit` ~ `GRE Score` + `TOEFL Score` +   
## `University Rating` + SOP + LOR + CGPA + Research)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.26259 -0.02103 0.01005 0.03628 0.15928   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.2594325 0.1247307 -10.097 < 2e-16 \*\*\*  
## `GRE Score` 0.0017374 0.0005979 2.906 0.00387 \*\*   
## `TOEFL Score` 0.0029196 0.0010895 2.680 0.00768 \*\*   
## `University Rating` 0.0057167 0.0047704 1.198 0.23150   
## SOP -0.0033052 0.0055616 -0.594 0.55267   
## LOR 0.0223531 0.0055415 4.034 6.6e-05 \*\*\*  
## CGPA 0.1189395 0.0122194 9.734 < 2e-16 \*\*\*  
## Research 0.0245251 0.0079598 3.081 0.00221 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06378 on 392 degrees of freedom  
## Multiple R-squared: 0.8035, Adjusted R-squared: 0.8   
## F-statistic: 228.9 on 7 and 392 DF, p-value: < 2.2e-16

anova (part2\_fit1)

## Analysis of Variance Table  
##   
## Response: Chance of Admit  
## Df Sum Sq Mean Sq F value Pr(>F)   
## `GRE Score` 1 5.2273 5.2273 1284.9016 < 2.2e-16 \*\*\*  
## `TOEFL Score` 1 0.3921 0.3921 96.3830 < 2.2e-16 \*\*\*  
## `University Rating` 1 0.2370 0.2370 58.2444 1.779e-13 \*\*\*  
## SOP 1 0.0757 0.0757 18.6143 2.028e-05 \*\*\*  
## LOR 1 0.1574 0.1574 38.6845 1.278e-09 \*\*\*  
## CGPA 1 0.3918 0.3918 96.2991 < 2.2e-16 \*\*\*  
## Research 1 0.0386 0.0386 9.4934 0.002208 \*\*   
## Residuals 392 1.5948 0.0041   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Re- fit first order model

par (mfrow = c(1, 2))  
plot (part2\_fit1, which = c(1, 2))

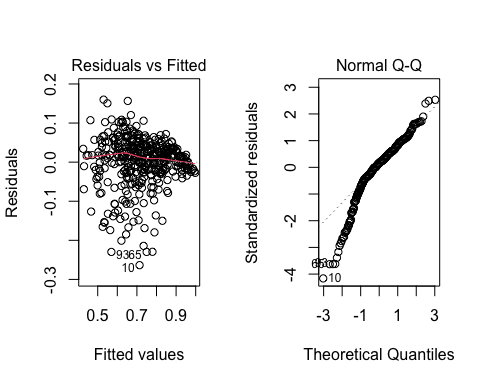


# Stepwise Regression

part2\_fit2=step(part2\_fit1)

## Start: AIC=-2193.9  
## `Chance of Admit` ~ `GRE Score` + `TOEFL Score` + `University Rating` +   
## SOP + LOR + CGPA + Research  
##   
## Df Sum of Sq RSS AIC  
## - SOP 1 0.00144 1.5962 -2195.5  
## - `University Rating` 1 0.00584 1.6006 -2194.4  
## <none> 1.5948 -2193.9  
## - `TOEFL Score` 1 0.02921 1.6240 -2188.6  
## - `GRE Score` 1 0.03435 1.6291 -2187.4  
## - Research 1 0.03862 1.6334 -2186.3  
## - LOR 1 0.06620 1.6609 -2179.6  
## - CGPA 1 0.38544 1.9802 -2109.3  
##   
## Step: AIC=-2195.54  
## `Chance of Admit` ~ `GRE Score` + `TOEFL Score` + `University Rating` +   
## LOR + CGPA + Research  
##   
## Df Sum of Sq RSS AIC  
## - `University Rating` 1 0.00464 1.6008 -2196.4  
## <none> 1.5962 -2195.5  
## - `TOEFL Score` 1 0.02806 1.6242 -2190.6  
## - `GRE Score` 1 0.03565 1.6318 -2188.7  
## - Research 1 0.03769 1.6339 -2188.2  
## - LOR 1 0.06983 1.6660 -2180.4  
## - CGPA 1 0.38660 1.9828 -2110.8  
##   
## Step: AIC=-2196.38  
## `Chance of Admit` ~ `GRE Score` + `TOEFL Score` + LOR + CGPA +   
## Research  
##   
## Df Sum of Sq RSS AIC  
## <none> 1.6008 -2196.4  
## - `TOEFL Score` 1 0.03292 1.6338 -2190.2  
## - `GRE Score` 1 0.03638 1.6372 -2189.4  
## - Research 1 0.03912 1.6400 -2188.7  
## - LOR 1 0.09133 1.6922 -2176.2  
## - CGPA 1 0.43201 2.0328 -2102.8

par (mfrow = c(1, 2))  
plot (part2\_fit2, which = c(1, 2))



The Residual vs Fitted plot shows good linearity and constant variables. The normal Q-Q plot aligns with normal distribution with a left tail.

# Centered Interaction effects

CGPA.c=CGPA-mean(CGPA)  
GreScore.c=`GRE Score`-mean(`GRE Score`)  
ToeflScore.c=`TOEFL Score`-mean(`TOEFL Score`)  
LOR.c=LOR-mean(LOR)  
Research.c=Research-mean(Research)  
  
part2\_fit3= lm ( `Chance of Admit`~ (CGPA.c + GreScore.c + ToeflScore.c + LOR.c + Research.c)^2 )  
summary (part2\_fit3)

##   
## Call:  
## lm(formula = `Chance of Admit` ~ (CGPA.c + GreScore.c + ToeflScore.c +   
## LOR.c + Research.c)^2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.256007 -0.022587 0.009704 0.035928 0.160275   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.7233378 0.0045013 160.694 < 2e-16 \*\*\*  
## CGPA.c 0.1205523 0.0121866 9.892 < 2e-16 \*\*\*  
## GreScore.c 0.0015825 0.0006425 2.463 0.014215 \*   
## ToeflScore.c 0.0031001 0.0010965 2.827 0.004940 \*\*   
## LOR.c 0.0227421 0.0048838 4.657 4.44e-06 \*\*\*  
## Research.c 0.0305427 0.0083425 3.661 0.000286 \*\*\*  
## CGPA.c:GreScore.c -0.0009908 0.0013541 -0.732 0.464796   
## CGPA.c:ToeflScore.c 0.0006427 0.0027054 0.238 0.812358   
## CGPA.c:LOR.c 0.0026550 0.0140034 0.190 0.849727   
## CGPA.c:Research.c 0.0209506 0.0284170 0.737 0.461418   
## GreScore.c:ToeflScore.c -0.0000159 0.0001166 -0.136 0.891625   
## GreScore.c:LOR.c -0.0001806 0.0009566 -0.189 0.850339   
## GreScore.c:Research.c 0.0008749 0.0014023 0.624 0.533042   
## ToeflScore.c:LOR.c -0.0005749 0.0015580 -0.369 0.712313   
## ToeflScore.c:Research.c 0.0006725 0.0024944 0.270 0.787623   
## LOR.c:Research.c 0.0021244 0.0124365 0.171 0.864458   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06396 on 384 degrees of freedom  
## Multiple R-squared: 0.8064, Adjusted R-squared: 0.7988   
## F-statistic: 106.6 on 15 and 384 DF, p-value: < 2.2e-16

