Predicting Graduate Student Perception of Admission Chances

Shiyu Murashima, Andrew Chen, Connie Ma, Daniel Kao, Rebekah Limb, Zoey Meng
Group 11 - Lec 1 - Section 1B

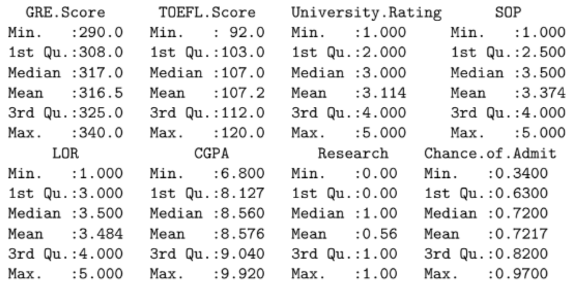
Introduction

The source of our data comes from students who are aspiring to obtain a Masters degree in the United States. Inspired by the UCLA Graduate Dataset: Engineering Students and created with the purpose of helping students in shortlisting potential universities, applicants reported relevant personal stats, such as Graduate Record Examination (GRE) and Cumulative Undergraduate GPA (CGPA), and were asked how confident they felt about their own admit chances. Other predictors include Test of English as a Foreign Language (TOEFL) score, University Rating, Statement of Purpose (SOP) and Letter of Recommendation (LOR) Strength, and Research Experience.

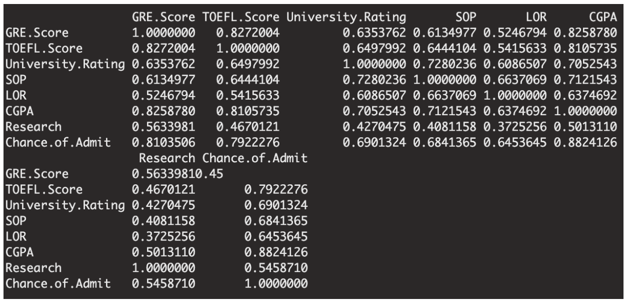
Our group was interested in how various factors such as test scores and grades would affect one's confidence in chances of admission. With the expectation that students with overall higher parameters would more favorably rate their chance of admit, we wanted to explore the validity of this assumption. Our final multiple linear regression model is a reduced model with Y transformed and the SOP predictor removed. In this report, we'll cover problematic errors such as multicollinearity, different transformation models tested, variable selection, and how we arrived at our final model. Additionally, we discuss how our model is relevant in a real world context, overall findings, and limitations of our analysis within our dataset.

Data Description

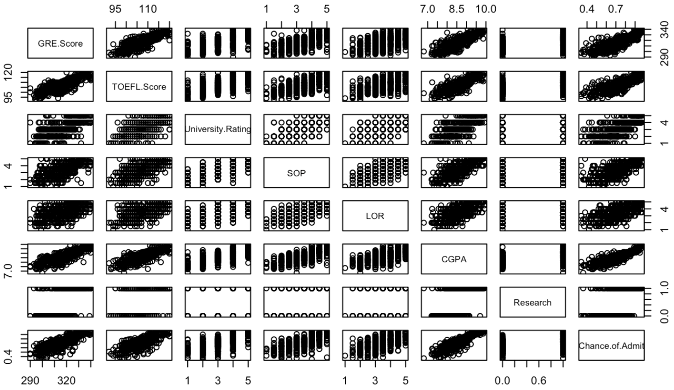
Variable	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
Std Dev	11.295	6.082	1.144	0.991	0.925	0.605	0.497	0.141



```
  GRE.Score      TOEFL.Score    University.Rating      SOP
 Min.   :290.0  Min.   : 92.0  Min.   :1.000     Min.   :1.000
 1st Qu.:308.0  1st Qu.:103.0  1st Qu.:2.000     1st Qu.:2.500
 Median :317.0  Median :107.0  Median :3.000     Median :3.500
 Mean   :316.5  Mean   :107.2  Mean   :3.114     Mean   :3.374
 3rd Qu.:325.0  3rd Qu.:112.0  3rd Qu.:4.000     3rd Qu.:4.000
 Max.   :340.0  Max.   :120.0  Max.   :5.000     Max.   :5.000
      LOR            CGPA          Research    Chance.of.Admit
 Min.   :1.000  Min.   :6.800  Min.   :0.00   Min.   :0.3400
 1st Qu.:3.000  1st Qu.:8.127  1st Qu.:0.00   1st Qu.:0.6300
 Median :3.500  Median :8.560  Median :1.00   Median :0.7200
 Mean   :3.484  Mean   :8.576  Mean   :0.56   Mean   :0.7217
 3rd Qu.:4.000  3rd Qu.:9.040  3rd Qu.:1.00   3rd Qu.:0.8200
 Max.   :5.000  Max.   :9.920  Max.   :1.00   Max.   :0.9700
```

