**ACADEMIC PERFORMANCES AND CHANCES OF GRADUATE ADMISSION**

Submitted to:

Statistics 467

July 13, 2018

**I. ABSTRACT**

Purpose: The purpose of this study is to investigate the association between chances of admission of Indian Americans, who finished their undergraduate study in the U.S., into graduate programs and graduate school applicants’ academic performances in colleges and on official exams like GRE or TOEFL. Chances of admission is the probability for a student to be accepted into a graduate program. This study helps people to understand what the important factors are that affect their chance of admission into graduate programs. The results of this study can help applicants to know what factors they need to work on more in order to increase their chance of admission.  
 Method: The evaluation of students’ academic performances in colleges is defined by varying aspects, including GPA, GRE scores, TOEFL scores, and strength of statement of purpose and letter of recommendation. Part of the data was collected by surveying the Indian-American applicants on what their scores on each aspect were and whether they were accepted or not with those metrics. The bivariate relationship between chances of admission into graduate programs and Indian-American students’ performances in colleges are investigated with univariate analysis by residual diagnostic plot, ANOVA, and simultaneous confidence intervals. Likelihood ratio tests, stepwise regression, and factor analysis were chosen for multivariate analysis.   
 Results: According to the scatter plots and histograms that shows each factor and the chance of admission of each applicant together, the students in the lower chances of admission appears to have relatively lower GPA, GRE scores, TOEFL scores, strength of statement of purpose, and letter of recommendation. On the contrary to this, the students who had higher chances of admission appears to have relatively higher GPA, GRE scores, TOEFL scores, strength of statement of purpose, and letter of recommendation. However, after the GPA, GRE scores, TOEFL scores, strength of statement of purpose and letter of recommendation reached certain point, the chance of acceptance stayed very similar.   
 Discussion: The study of the relationship between academic performances and the chance of admission to graduate programs is able to imply the preferences of admission committee when reviewing thousands of applications, therefore, offering future applicants a direction about which factors should they put most efforts on to improve. Even though this dataset is from the perspective of Indian American students, the results of this study could be combined with that of Europe and provide a bigger and more general data.

Key words: GRE, TOEFL, SOP, LOR, GPA, Graduation, Admission.

**II. INTRODUCTION**

The time has changed and the present work fields want and prefer people with more than just with Bachelor’s degree. People now go deeper in their chosen field of study and strengthen their knowledge. In order to gain more knowledge in a field and to be recognized, many people decide to continue their education in graduate schools with systematic programs. Therefore, the admission of graduate programs has become more competitive. To increase their chance to be accepted into graduate programs, applicants need to know what things they need to improve on.  
 To begin the test, some data were collected from the survey of the second generation of Indian-American college students who applied to graduate programs. The GRE scores, TOEFL score, college rating, strength of their statement of purpose and letter of recommendation of Indian undergraduate students, were collected. So, each factor and the acceptance rate were calculated to find any association between the chance of graduate school admissions of a college student and the academic performance. We proclaimed and tested in the following section the predictors, the academic performance of the college Indian students, have a positive correlation with the dependent variable, which is the chance of graduate school admissions of a college student, based on the preponderance of evidence.   
 To help people to understand why GPA or grade point average is important in the perspective of admission committees, Tara Kuther, a professor at graduate programs in Lehman college, CUNY, Iona College, Fordham University, and Teachers College, Columbia University, explained that not because it signifies your intelligence, but because it is a long-term indicator how well you perform your job as a student. Grade reflect your motivation and your ability to do consistently good or bad work (Kuther et al, 2018), which could also imply your performance of graduate study.

According to this article from the official website of National Center for Biotechnology Information, a study on the correlation of admissions and some quantifiable factors, like GPA, performance on the program exit exam, performance in oral review sessions, and faculty rating, was done by some professors. In the article, it says that the quantitative GRE score was correlated with student performance in some courses and with the exit exam. It is important to note that a high degree of selection on undergraduate GPA and quantitative GRE score takes place through the admissions process. (Burmeister et al, 2014) From their study, the authors of the article also saw the important role that GPA and GRE scores play during the admission process. Depends on the type of the graduate programs, there could be more aspects that the admission committee require, but GPA and GRE scores are always required by most admission committees.

Also, there is a research done by Ann L. Mullen, that emphasizes the influence of GPA. In the article, it states that for all types of programs, GPA is a strong determinant of continuation into graduate education. Students who are the most likely to enroll in graduate education of any kind are those who performed well in college. (Mullen et al, 2003) Also, GPA has a different impact varying from different types of graduation programs. From master’s programs to doctoral program, the effects of GPA increase in strength. Each one decile increase in GPA raises one’s odds of entering a master’s program by 13 percent, an MBA program by 20 percent, and a doctoral program by 37 percent (Mullen et al, 2003).

Also, most graduate programs have minimum requirements and applicants need to meet or be better than the minimum requirements. On the website of graduate program at the university of Texas at Austin, it says that to be eligible for admission consideration, applicants must meet the Graduate School’s minimum requirements. (“University of Texas at Austin”, n.d.) The minimum requirements of the university of Texas at Austin, are Bachelor’s degree, 3.0 GPA, official test scores like GRE, GMAT, TOEFL, IELTS, and additional program requirements. By looking at the required minimum GPA is not really low, it is clear that GPA is one of the many important factor that affects the chance of admission into a graduate program and the chance of admission increases with higher GPA.