# Stepwise by AIC

part2\_fit\_AIC = step (part2\_fit3, direction="both")

## Start: AIC=-2183.92  
## `Chance of Admit` ~ (CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c)^2  
##   
## Df Sum of Sq RSS AIC  
## - GreScore.c:ToeflScore.c 1 0.00007604 1.5710 -2185.9  
## - LOR.c:Research.c 1 0.00011937 1.5710 -2185.9  
## - GreScore.c:LOR.c 1 0.00014584 1.5711 -2185.9  
## - CGPA.c:LOR.c 1 0.00014705 1.5711 -2185.9  
## - CGPA.c:ToeflScore.c 1 0.00023085 1.5712 -2185.9  
## - ToeflScore.c:Research.c 1 0.00029732 1.5712 -2185.8  
## - ToeflScore.c:LOR.c 1 0.00055710 1.5715 -2185.8  
## - GreScore.c:Research.c 1 0.00159256 1.5725 -2185.5  
## - CGPA.c:GreScore.c 1 0.00219024 1.5731 -2185.4  
## - CGPA.c:Research.c 1 0.00222360 1.5731 -2185.4  
## <none> 1.5709 -2183.9  
##   
## Step: AIC=-2185.9  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:LOR.c +   
## CGPA.c:Research.c + GreScore.c:LOR.c + GreScore.c:Research.c +   
## ToeflScore.c:LOR.c + ToeflScore.c:Research.c + LOR.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - LOR.c:Research.c 1 0.00011432 1.5711 -2187.9  
## - CGPA.c:ToeflScore.c 1 0.00015992 1.5712 -2187.9  
## - GreScore.c:LOR.c 1 0.00017623 1.5712 -2187.9  
## - ToeflScore.c:Research.c 1 0.00023306 1.5712 -2187.8  
## - CGPA.c:LOR.c 1 0.00024504 1.5712 -2187.8  
## - ToeflScore.c:LOR.c 1 0.00061945 1.5716 -2187.8  
## - GreScore.c:Research.c 1 0.00154088 1.5725 -2187.5  
## - CGPA.c:GreScore.c 1 0.00275651 1.5737 -2187.2  
## - CGPA.c:Research.c 1 0.00284461 1.5738 -2187.2  
## <none> 1.5710 -2185.9  
## + GreScore.c:ToeflScore.c 1 0.00007604 1.5709 -2183.9  
##   
## Step: AIC=-2187.87  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:LOR.c +   
## CGPA.c:Research.c + GreScore.c:LOR.c + GreScore.c:Research.c +   
## ToeflScore.c:LOR.c + ToeflScore.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - GreScore.c:LOR.c 1 0.0000904 1.5712 -2189.8  
## - CGPA.c:ToeflScore.c 1 0.0001670 1.5713 -2189.8  
## - CGPA.c:LOR.c 1 0.0002449 1.5714 -2189.8  
## - ToeflScore.c:Research.c 1 0.0002574 1.5714 -2189.8  
## - ToeflScore.c:LOR.c 1 0.0006842 1.5718 -2189.7  
## - GreScore.c:Research.c 1 0.0014441 1.5725 -2189.5  
## - CGPA.c:GreScore.c 1 0.0032045 1.5743 -2189.1  
## - CGPA.c:Research.c 1 0.0042300 1.5753 -2188.8  
## <none> 1.5711 -2187.9  
## + LOR.c:Research.c 1 0.0001143 1.5710 -2185.9  
## + GreScore.c:ToeflScore.c 1 0.0000710 1.5710 -2185.9  
##   
## Step: AIC=-2189.85  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:LOR.c +   
## CGPA.c:Research.c + GreScore.c:Research.c + ToeflScore.c:LOR.c +   
## ToeflScore.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - CGPA.c:LOR.c 1 0.0001686 1.5714 -2191.8  
## - ToeflScore.c:Research.c 1 0.0002883 1.5715 -2191.8  
## - CGPA.c:ToeflScore.c 1 0.0003634 1.5716 -2191.8  
## - GreScore.c:Research.c 1 0.0013538 1.5725 -2191.5  
## - ToeflScore.c:LOR.c 1 0.0015958 1.5728 -2191.4  
## - CGPA.c:Research.c 1 0.0044541 1.5756 -2190.7  
## - CGPA.c:GreScore.c 1 0.0050925 1.5763 -2190.6  
## <none> 1.5712 -2189.8  
## + GreScore.c:ToeflScore.c 1 0.0000957 1.5711 -2187.9  
## + GreScore.c:LOR.c 1 0.0000904 1.5711 -2187.9  
## + LOR.c:Research.c 1 0.0000285 1.5712 -2187.9  
##   
## Step: AIC=-2191.81  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:Research.c +   
## GreScore.c:Research.c + ToeflScore.c:LOR.c + ToeflScore.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - ToeflScore.c:Research.c 1 0.0002523 1.5716 -2193.7  
## - CGPA.c:ToeflScore.c 1 0.0002974 1.5717 -2193.7  
## - GreScore.c:Research.c 1 0.0012595 1.5726 -2193.5  
## - ToeflScore.c:LOR.c 1 0.0019524 1.5733 -2193.3  
## - CGPA.c:GreScore.c 1 0.0049846 1.5763 -2192.5  
## - CGPA.c:Research.c 1 0.0052718 1.5766 -2192.5  
## <none> 1.5714 -2191.8  
## + GreScore.c:ToeflScore.c 1 0.0001693 1.5712 -2189.8  
## + CGPA.c:LOR.c 1 0.0001686 1.5712 -2189.8  
## + LOR.c:Research.c 1 0.0000580 1.5713 -2189.8  
## + GreScore.c:LOR.c 1 0.0000141 1.5714 -2189.8  
##   
## Step: AIC=-2193.74  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:Research.c +   
## GreScore.c:Research.c + ToeflScore.c:LOR.c  
##   
## Df Sum of Sq RSS AIC  
## - CGPA.c:ToeflScore.c 1 0.0005120 1.5721 -2195.6  
## - ToeflScore.c:LOR.c 1 0.0018361 1.5735 -2195.3  
## - GreScore.c:Research.c 1 0.0026296 1.5742 -2195.1  
## - CGPA.c:GreScore.c 1 0.0062708 1.5779 -2194.2  
## - CGPA.c:Research.c 1 0.0069380 1.5785 -2194.0  
## <none> 1.5716 -2193.7  
## + ToeflScore.c:Research.c 1 0.0002523 1.5714 -2191.8  
## + CGPA.c:LOR.c 1 0.0001326 1.5715 -2191.8  
## + LOR.c:Research.c 1 0.0000592 1.5716 -2191.8  
## + GreScore.c:ToeflScore.c 1 0.0000535 1.5716 -2191.8  
## + GreScore.c:LOR.c 1 0.0000315 1.5716 -2191.8  
##   
## Step: AIC=-2195.61  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c + GreScore.c:Research.c +   
## ToeflScore.c:LOR.c  
##   
## Df Sum of Sq RSS AIC  
## - ToeflScore.c:LOR.c 1 0.0013242 1.5735 -2197.3  
## - GreScore.c:Research.c 1 0.0022265 1.5744 -2197.1  
## <none> 1.5721 -2195.6  
## - CGPA.c:Research.c 1 0.0081335 1.5803 -2195.6  
## - CGPA.c:GreScore.c 1 0.0094446 1.5816 -2195.2  
## + CGPA.c:ToeflScore.c 1 0.0005120 1.5716 -2193.7  
## + ToeflScore.c:Research.c 1 0.0004670 1.5717 -2193.7  
## + GreScore.c:LOR.c 1 0.0002596 1.5719 -2193.7  
## + CGPA.c:LOR.c 1 0.0000501 1.5721 -2193.6  
## + GreScore.c:ToeflScore.c 1 0.0000260 1.5721 -2193.6  
## + LOR.c:Research.c 1 0.0000060 1.5721 -2193.6  
##   
## Step: AIC=-2197.28  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c + GreScore.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - GreScore.c:Research.c 1 0.002325 1.5758 -2198.7  
## - CGPA.c:Research.c 1 0.007436 1.5809 -2197.4  
## <none> 1.5735 -2197.3  
## + GreScore.c:LOR.c 1 0.001432 1.5720 -2195.6  
## + ToeflScore.c:LOR.c 1 0.001324 1.5721 -2195.6  
## + CGPA.c:LOR.c 1 0.000478 1.5730 -2195.4  
## + LOR.c:Research.c 1 0.000153 1.5733 -2195.3  
## + ToeflScore.c:Research.c 1 0.000120 1.5733 -2195.3  
## + GreScore.c:ToeflScore.c 1 0.000087 1.5734 -2195.3  
## + CGPA.c:ToeflScore.c 1 0.000000 1.5735 -2195.3  
## - CGPA.c:GreScore.c 1 0.017798 1.5913 -2194.8  
## - ToeflScore.c 1 0.033803 1.6073 -2190.8  
## - LOR.c 1 0.091654 1.6651 -2176.6  
##   
## Step: AIC=-2198.69  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## <none> 1.5758 -2198.7  
## + GreScore.c:Research.c 1 0.002325 1.5735 -2197.3  
## + ToeflScore.c:LOR.c 1 0.001423 1.5744 -2197.1  
## + ToeflScore.c:Research.c 1 0.000996 1.5748 -2196.9  
## + GreScore.c:LOR.c 1 0.000890 1.5749 -2196.9  
## + CGPA.c:LOR.c 1 0.000762 1.5750 -2196.9  
## + LOR.c:Research.c 1 0.000181 1.5756 -2196.7  
## + CGPA.c:ToeflScore.c 1 0.000108 1.5757 -2196.7  
## + GreScore.c:ToeflScore.c 1 0.000000 1.5758 -2196.7  
## - CGPA.c:GreScore.c 1 0.015975 1.5917 -2196.7  
## - CGPA.c:Research.c 1 0.022840 1.5986 -2194.9  
## - ToeflScore.c 1 0.035274 1.6111 -2191.8  
## - LOR.c 1 0.091911 1.6677 -2178.0