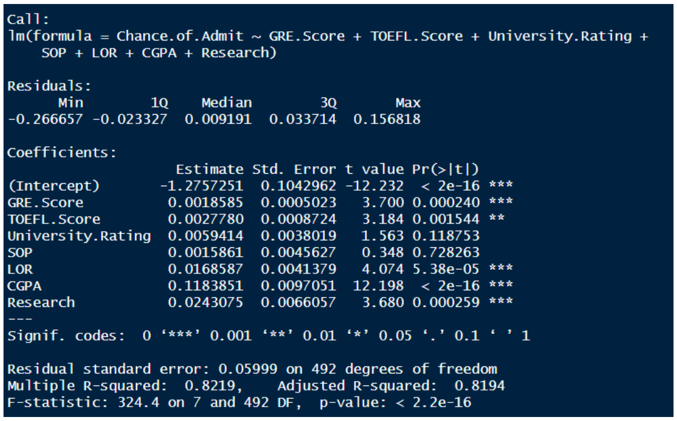


```
                  GRE.Score TOEFL.Score University.Rating       SOP       LOR      CGPA
GRE.Score         1.0000000   0.8272004         0.6353762 0.6134977 0.5246794 0.8258780
TOEFL.Score       0.8272004   1.0000000         0.6497992 0.6444104 0.5415633 0.8105735
University.Rating 0.6353762   0.6497992         1.0000000 0.7280236 0.6086507 0.7052543
SOP               0.6134977   0.6444104         0.7280236 1.0000000 0.6637069 0.7121543
LOR               0.5246794   0.5415633         0.6086507 0.6637069 1.0000000 0.6374692
CGPA              0.8258780   0.8105735         0.7052543 0.7121543 0.6374692 1.0000000
Research          0.5633981   0.4670121         0.4270475 0.4081158 0.3725256 0.5013110
Chance.of.Admit   0.8103506   0.7922276         0.6901324 0.6841365 0.6453645 0.8824126
                   Research Chance.of.Admit
GRE.Score         0.56339810.45
TOEFL.Score       0.4670121       0.7922276
University.Rating 0.4270475       0.6901324
SOP               0.4081158       0.6841365
LOR               0.3725256       0.6453645
CGPA              0.5013110       0.8824126
Research          1.0000000       0.5458710
Chance.of.Admit   0.5458710       1.0000000
```

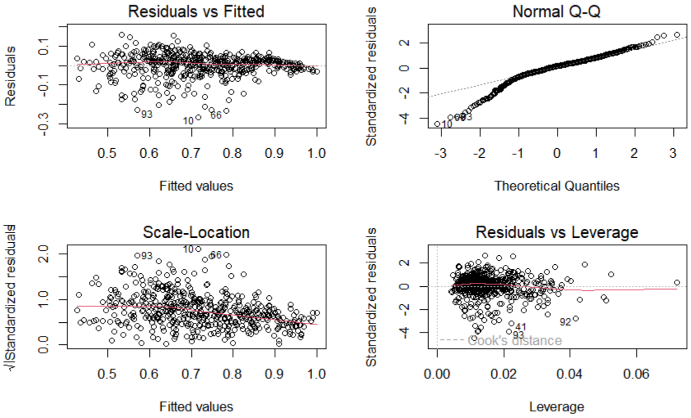


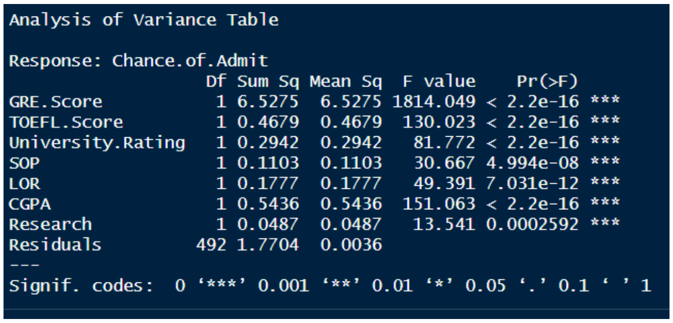
Our scatterplot matrix shows that many of the variables seem to have a linear correlation, however, both with our target variable, chance of admit, as well as with each other. A linear association with our target variable is good as it allows us to predict, but in the case of the variables being linearly associated with each other, we would need to conduct further tests to account for multicollinearity to see if our model results will be negatively impacted. We also notice that Research is a categorical variable of 0 and 1, and University Rating, SOP, and LOR only take discrete values between 1 and 5 with 1 being the worst and 5 being the best. Based on our correlation matrix, we do see that there are many highly correlated variables. This proves what we noticed in the scatterplots and further suggests that using things such as the VIF to account for multicollinearity is needed.

Results & Interpretation



```
Call:
lm(formula = Chance.of.Admit ~ GRE.Score + TOEFL.Score + University.Rating +
    SOP + LOR + CGPA + Research)

Residuals:
      Min        1Q    Median        3Q       Max
-0.266657 -0.023327  0.009191  0.033714  0.156818

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -1.2757251  0.1042962 -12.232  < 2e-16 ***
GRE.Score          0.0018585  0.0005023   3.700 0.000240 ***
TOEFL.Score        0.0027780  0.0008724   3.184 0.001544 **
University.Rating  0.0059414  0.0038019   1.563 0.118753
SOP                0.0015861  0.0045627   0.348 0.728263
LOR                0.0168587  0.0041379   4.074 5.38e-05 ***
CGPA               0.1183851  0.0097051  12.198  < 2e-16 ***
Research           0.0243075  0.0066057   3.680 0.000259 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05999 on 492 degrees of freedom
Multiple R-squared:  0.8219,    Adjusted R-squared:  0.8194
F-statistic: 324.4 on 7 and 492 DF,  p-value: < 2.2e-16
```





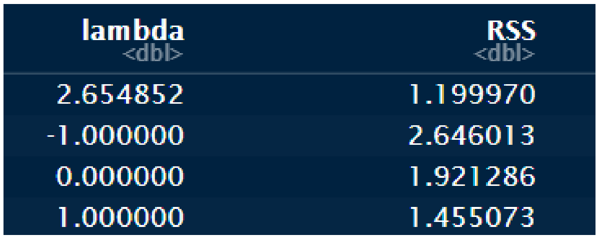
```
Analysis of Variance Table

Response: Chance.of.Admit
                   Df Sum Sq Mean Sq  F value    Pr(>F)
GRE.Score           1 6.5275  6.5275 1814.049 < 2.2e-16 ***
TOEFL.Score         1 0.4679  0.4679  130.023 < 2.2e-16 ***
University.Rating   1 0.2942  0.2942   81.772 < 2.2e-16 ***
SOP                 1 0.1103  0.1103   30.667 4.994e-08 ***
LOR                 1 0.1777  0.1777   49.391 7.031e-12 ***
CGPA                1 0.5436  0.5436  151.063 < 2.2e-16 ***
Research            1 0.0487  0.0487   13.541 0.0002592 ***
Residuals         492 1.7704  0.0036
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Our first full model without any transformations or variable selection has fairly high predictive power with an $R^2 = 0.8194$. With this model, our regression equation was:

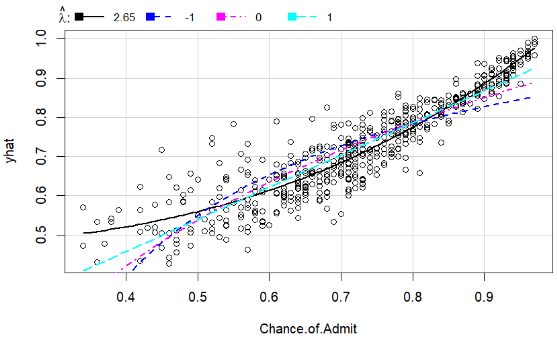
Chance of Admit = -1.2757 + 0.0019*GRE + 0.0028*TOEFL + 0.0059*University Rating + 0.0016*SOP + 0.0017*LOR + 0.1184*CGPA + 0.0243*Research

The residual plots look like mostly random scatter around 0, but the Q-Q plot does not follow a straight line meaning the errors are not entirely normal and there are some bad leverage points. Overall, the model is not entirely invalid, but there are some signs that more investigation is needed.

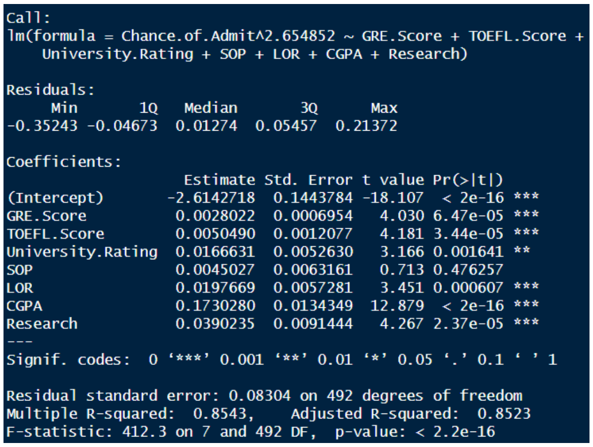
We tried doing some transformations to improve the fit and validity of the model, starting first with an inverse response plot to transform the response variable.



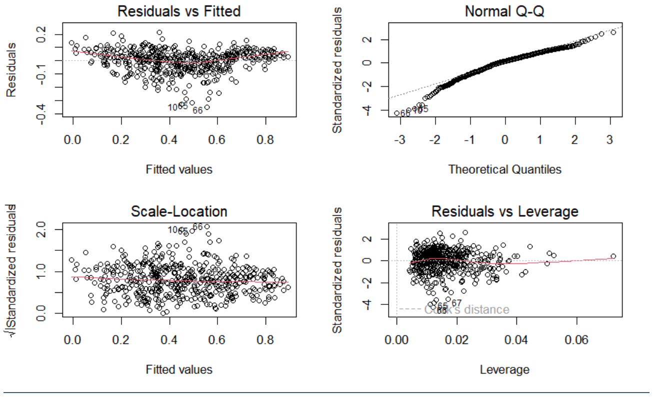
lambda <dbl>	RSS <dbl>
2.654852	1.199970
-1.000000	2.646013
0.000000	1.921286
1.000000	1.455073



Based on these results, we made a new model transforming y:

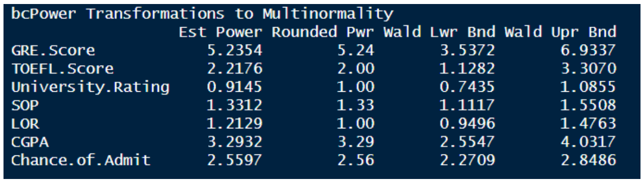