summary (part2\_fit\_AIC)

##   
## Call:  
## lm(formula = `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c +   
## LOR.c + Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.259165 -0.022316 0.009348 0.036366 0.160108   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.7243959 0.0040496 178.879 < 2e-16 \*\*\*  
## CGPA.c 0.1208875 0.0116826 10.348 < 2e-16 \*\*\*  
## GreScore.c 0.0015541 0.0005996 2.592 0.009901 \*\*   
## ToeflScore.c 0.0031474 0.0010625 2.962 0.003240 \*\*   
## LOR.c 0.0228900 0.0047870 4.782 2.46e-06 \*\*\*  
## Research.c 0.0300528 0.0081828 3.673 0.000273 \*\*\*  
## CGPA.c:GreScore.c -0.0010408 0.0005221 -1.994 0.046897 \*   
## CGPA.c:Research.c 0.0379682 0.0159286 2.384 0.017617 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.0634 on 392 degrees of freedom  
## Multiple R-squared: 0.8058, Adjusted R-squared: 0.8023   
## F-statistic: 232.4 on 7 and 392 DF, p-value: < 2.2e-16

# Stepwise by BIC

part2\_fit\_BIC = step (part2\_fit3, direction="both", k=log (part2\_fit3$rank + part2\_fit3$df.residual))