```
Call:
lm(formula = Chance.of.Admit^2.654852 ~ GRE.Score + TOEFL.Score +
    University.Rating + SOP + LOR + CGPA + Research)

Residuals:
     Min       1Q   Median       3Q      Max
-0.35243 -0.04673  0.01274  0.05457  0.21372

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -2.6142718  0.1443784 -18.107  < 2e-16 ***
GRE.Score          0.0028022  0.0006954   4.030 6.47e-05 ***
TOEFL.Score        0.0050490  0.0012077   4.181 3.44e-05 ***
University.Rating  0.0166631  0.0052630   3.166 0.001641 **
SOP                0.0045027  0.0063161   0.713 0.476257
LOR                0.0197669  0.0057281   3.451 0.000607 ***
CGPA               0.1730280  0.0134349  12.879  < 2e-16 ***
Research           0.0390235  0.0091444   4.267 2.37e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08304 on 492 degrees of freedom
Multiple R-squared:  0.8543,    Adjusted R-squared:  0.8523
F-statistic: 412.3 on 7 and 492 DF,  p-value: < 2.2e-16
```

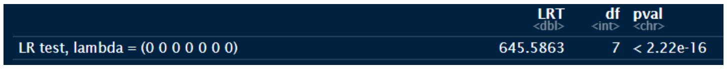


The R^2 of this model was better with R^2 = 0.8523, suggesting that this model has better fit. Additionally, the diagnostic plots look better since there are less bad leverage points, the residual plots are still random scatter around 0, and the errors look more normally distributed. Overall, this transformation looks successful.

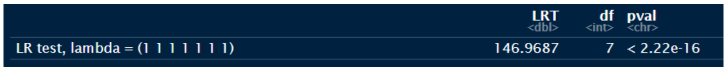
We then tried to transform both X and Y using the Box Cox powertransform:



```
bcPower Transformations to Multinormality
                  Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
GRE.Score            5.2354        5.24       3.5372       6.9337
TOEFL.Score          2.2176        2.00       1.1282       3.3070
University.Rating    0.9145        1.00       0.7435       1.0855
SOP                  1.3312        1.33       1.1117       1.5508
LOR                  1.2129        1.00       0.9496       1.4763
CGPA                 3.2932        3.29       2.5547       4.0317
Chance.of.Admit      2.5597        2.56       2.2709       2.8486
```

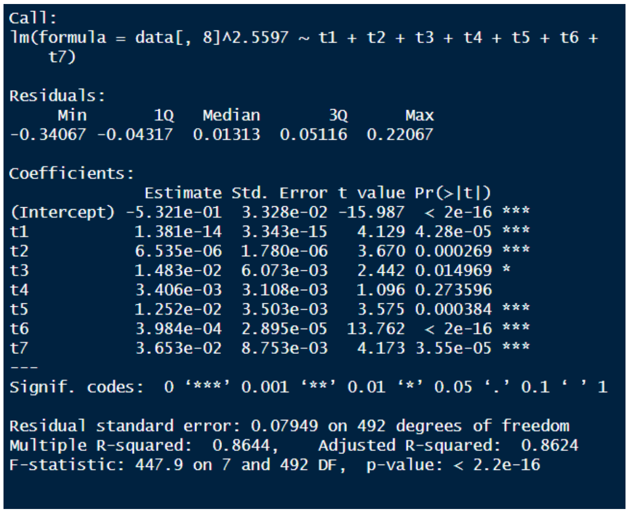


	LRT <dbl>	df <int>	pval <chr>
LR test, lambda = (0 0 0 0 0 0 0)	645.5863	7	< 2.22e-16

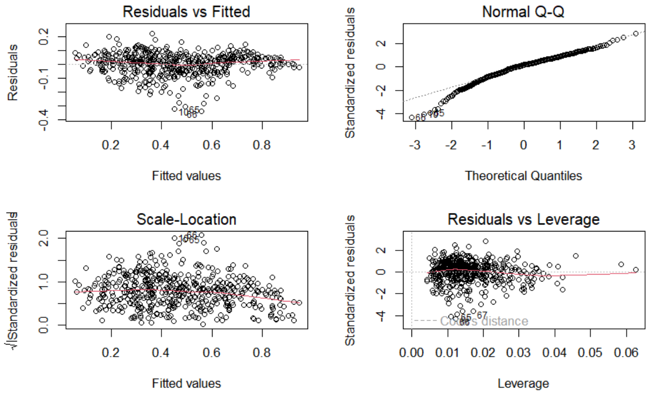


	LRT <dbl>	df <int>	pval <chr>
LR test, lambda = (1 1 1 1 1 1 1)	146.9687	7	< 2.22e-16

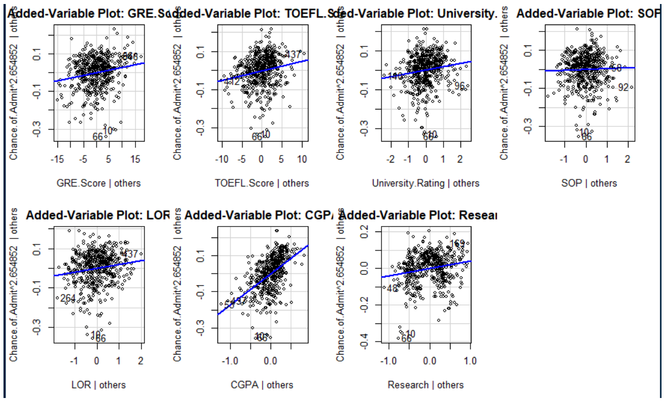
We constructed a model using these transformations.



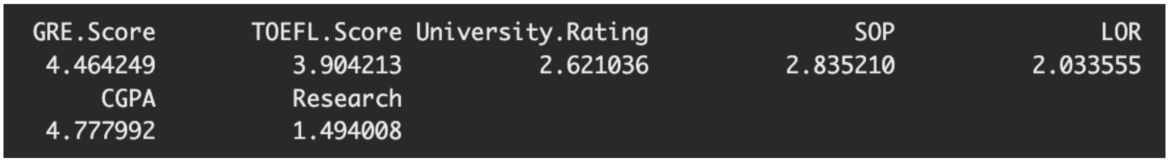
```
Call:
lm(formula = data[, 8]^2.5597 ~ t1 + t2 + t3 + t4 + t5 + t6 +
    t7)

Residuals:
     Min       1Q   Median       3Q      Max
-0.34067 -0.04317  0.01313  0.05116  0.22067

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.321e-01  3.328e-02 -15.987  < 2e-16 ***
t1           1.381e-14  3.343e-15   4.129 4.28e-05 ***
t2           6.535e-06  1.780e-06   3.670 0.000269 ***
t3           1.483e-02  6.073e-03   2.442 0.014969 *
t4           3.406e-03  3.108e-03   1.096 0.273596
t5           1.252e-02  3.503e-03   3.575 0.000384 ***
t6           3.984e-04  2.895e-05  13.762  < 2e-16 ***
t7           3.653e-02  8.753e-03   4.173 3.55e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07949 on 492 degrees of freedom
Multiple R-squared:  0.8644,    Adjusted R-squared:  0.8624
F-statistic: 447.9 on 7 and 492 DF,  p-value: < 2.2e-16
```



The R^2 of this model was again slightly better with R^2 = 0.8674, suggesting that this model has better fit. However, the diagnostic plots had no visible improvement over the previous transformation. Therefore, we decided to only use the transformation in the response variable to help with interpretability and decrease complexity. Transforming X as well only marginally improved the fit of the model.



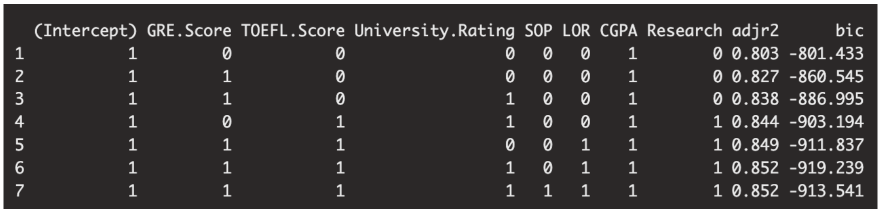
For variable selection, we first looked at the added-variable plots of each variable to see which ones should be investigated.
All the slopes besides SOP were significant, which implies that SOP should be investigated. It could either be meaningless to the model or have multicollinearity with another variable, which we saw before as a possibility when looking at the overall data.



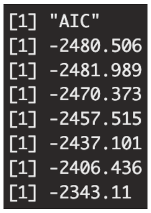
```
GRE.Score       TOEFL.Score University.Rating               SOP               LOR 
 4.464249          3.904213          2.621036          2.835210          2.033555 
     CGPA          Research 
 4.777992          1.494008 
```

To check for multicollinearity, we checked the VIFs of the variables and found that none of the VIFs for the predictor variables were greater than 5, so we can conclude that there is no multicollinearity between our variables.

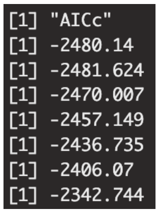
Nonetheless, because the SOP predictor variable was insignificant in our AV-plots, we conducted variable selection and tested goodness of fit by considering all possible subsets.



```
  (Intercept) GRE.Score TOEFL.Score University.Rating SOP LOR CGPA Research adjr2      bic
1           1         0           0                 0   0   0    1        0 0.803 -801.433
2           1         1           0                 0   0   0    1        0 0.827 -860.545
3           1         1           0                 1   0   0    1        0 0.838 -886.995
4           1         0           1                 1   0   0    1        1 0.844 -903.194
5           1         1           1                 0   0   1    1        1 0.849 -911.837
6           1         1           1                 1   0   1    1        1 0.852 -919.239
7           1         1           1                 1   1   1    1        1 0.852 -913.541
```



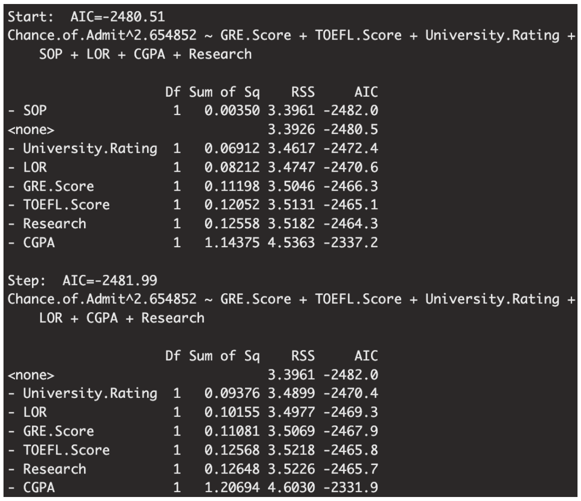
```
[1] "AIC"
[1] -2480.506
[1] -2481.989
[1] -2470.373
[1] -2457.515
[1] -2437.101
[1] -2406.436
[1] -2343.11
```