## Start: AIC=-2120.06  
## `Chance of Admit` ~ (CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c)^2  
##   
## Df Sum of Sq RSS AIC  
## - GreScore.c:ToeflScore.c 1 0.00007604 1.5710 -2126.0  
## - LOR.c:Research.c 1 0.00011937 1.5710 -2126.0  
## - GreScore.c:LOR.c 1 0.00014584 1.5711 -2126.0  
## - CGPA.c:LOR.c 1 0.00014705 1.5711 -2126.0  
## - CGPA.c:ToeflScore.c 1 0.00023085 1.5712 -2126.0  
## - ToeflScore.c:Research.c 1 0.00029732 1.5712 -2126.0  
## - ToeflScore.c:LOR.c 1 0.00055710 1.5715 -2125.9  
## - GreScore.c:Research.c 1 0.00159256 1.5725 -2125.7  
## - CGPA.c:GreScore.c 1 0.00219024 1.5731 -2125.5  
## - CGPA.c:Research.c 1 0.00222360 1.5731 -2125.5  
## <none> 1.5709 -2120.1  
##   
## Step: AIC=-2126.03  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:LOR.c +   
## CGPA.c:Research.c + GreScore.c:LOR.c + GreScore.c:Research.c +   
## ToeflScore.c:LOR.c + ToeflScore.c:Research.c + LOR.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - LOR.c:Research.c 1 0.00011432 1.5711 -2132.0  
## - CGPA.c:ToeflScore.c 1 0.00015992 1.5712 -2132.0  
## - GreScore.c:LOR.c 1 0.00017623 1.5712 -2132.0  
## - ToeflScore.c:Research.c 1 0.00023306 1.5712 -2132.0  
## - CGPA.c:LOR.c 1 0.00024504 1.5712 -2132.0  
## - ToeflScore.c:LOR.c 1 0.00061945 1.5716 -2131.9  
## - GreScore.c:Research.c 1 0.00154088 1.5725 -2131.6  
## - CGPA.c:GreScore.c 1 0.00275651 1.5737 -2131.3  
## - CGPA.c:Research.c 1 0.00284461 1.5738 -2131.3  
## <none> 1.5710 -2126.0  
## + GreScore.c:ToeflScore.c 1 0.00007604 1.5709 -2120.1  
##   
## Step: AIC=-2131.99  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:LOR.c +   
## CGPA.c:Research.c + GreScore.c:LOR.c + GreScore.c:Research.c +   
## ToeflScore.c:LOR.c + ToeflScore.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - GreScore.c:LOR.c 1 0.0000904 1.5712 -2138.0  
## - CGPA.c:ToeflScore.c 1 0.0001670 1.5713 -2137.9  
## - CGPA.c:LOR.c 1 0.0002449 1.5714 -2137.9  
## - ToeflScore.c:Research.c 1 0.0002574 1.5714 -2137.9  
## - ToeflScore.c:LOR.c 1 0.0006842 1.5718 -2137.8  
## - GreScore.c:Research.c 1 0.0014441 1.5725 -2137.6  
## - CGPA.c:GreScore.c 1 0.0032045 1.5743 -2137.2  
## - CGPA.c:Research.c 1 0.0042300 1.5753 -2136.9  
## <none> 1.5711 -2132.0  
## + LOR.c:Research.c 1 0.0001143 1.5710 -2126.0  
## + GreScore.c:ToeflScore.c 1 0.0000710 1.5710 -2126.0  
##   
## Step: AIC=-2137.96  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:LOR.c +   
## CGPA.c:Research.c + GreScore.c:Research.c + ToeflScore.c:LOR.c +   
## ToeflScore.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - CGPA.c:LOR.c 1 0.0001686 1.5714 -2143.9  
## - ToeflScore.c:Research.c 1 0.0002883 1.5715 -2143.9  
## - CGPA.c:ToeflScore.c 1 0.0003634 1.5716 -2143.9  
## - GreScore.c:Research.c 1 0.0013538 1.5725 -2143.6  
## - ToeflScore.c:LOR.c 1 0.0015958 1.5728 -2143.6  
## - CGPA.c:Research.c 1 0.0044541 1.5756 -2142.8  
## - CGPA.c:GreScore.c 1 0.0050925 1.5763 -2142.7  
## <none> 1.5712 -2138.0  
## + GreScore.c:ToeflScore.c 1 0.0000957 1.5711 -2132.0  
## + GreScore.c:LOR.c 1 0.0000904 1.5711 -2132.0  
## + LOR.c:Research.c 1 0.0000285 1.5712 -2132.0  
##   
## Step: AIC=-2143.91  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:Research.c +   
## GreScore.c:Research.c + ToeflScore.c:LOR.c + ToeflScore.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - ToeflScore.c:Research.c 1 0.0002523 1.5716 -2149.8  
## - CGPA.c:ToeflScore.c 1 0.0002974 1.5717 -2149.8  
## - GreScore.c:Research.c 1 0.0012595 1.5726 -2149.6  
## - ToeflScore.c:LOR.c 1 0.0019524 1.5733 -2149.4  
## - CGPA.c:GreScore.c 1 0.0049846 1.5763 -2148.6  
## - CGPA.c:Research.c 1 0.0052718 1.5766 -2148.6  
## <none> 1.5714 -2143.9  
## + GreScore.c:ToeflScore.c 1 0.0001693 1.5712 -2138.0  
## + CGPA.c:LOR.c 1 0.0001686 1.5712 -2138.0  
## + LOR.c:Research.c 1 0.0000580 1.5713 -2137.9  
## + GreScore.c:LOR.c 1 0.0000141 1.5714 -2137.9  
##   
## Step: AIC=-2149.84  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:ToeflScore.c + CGPA.c:Research.c +   
## GreScore.c:Research.c + ToeflScore.c:LOR.c  
##   
## Df Sum of Sq RSS AIC  
## - CGPA.c:ToeflScore.c 1 0.0005120 1.5721 -2155.7  
## - ToeflScore.c:LOR.c 1 0.0018361 1.5735 -2155.4  
## - GreScore.c:Research.c 1 0.0026296 1.5742 -2155.2  
## - CGPA.c:GreScore.c 1 0.0062708 1.5779 -2154.2  
## - CGPA.c:Research.c 1 0.0069380 1.5785 -2154.1  
## <none> 1.5716 -2149.8  
## + ToeflScore.c:Research.c 1 0.0002523 1.5714 -2143.9  
## + CGPA.c:LOR.c 1 0.0001326 1.5715 -2143.9  
## + LOR.c:Research.c 1 0.0000592 1.5716 -2143.9  
## + GreScore.c:ToeflScore.c 1 0.0000535 1.5716 -2143.9  
## + GreScore.c:LOR.c 1 0.0000315 1.5716 -2143.8  
##   
## Step: AIC=-2155.7  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c + GreScore.c:Research.c +   
## ToeflScore.c:LOR.c  
##   
## Df Sum of Sq RSS AIC  
## - ToeflScore.c:LOR.c 1 0.0013242 1.5735 -2161.3  
## - GreScore.c:Research.c 1 0.0022265 1.5744 -2161.1  
## - CGPA.c:Research.c 1 0.0081335 1.5803 -2159.6  
## - CGPA.c:GreScore.c 1 0.0094446 1.5816 -2159.3  
## <none> 1.5721 -2155.7  
## + CGPA.c:ToeflScore.c 1 0.0005120 1.5716 -2149.8  
## + ToeflScore.c:Research.c 1 0.0004670 1.5717 -2149.8  
## + GreScore.c:LOR.c 1 0.0002596 1.5719 -2149.8  
## + CGPA.c:LOR.c 1 0.0000501 1.5721 -2149.7  
## + GreScore.c:ToeflScore.c 1 0.0000260 1.5721 -2149.7  
## + LOR.c:Research.c 1 0.0000060 1.5721 -2149.7  
##   
## Step: AIC=-2161.35  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c + GreScore.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - GreScore.c:Research.c 1 0.002325 1.5758 -2166.8  
## - CGPA.c:Research.c 1 0.007436 1.5809 -2165.5  
## - CGPA.c:GreScore.c 1 0.017798 1.5913 -2162.8  
## <none> 1.5735 -2161.3  
## - ToeflScore.c 1 0.033803 1.6073 -2158.8  
## + GreScore.c:LOR.c 1 0.001432 1.5720 -2155.7  
## + ToeflScore.c:LOR.c 1 0.001324 1.5721 -2155.7  
## + CGPA.c:LOR.c 1 0.000478 1.5730 -2155.5  
## + LOR.c:Research.c 1 0.000153 1.5733 -2155.4  
## + ToeflScore.c:Research.c 1 0.000120 1.5733 -2155.4  
## + GreScore.c:ToeflScore.c 1 0.000087 1.5734 -2155.4  
## + CGPA.c:ToeflScore.c 1 0.000000 1.5735 -2155.4  
## - LOR.c 1 0.091654 1.6651 -2144.7  
##   
## Step: AIC=-2166.75  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - CGPA.c:GreScore.c 1 0.015975 1.5917 -2168.7  
## - CGPA.c:Research.c 1 0.022840 1.5986 -2167.0  
## <none> 1.5758 -2166.8  
## - ToeflScore.c 1 0.035274 1.6111 -2163.9  
## + GreScore.c:Research.c 1 0.002325 1.5735 -2161.3  
## + ToeflScore.c:LOR.c 1 0.001423 1.5744 -2161.1  
## + ToeflScore.c:Research.c 1 0.000996 1.5748 -2161.0  
## + GreScore.c:LOR.c 1 0.000890 1.5749 -2161.0  
## + CGPA.c:LOR.c 1 0.000762 1.5750 -2161.0  
## + LOR.c:Research.c 1 0.000181 1.5756 -2160.8  
## + CGPA.c:ToeflScore.c 1 0.000108 1.5757 -2160.8  
## + GreScore.c:ToeflScore.c 1 0.000000 1.5758 -2160.8  
## - LOR.c 1 0.091911 1.6677 -2150.1  
##   
## Step: AIC=-2168.71  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:Research.c  
##   
## Df Sum of Sq RSS AIC  
## - CGPA.c:Research.c 1 0.009082 1.6008 -2172.4  
## <none> 1.5917 -2168.7  
## + CGPA.c:GreScore.c 1 0.015975 1.5758 -2166.8  
## - ToeflScore.c 1 0.032106 1.6239 -2166.7  
## - GreScore.c 1 0.033689 1.6254 -2166.3  
## + CGPA.c:ToeflScore.c 1 0.012084 1.5797 -2165.8  
## + GreScore.c:ToeflScore.c 1 0.010438 1.5813 -2165.3  
## + GreScore.c:LOR.c 1 0.010398 1.5814 -2165.3  
## + ToeflScore.c:LOR.c 1 0.009292 1.5825 -2165.1  
## + CGPA.c:LOR.c 1 0.008402 1.5834 -2164.8  
## + GreScore.c:Research.c 1 0.000502 1.5913 -2162.8  
## + LOR.c:Research.c 1 0.000402 1.5914 -2162.8  
## + ToeflScore.c:Research.c 1 0.000398 1.5914 -2162.8  
## - LOR.c 1 0.094164 1.6859 -2151.7  
##   
## Step: AIC=-2172.43  
## `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c  
##   
## Df Sum of Sq RSS AIC  
## <none> 1.6008 -2172.4  
## - ToeflScore.c 1 0.03292 1.6338 -2170.3  
## - GreScore.c 1 0.03638 1.6372 -2169.4  
## - Research.c 1 0.03912 1.6400 -2168.8  
## + CGPA.c:Research.c 1 0.00908 1.5917 -2168.7  
## + GreScore.c:Research.c 1 0.00744 1.5934 -2168.3  
## + ToeflScore.c:Research.c 1 0.00724 1.5936 -2168.2  
## + CGPA.c:GreScore.c 1 0.00222 1.5986 -2167.0  
## + GreScore.c:ToeflScore.c 1 0.00196 1.5989 -2166.9  
## + GreScore.c:LOR.c 1 0.00185 1.5990 -2166.9  
## + ToeflScore.c:LOR.c 1 0.00176 1.5991 -2166.9  
## + LOR.c:Research.c 1 0.00163 1.5992 -2166.8  
## + CGPA.c:ToeflScore.c 1 0.00134 1.5995 -2166.8  
## + CGPA.c:LOR.c 1 0.00085 1.6000 -2166.7  
## - LOR.c 1 0.09133 1.6922 -2156.2  
## - CGPA.c 1 0.43201 2.0328 -2082.8