```
[1] "AICc"
[1] -2480.14
[1] -2481.624
[1] -2470.007
[1] -2457.149
[1] -2436.735
[1] -2406.07
[1] -2342.744
```

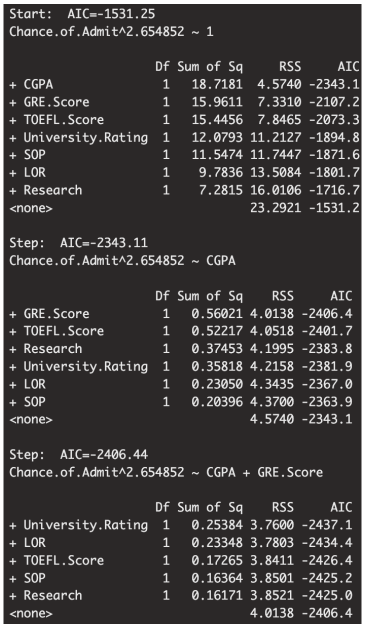
With an Adjusted $R^2 = 0.852$, this suggests that either the full model or a reduced model with SOP removed were the best fitting models. BIC, AIC, and AIC corrected with respective values -919.239, -2406.436, and -2342.744, suggest that the best model is just with SOP removed.

Furthermore, we also used forward and backward stepwise regression to check the AIC.

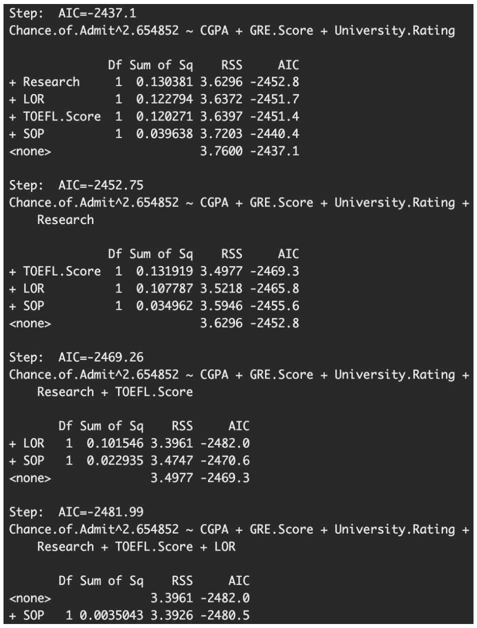


```
Start:  AIC=-2480.51
Chance.of.Admit^2.654852 ~ GRE.Score + TOEFL.Score + University.Rating + 
    SOP + LOR + CGPA + Research

                    Df Sum of Sq    RSS     AIC
- SOP                1   0.00350 3.3961 -2482.0
<none>                           3.3926 -2480.5
- University.Rating  1   0.06912 3.4617 -2472.4
- LOR                1   0.08212 3.4747 -2470.6
- GRE.Score          1   0.11198 3.5046 -2466.3
- TOEFL.Score        1   0.12052 3.5131 -2465.1
- Research           1   0.12558 3.5182 -2464.3
- CGPA               1   1.14375 4.5363 -2337.2

Step:  AIC=-2481.99
Chance.of.Admit^2.654852 ~ GRE.Score + TOEFL.Score + University.Rating + 
    LOR + CGPA + Research

                    Df Sum of Sq    RSS     AIC
<none>                           3.3961 -2482.0
- University.Rating  1   0.09376 3.4899 -2470.4
- LOR                1   0.10155 3.4977 -2469.3
- GRE.Score          1   0.11081 3.5069 -2467.9
- TOEFL.Score        1   0.12568 3.5218 -2465.8
- Research           1   0.12648 3.5226 -2465.7
- CGPA               1   1.20694 4.6030 -2331.9
```

Backward stepwise led to an AIC suggesting the removal of SOP again.



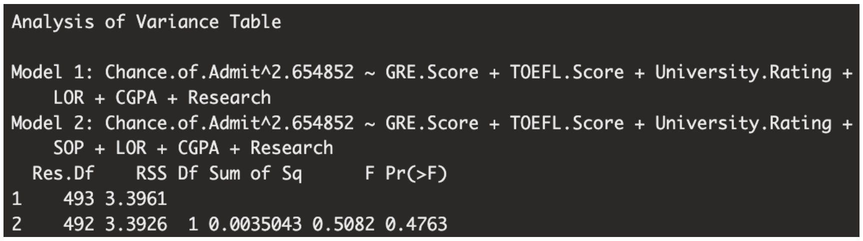
```
Start:  AIC=-1531.25
Chance.of.Admit^2.654852 ~ 1

                   Df Sum of Sq     RSS     AIC
+ CGPA              1   18.7181  4.5740 -2343.1
+ GRE.Score         1   15.9611  7.3310 -2107.2
+ TOEFL.Score       1   15.4456  7.8465 -2073.3
+ University.Rating 1   12.0793 11.2127 -1894.8
+ SOP               1   11.5474 11.7447 -1871.6
+ LOR               1    9.7836 13.5084 -1801.7
+ Research          1    7.2815 16.0106 -1716.7
<none>                          23.2921 -1531.2

Step:  AIC=-2343.11
Chance.of.Admit^2.654852 ~ CGPA

                   Df Sum of Sq    RSS     AIC
+ GRE.Score         1   0.56021 4.0138 -2406.4
+ TOEFL.Score       1   0.52217 4.0518 -2401.7
+ Research          1   0.37453 4.1995 -2383.8
+ University.Rating 1   0.35818 4.2158 -2381.9
+ LOR               1   0.23050 4.3435 -2367.0
+ SOP               1   0.20396 4.3700 -2363.9
<none>                          4.5740 -2343.1

Step:  AIC=-2406.44
Chance.of.Admit^2.654852 ~ CGPA + GRE.Score

                   Df Sum of Sq    RSS     AIC
+ University.Rating 1   0.25384 3.7600 -2437.1
+ LOR               1   0.23348 3.7803 -2434.4
+ TOEFL.Score       1   0.17265 3.8411 -2426.4
+ SOP               1   0.16364 3.8501 -2425.2
+ Research          1   0.16171 3.8521 -2425.0
<none>                          4.0138 -2406.4
```



```
Step:  AIC=-2437.1
Chance.of.Admit^2.654852 ~ CGPA + GRE.Score + University.Rating

              Df Sum of Sq    RSS     AIC
+ Research     1  0.130381 3.6296 -2452.8
+ LOR          1  0.122794 3.6372 -2451.7
+ TOEFL.Score  1  0.120271 3.6397 -2451.4
+ SOP          1  0.039638 3.7203 -2440.4
<none>                     3.7600 -2437.1

Step:  AIC=-2452.75
Chance.of.Admit^2.654852 ~ CGPA + GRE.Score + University.Rating +
    Research

              Df Sum of Sq    RSS     AIC
+ TOEFL.Score  1  0.131919 3.4977 -2469.3
+ LOR          1  0.107787 3.5218 -2465.8
+ SOP          1  0.034962 3.5946 -2455.6
<none>                     3.6296 -2452.8

Step:  AIC=-2469.26
Chance.of.Admit^2.654852 ~ CGPA + GRE.Score + University.Rating +
    Research + TOEFL.Score

       Df Sum of Sq    RSS     AIC
+ LOR   1  0.101546 3.3961 -2482.0
+ SOP   1  0.022935 3.4747 -2470.6
<none>              3.4977 -2469.3

Step:  AIC=-2481.99
Chance.of.Admit^2.654852 ~ CGPA + GRE.Score + University.Rating +
    Research + TOEFL.Score + LOR

       Df Sum of Sq    RSS     AIC
<none>              3.3961 -2482.0
+ SOP   1 0.0035043 3.3926 -2480.5
```

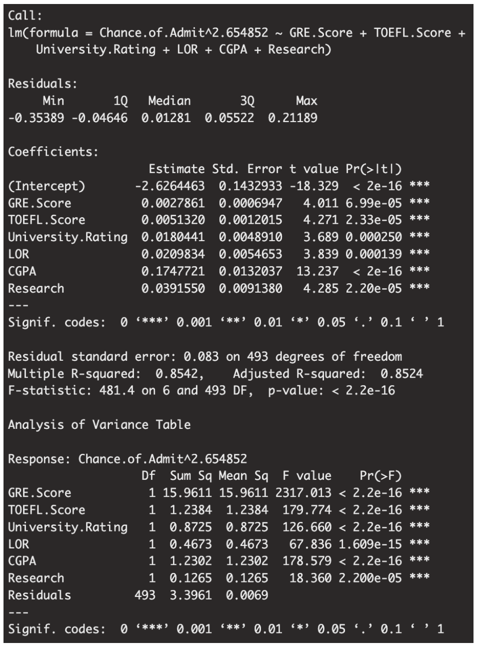
Forward stepwise also led to an AIC suggesting the removal of SOP.

With the exception of Adjusted R^2 which also suggested the removal of no predictors, each goodness of fit criteria we tested suggested the removal of the SOP predictor. Thus, we decided to perform a partial F-test with the SOP predictor removed.

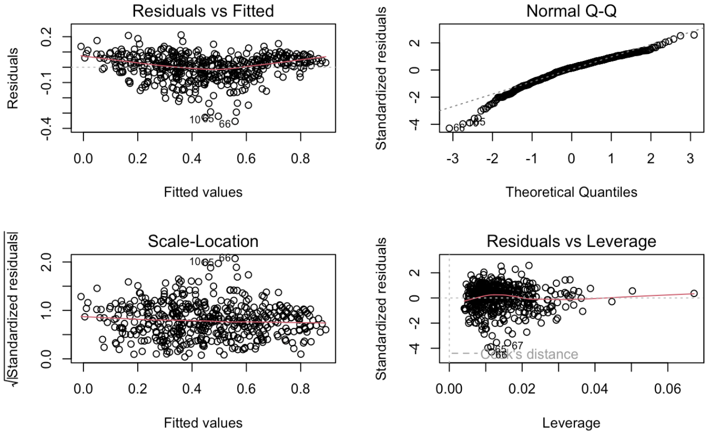