summary (part2\_fit\_BIC)

##   
## Call:  
## lm(formula = `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c +   
## LOR.c + Research.c)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.263542 -0.023297 0.009879 0.038078 0.159897   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.7243500 0.0031871 227.276 < 2e-16 \*\*\*  
## CGPA.c 0.1210042 0.0117349 10.312 < 2e-16 \*\*\*  
## GreScore.c 0.0017820 0.0005955 2.992 0.00294 \*\*   
## ToeflScore.c 0.0030320 0.0010651 2.847 0.00465 \*\*   
## LOR.c 0.0227762 0.0048039 4.741 2.97e-06 \*\*\*  
## Research.c 0.0245769 0.0079203 3.103 0.00205 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06374 on 394 degrees of freedom  
## Multiple R-squared: 0.8027, Adjusted R-squared: 0.8002   
## F-statistic: 320.6 on 5 and 394 DF, p-value: < 2.2e-16

## Results from AIC & BIC

summary (part2\_fit\_AIC)

##   
## Call:  
## lm(formula = `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c +   
## LOR.c + Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.259165 -0.022316 0.009348 0.036366 0.160108   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.7243959 0.0040496 178.879 < 2e-16 \*\*\*  
## CGPA.c 0.1208875 0.0116826 10.348 < 2e-16 \*\*\*  
## GreScore.c 0.0015541 0.0005996 2.592 0.009901 \*\*   
## ToeflScore.c 0.0031474 0.0010625 2.962 0.003240 \*\*   
## LOR.c 0.0228900 0.0047870 4.782 2.46e-06 \*\*\*  
## Research.c 0.0300528 0.0081828 3.673 0.000273 \*\*\*  
## CGPA.c:GreScore.c -0.0010408 0.0005221 -1.994 0.046897 \*   
## CGPA.c:Research.c 0.0379682 0.0159286 2.384 0.017617 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.0634 on 392 degrees of freedom  
## Multiple R-squared: 0.8058, Adjusted R-squared: 0.8023   
## F-statistic: 232.4 on 7 and 392 DF, p-value: < 2.2e-16

summary (part2\_fit\_BIC)

##   
## Call:  
## lm(formula = `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c +   
## LOR.c + Research.c)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.263542 -0.023297 0.009879 0.038078 0.159897   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.7243500 0.0031871 227.276 < 2e-16 \*\*\*  
## CGPA.c 0.1210042 0.0117349 10.312 < 2e-16 \*\*\*  
## GreScore.c 0.0017820 0.0005955 2.992 0.00294 \*\*   
## ToeflScore.c 0.0030320 0.0010651 2.847 0.00465 \*\*   
## LOR.c 0.0227762 0.0048039 4.741 2.97e-06 \*\*\*  
## Research.c 0.0245769 0.0079203 3.103 0.00205 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06374 on 394 degrees of freedom  
## Multiple R-squared: 0.8027, Adjusted R-squared: 0.8002   
## F-statistic: 320.6 on 5 and 394 DF, p-value: < 2.2e-16

anova(part2\_fit\_AIC,part2\_fit\_BIC)

## Analysis of Variance Table  
##   
## Model 1: `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c  
## Model 2: `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c + LOR.c +   
## Research.c  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 392 1.5758   
## 2 394 1.6008 -2 -0.025058 3.1168 0.0454 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

We conclude from the ANOVA table that the two models are not similar since the p-value of the f-test is 0.0454 which is less than 0.05 and therefore we reject the null hypothesis which says that there is no statistical difference between the two models.