```
Analysis of Variance Table

Model 1: Chance.of.Admit^2.654852 ~ GRE.Score + TOEFL.Score + University.Rating +
    LOR + CGPA + Research
Model 2: Chance.of.Admit^2.654852 ~ GRE.Score + TOEFL.Score + University.Rating +
    SOP + LOR + CGPA + Research
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    493 3.3961
2    492 3.3926  1 0.0035043 0.5082 0.4763
```

We can see from the ANOVA table above that removing the SOP predictor is sensible as the p-value is 0.4763, which is greater than 0.05, and thus does not show a statistically significant result between the full model and reduced model in predicting confidence of chance of admission. Therefore, we will remove the SOP predictor as it can reduce our model complexity.



```
Call:
lm(formula = Chance.of.Admit^2.654852 ~ GRE.Score + TOEFL.Score +
    University.Rating + LOR + CGPA + Research)

Residuals:
     Min       1Q   Median       3Q      Max
-0.35389 -0.04646  0.01281  0.05522  0.21189

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.6264463  0.1432933 -18.329  < 2e-16 ***
GRE.Score         0.0027861  0.0006947   4.011 6.99e-05 ***
TOEFL.Score       0.0051320  0.0012015   4.271 2.33e-05 ***
University.Rating  0.0180441  0.0048910   3.689 0.000250 ***
LOR               0.0209834  0.0054653   3.839 0.000139 ***
CGPA              0.1747721  0.0132037  13.237  < 2e-16 ***
Research          0.0391550  0.0091380   4.285 2.20e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.083 on 493 degrees of freedom
Multiple R-squared:  0.8542,    Adjusted R-squared:  0.8524
F-statistic: 481.4 on 6 and 493 DF,  p-value: < 2.2e-16

Analysis of Variance Table

Response: Chance.of.Admit^2.654852
                   Df  Sum Sq Mean Sq  F value    Pr(>F)
GRE.Score           1 15.9611 15.9611 2317.013 < 2.2e-16 ***
TOEFL.Score         1  1.2384  1.2384  179.774 < 2.2e-16 ***
University.Rating   1  0.8725  0.8725  126.660 < 2.2e-16 ***
LOR                 1  0.4673  0.4673   67.836 1.609e-15 ***
CGPA                1  1.2302  1.2302  178.579 < 2.2e-16 ***
Research            1  0.1265  0.1265   18.360 2.200e-05 ***
Residuals         493  3.3961  0.0069
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



Our final linear regression equation is thus:

Chance of Admit$^{2.6549}$ = -1.2757 + 0.0019*GRE + 0.0028*TOEFL + 0.0059*University Rating + 0.0016*LOR + 0.1184*CGPA + 0.0243*Research

The final model we chose is a reduced model with Y transformed and the SOP predictor removed. Compared to our first full model, our Adjusted R^2 value has improved slightly from 0.8194 to 0.8524. Our diagnostic plots have also slightly improved. The line of the Residuals vs Fitted model is straight around 0, meaning that the relationship between the predictors and response variable is linear. The line in the Normal Q-Q plot is also very slightly more aligned with the points now, meaning that the errors are likely more normal. The scatter in the Scale-Location model has become slightly more random, so the variance is likely constant. Lastly, the Residuals vs Leverage plot now has less bad leverage points.

Discussion

In summary, our study sought out to conduct a predictive model of master's program student applications to determine how factors: GRE score, TOEFL score, undergraduate university rating, statement of purpose, letter of recommendation rating, GPA, research experience, could affect one's confidence of chances of admission.

As for our main findings of this study, since not all of our predictors were recorded on the same scale it was not meaningful to compare their effects. However, for all the individual predictors, all score based, except the research categorical variable, higher values were considered better scores. In turn, the 6 predictor variables we used in the model all had positive coefficients which would make sense in the real world context of having higher scores and a stronger application leading to higher confidence.

It was interesting to us to discover that the statement of purpose variable (SOP) was not significant enough to contribute to the linear model. In the real world situation, one could hypothesize that this factor, since it is self-written and a subjective personal measure, may not affect one's confidence in admissions as much as objective scores like the GRE or GPA does.

But overall, our main finding was the significant positive linear relationship among the selected predictors and the confidence of admission. This was to be expected, as generally you can find in graduate school application requirements, that higher GPA, more research experience, etc. make a stronger application, and thus a stronger confidence in admission for the applicant.

As for the limitations of our analysis, due to the lack of detail in how the data was collected we didn't know the entire context behind the data. For example, it was unclear how certain subjective scores such as letter of recommendation rating or undergraduate school rating were determined, and thus it is less reliable as a model to make predictions based on those variables. So for further work, this study could be improved by attaining more descriptive data to make more meaningful conclusions about predicting confidence of admission.

Works Cited

Acharya, Mohan S. “Graduate Admission 2.” Kaggle, December 28, 2018. https://www.kaggle.com/datasets/mohansacharya/graduate-admissions?select=Admission_Predict_Ver1.1.csv.