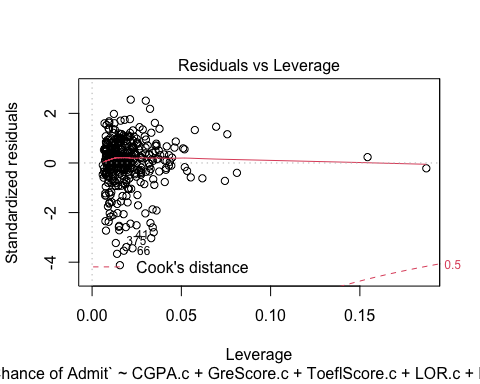
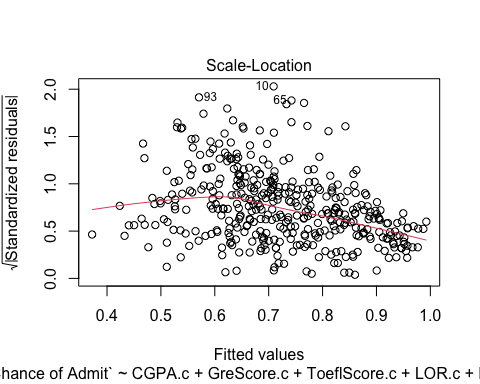
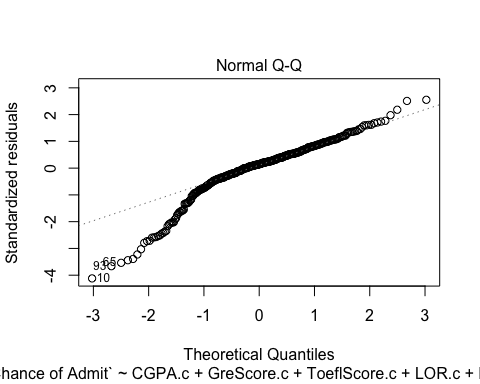
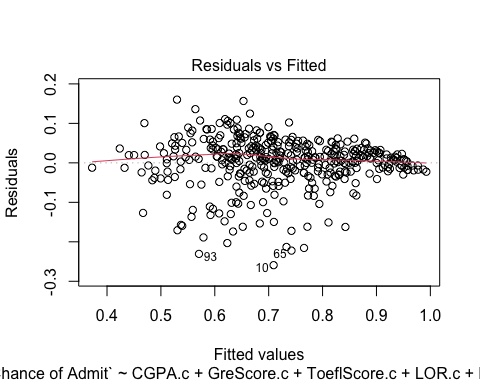
The the Adjusted R-squared value and the Residual Standard error for both the AIC & BIC Model are as follows:

|  |  |  |
| --- | --- | --- |
| Model | Adjusted R-square | Residual Standard error |
| AIC | 0.8023 | 0.0634 |
| BIC | 0.8002 | 0.06374 |

The AIC model is the best model because it has a higher Adjusted R-squared value and a lower Residual Standard error.

## Residual Diagnostics for Final Model

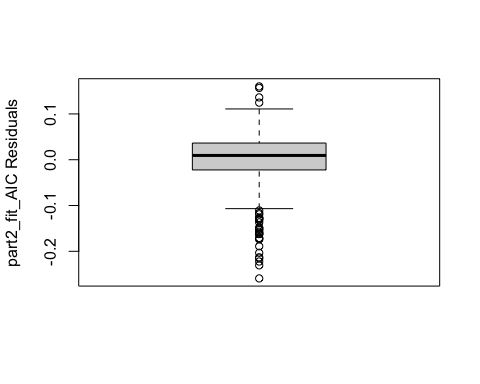
#par (mfrow = c(1, 2))  
plot (part2\_fit\_AIC)



1. The linearity assumption is reasonable in the Residual vs. Fitted plot, however there is a non constant spread.
2. The Q-Q plot shows that there is some deviation from the lower tale, this might affect the normality assumption.
3. The scale location plot also shows a decrease in the residuals variance with the decrease of fitted values.
4. The Residuals vs Leverage plot does show some concerns, and there are two leverage points and three potential outliers.

# Box plot of residuals

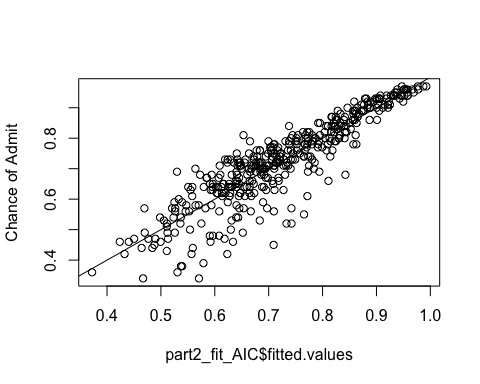
boxplot(resid(part2\_fit\_AIC), ylab="part2\_fit\_AIC Residuals")



The box-plot of residuals looks symmetric and there some outliers on the lower tale #(Ask professor about box-plot Question )and three in the upper.

# b. Plot response variable vs fitted values, add a line with intercept 0 and slope 1

plot(`Chance of Admit`~part2\_fit\_AIC$fitted.values)  
abline(0,1)



The plot of the response variable chance of Admit vs.fitted values shows a good fit of the model to the data.

# Variance Inflation

car::vif(part2\_fit\_AIC)

## CGPA.c GreScore.c ToeflScore.c LOR.c   
## 4.817218 4.697499 4.127852 1.836170   
## Research.c CGPA.c:GreScore.c CGPA.c:Research.c   
## 1.650650 1.536635 1.591094

All of the VIF are less than 5, which means that multicollinearity is low in our model.

# Influence analysis

Save the residuals and the leverage values in the data frame

Admission\_Predict$Residual = round (part2\_fit\_AIC$residuals, 4)  
Admission\_Predict$leverage = round (hatvalues(part2\_fit\_AIC), 4)  
Admission\_Predict$rstudent = round (rstudent(part2\_fit\_AIC), 4)

Pull out the cases with high leverage or a large residual

high.levg.resd = Admission\_Predict [Admission\_Predict$leverage > 2\*(part2\_fit\_AIC$rank) / (part2\_fit\_AIC$rank + part2\_fit\_AIC$df.residual) |  
 abs (part2\_fit\_AIC$residuals) > 30 , ]  
  
high.levg.resd[order(-high.levg.resd$rstudent),][c(1,9:12)]

## # A tibble: 27 x 5  
## `Serial No.` `Chance of Admit` Residual leverage rstudent  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 169 0.64 0.0893 0.0695 1.46   
## 2 57 0.64 0.0817 0.0576 1.33   
## 3 53 0.78 0.0706 0.0758 1.16   
## 4 44 0.87 0.0535 0.0469 0.864  
## 5 30 0.54 0.0416 0.0418 0.670  
## 6 56 0.64 0.0383 0.0462 0.618  
## 7 29 0.46 0.0363 0.051 0.588  
## 8 298 0.86 0.034 0.0518 0.551  
## 9 149 0.96 0.0213 0.0404 0.343  
## 10 80 0.46 0.0195 0.0431 0.314  
## # … with 17 more rows

The three highest leverage points bigger than the Leverage Cutoff value for 2(k+1)/n which is at 0.04 for this data. Row 59 has the highest leverage at 0.187 followed by row 48 at 0.154 and row 79 at 0.0811. These points are the highest from the mean

high.rstudntLowEnd = Admission\_Predict [Admission\_Predict$rstudent < -3.0 & is.na(Admission\_Predict$rstudent) == FALSE, ]  
high.rstudntLowEnd[order(high.rstudntLowEnd$rstudent),][c(1,9:12)]

## # A tibble: 7 x 5  
## `Serial No.` `Chance of Admit` Residual leverage rstudent  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10 0.45 -0.259 0.0156 -4.21  
## 2 93 0.34 -0.231 0.014 -3.72  
## 3 65 0.52 -0.222 0.0178 -3.59  
## 4 66 0.55 -0.216 0.0226 -3.49  
## 5 11 0.52 -0.213 0.0192 -3.44  
## 6 60 0.42 -0.203 0.0133 -3.27  
## 7 375 0.39 -0.189 0.0331 -3.06

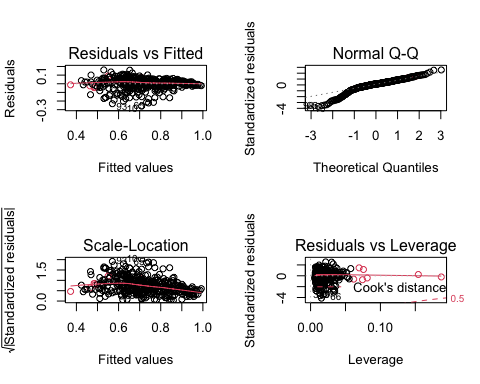
The three lowest studentized residual points lower than the -3.0 are Row 10 has the lowest studentized residual at -4.21 followed by row 93 at -3.72 and row 65 at -3.59 These points have the highest left tail deviation from the Normal distribution. It's important to point out only 7 observations have an studentized residual lower than -3.0

high.rstudntHighEnd = Admission\_Predict [Admission\_Predict$rstudent > 2.0, ]  
high.rstudntHighEnd[order(-high.rstudntHighEnd$rstudent),][c(1,9:12)]

## # A tibble: 3 x 5  
## `Serial No.` `Chance of Admit` Residual leverage rstudent  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 328 0.69 0.160 0.0217 2.57  
## 2 360 0.81 0.157 0.0301 2.53  
## 3 359 0.7 0.136 0.0325 2.19

There are no observations that have studentized residuals higher than 3.0.

# Residual plots   
  
par (mfrow=c(2,2))  
plot (part2\_fit\_AIC, col= (Admission\_Predict$leverage > 0.06)+1)



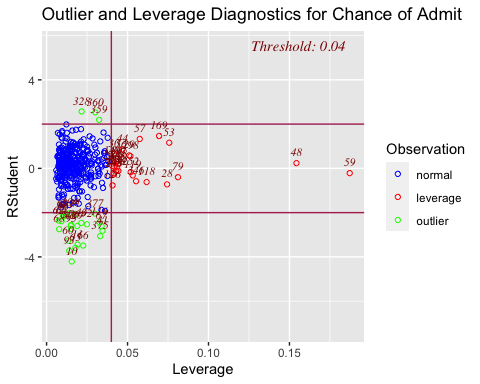
library(olsrr)

##   
## Attaching package: 'olsrr'

## The following object is masked from 'package:MASS':  
##   
## cement

## The following object is masked from 'package:datasets':  
##   
## rivers

ols\_plot\_resid\_lev(part2\_fit\_AIC, print\_plot = TRUE)



This plot illustrates that there are no points that have both high leverage and have studentized residuals lower than -3 or higher than 3. And it shows as indicated earlier that there are 7 outlier points with studentized residuals lower than -3 namely row 10,93,65,66, 11, 60 & 375.

Now we will see whether any of those outliers are influential points using Cook's Distance.

Admission\_Predict$cookd = round (cooks.distance(part2\_fit\_AIC), 4)

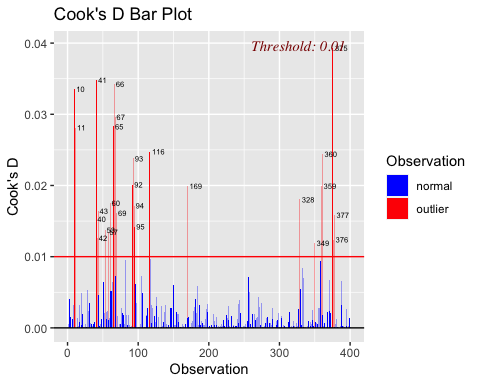
high.cooksDveryInf = Admission\_Predict [Admission\_Predict$cookd >1.0 , ]  
high.cooksDModerateInf = Admission\_Predict [Admission\_Predict$cookd >0.5 , ]  
highest.cooksDinData = Admission\_Predict [Admission\_Predict$cookd >0.03 , ]

There are no cook's distance bigger than 0.5. Therefore no influential points.

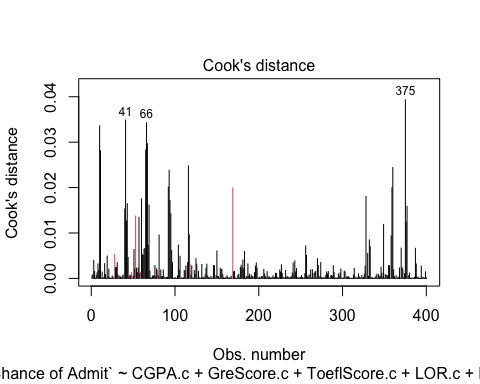
highest.cooksDinData[order(-highest.cooksDinData$cookd),][c(1,9:13)]

## # A tibble: 4 x 6  
## `Serial No.` `Chance of Admit` Residual leverage rstudent cookd  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 375 0.39 -0.189 0.0331 -3.06 0.0393  
## 2 41 0.46 -0.174 0.0345 -2.82 0.0348  
## 3 66 0.55 -0.216 0.0226 -3.49 0.0343  
## 4 10 0.45 -0.259 0.0156 -4.21 0.0336

ols\_plot\_cooksd\_bar(part2\_fit\_AIC, print\_plot = TRUE)



# Residual plots   
  
#par (mfrow=c(1,2))  
plot (part2\_fit\_AIC, which=4, col= (Admission\_Predict$leverage > 0.06)+1)



The all the outlier rows with studentized residuals less than -3.0 have a cook's distance less than 0.5 and therefore neither of them are influential. Row 375 has the highest cook's distance at 0.0393.

Our residual and influential analysis shows no outliers that are influential points therefore our model does not need any transformation.

1. Interpretation
2. Meaning/interpretation of regression parameters
3. Qualitative, directional statements can be acceptable
4. Note: Parameter estimates of individual predictor variables that are involved in an interaction effect or a quadratic effect are not interpret able – Plot and summarize the interaction or quadratic effect from the plot, instead
5. Make some example response predictions with confidence intervals and interpret

summary(part2\_fit\_AIC)

##   
## Call:  
## lm(formula = `Chance of Admit` ~ CGPA.c + GreScore.c + ToeflScore.c +   
## LOR.c + Research.c + CGPA.c:GreScore.c + CGPA.c:Research.c)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.259165 -0.022316 0.009348 0.036366 0.160108   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.7243959 0.0040496 178.879 < 2e-16 \*\*\*  
## CGPA.c 0.1208875 0.0116826 10.348 < 2e-16 \*\*\*  
## GreScore.c 0.0015541 0.0005996 2.592 0.009901 \*\*   
## ToeflScore.c 0.0031474 0.0010625 2.962 0.003240 \*\*   
## LOR.c 0.0228900 0.0047870 4.782 2.46e-06 \*\*\*  
## Research.c 0.0300528 0.0081828 3.673 0.000273 \*\*\*  
## CGPA.c:GreScore.c -0.0010408 0.0005221 -1.994 0.046897 \*   
## CGPA.c:Research.c 0.0379682 0.0159286 2.384 0.017617 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.0634 on 392 degrees of freedom  
## Multiple R-squared: 0.8058, Adjusted R-squared: 0.8023   
## F-statistic: 232.4 on 7 and 392 DF, p-value: < 2.2e-16

The intercept which is 0.7243959 does not have any meaning because a student does not apply they will not have any chance of admission.

# Conclusions

Our model produces the following formula for admitting students

Our final model is:

For every increase in 1 CGPA point the chance of admission for a student would increase by 12.08875%. Students who do research have a 3% higher chance of admission than students who do not. In addition, a point increase in a student's LOR (Letter of Recommendation) increases their chance of admission by 2.29%. This is surprisingly higher than a point's increase in any of the standardized test scores. A point increase in a student's TOEFL increases their chance of admission by 0.31% which is 0.16% than a point increase in GRE Score, that makes the TOEFL score more important for Grad School admission than a GRE score, that is surprising because TOEFL is just a language competency test.

The final model has Adjusted R square= 0.8023, which means that 80% of the variation on chance of admission is explained by the model. The residual standard error is 0.0634.

# ii. Make some example response predictions with confidence intervals and interpret those results.

predch= predict (part2\_fit\_AIC, interval = 'confidence')  
head(predch)

## fit lwr upr  
## 1 0.9497144 0.9316223 0.9678065  
## 2 0.8072569 0.7936853 0.8208285  
## 3 0.6438988 0.6257228 0.6620747  
## 4 0.7418424 0.7271054 0.7565794  
## 5 0.6392735 0.6270956 0.6514513  
## 6 0.8641715 0.8482569 0.8800861

Admission\_Predict$pred.ChanceOfAdmission = predch[,1]  
Admission\_Predict$pred.lower = predch[,2]  
Admission\_Predict$pred.upper = predch[,3]  
  
  
  
head(Admission\_Predict)

## # A tibble: 6 x 16  
## `Serial No.` `GRE Score` `TOEFL Score` `University Rat… SOP LOR CGPA  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 337 118 4 4.5 4.5 9.65  
## 2 2 324 107 4 4 4.5 8.87  
## 3 3 316 104 3 3 3.5 8   
## 4 4 322 110 3 3.5 2.5 8.67  
## 5 5 314 103 2 2 3 8.21  
## 6 6 330 115 5 4.5 3 9.34  
## # … with 9 more variables: Research <dbl>, `Chance of Admit` <dbl>,  
## # Residual <dbl>, leverage <dbl>, rstudent <dbl>, cookd <dbl>,  
## # pred.ChanceOfAdmission <dbl>, pred.lower <dbl>, pred.upper <dbl>

head(Admission\_Predict)

## # A tibble: 6 x 16  
## `Serial No.` `GRE Score` `TOEFL Score` `University Rat… SOP LOR CGPA  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 337 118 4 4.5 4.5 9.65  
## 2 2 324 107 4 4 4.5 8.87  
## 3 3 316 104 3 3 3.5 8   
## 4 4 322 110 3 3.5 2.5 8.67  
## 5 5 314 103 2 2 3 8.21  
## 6 6 330 115 5 4.5 3 9.34  
## # … with 9 more variables: Research <dbl>, `Chance of Admit` <dbl>,  
## # Residual <dbl>, leverage <dbl>, rstudent <dbl>, cookd <dbl>,  
## # pred.ChanceOfAdmission <dbl>, pred.lower <dbl>, pred.upper <dbl>

Admission\_Predict[c(115,19,47,66,96,375,215),c(1,9,14:16)]

## # A tibble: 7 x 5  
## `Serial No.` `Chance of Admit` pred.ChanceOfAdmission pred.lower pred.upper  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 115 0.59 0.690 0.677 0.703  
## 2 19 0.63 0.727 0.712 0.743  
## 3 47 0.86 0.901 0.888 0.914  
## 4 66 0.55 0.766 0.747 0.784  
## 5 96 0.42 0.557 0.544 0.569  
## 6 375 0.39 0.579 0.556 0.602  
## 7 215 0.94 0.927 0.912 0.942

Admission\_Predict$in.interval= ifelse (Admission\_Predict$pred.lower <= Admission\_Predict$pred.ChanceOfAdmission &  
 Admission\_Predict$pred.ChanceOfAdmission <= Admission\_Predict$pred.upper,1,0)  
mean(Admission\_Predict$in.interval)

## [1] 1

Among the seven example Serial NO whose prediction are shown above, all the predicted chance of admission fall between the lower and upper prediction. However, among all of the Serial NO in the data set, 100% have prediction intervals that contain the observed serial NO.We will expect 100%.