Along with GPA and official test scores, essays are very important. The reason why graduate schools ask for a personal statement is to get to know about the applicant, to see if you can think logically, to see if you can write well, and to determine if the applicant is likely to succeed. (Sokolove et al, 2018)Also, Harvard University emphasizes the importance of personal essay. One of the article from the university says that a personal Statement should be a story describing how your personality, motivations, and experiences all come together to make you a great candidate for the program. Your statement should explain how you will contribute positively and effectively to the specific department, and to the academic discipline itself. (“Harvard University”, n.d.)

When other aspects that are needed when applying to a graduate school are all numbers that indicates your academic ability, personal statements is a great way to really express applicant’s personality, experiences, and imply of his or her contribution to the community. According to an article from *Huffingtonpost,* the admission rate for international students was 3 percent, considerably lower than the domestic student admission rate of 10.8 percent. (Cohen, 2013) In order to stand out in many piles of applications, applicants should try to express themselves on their personal statements.

**Research Question Statement:**

1. Is there any relationship, and if there is, what is the relationship, between the chances of admission for an Indian American student to a graduate program and his or her performance on GRE Score, TOEFL Score, University Rating, Statement of Purpose Strength, Letter of Recommendation Strength, and Undergraduate GPA?

-There is a positive correlation between the chances of admission and each of the performance on GPA, GRE scores, TOEFL scores, statement of purpose strength, letter of recommendation strength. As each aspect of the applicants’ academic performance increases, the chance of acceptance into a graduate program increases. However, after each aspect reaches certain point, the chance of admission is similar.

1. Does the academic performance of an Indian American student in University measured by Letter of Recommendation Strength influence his or her chance of getting admission?

-Yes, the applicants with higher strength of letter of recommendation tend to have higher chance of admission than the applicants with lower strength of letter of recommendation

**III. METHODS**

1. **Measures of Variables**

In order to help students shortlist universities with their profiles, this dataset owned by Mohan S Acharya is created to predict the graduate admissions from the perspective of Indian students, which is inspired by the UCLA Graduate Dataset. The dataset aims to output predictions that gives students a fair idea about their chances for a certain university. There are 400 observations in the dataset with the unit of analysis being students.

Our object is to characterize the dependent variable Chance of Admission (CoA). Along with the value of Chance of Admit for 400 students, we had been provided with the value of seven independent variables: Serial No., GRE Score (GS), Toefl Score (TS), University Rating (UR), Statement of Purpose Strength (SOP), Letter of Recommendation Strength (LOR), and Undergraduate GPA (GPA). The GRE scores and GPA are in the older format.

**Dataset**: the data has been entered into the computer and printed out. The data has been checked for accuracy and has been verified to be the same as the data provided to us.

**Chance of Admission (CoA):** the continuous dependent variable. The probability for a student to be accepted by a graduate program. Part of the data was collected by surveying the applicants how sure they were of getting an admission in term of percentage, while adding an extra decimal to increase the accuracy. Regression methods were used to achieve the rest of the data. Has an average of 0.72, standard deviation of 0.14, ranges from a minimum value of 0.34 to a maximum of 0.97. The shape of the distribution appears unimodal skewed to the left.

**Undergraduate GPA (GPA):** an independent variable. The student’s normalized grade point average, on a scale of 1 to 10, was taken from the student’s college transcript. These scores were then grouped into deciles. Has an average of 8.60, standard deviation of 0.60, ranges from a minimum value of 6.8 to a maximum value of 9.92. The shape of the distribution appears unimodal and slightly skewed to the left

**Letter of Recommendation Strength (LOR):** a discrete independent variable. Rating the strength of the content of recommendation letter from 1 to 5. Has an average of 3.45, standard deviation of 0.89, ranges from a minimum value of 1 to a maximum value of 5. The shape of the distribution appears unimodal.

**University Rating (UR)**: a discrete independent variable. Using QS Star rating system ranges from 1 to 5 that provides a detailed look at an institution to identify which universities the best qualified in certain field. Has an average of 3.09, standard deviation of 1.14, ranges from a minimum value of 1 to 5. The shape of the distribution appears unimodal but disconnected between bars.

**Serial No.:** a discrete independent variable as an identifier for each observation Indian American student. Marking the observations from 1 to 400. Has an average of 200.5, standard deviation of 115.61, ranges from a minimum value of 1 to 400. The shape of the distribution appears uniform.

**Statement of Purpose Strength (SOP):** a discrete independent variable. Rating the strength of the content of statement of purpose from 1 to 5. Has an average of 3.4, standard deviation of 1.00, ranges from a minimum value of 1 to a maximum value of 5. The shape of the distribution appears unimodal.

**TOEFL Scores:** a discrete independent variable. Collected by surveying the TOEFL scores of the applicants. Has an average of 107.41, standard deviation of 6.07, ranges from a minimum value of 92 to a maximum value of 102. The shape of the distribution appears unimodal.

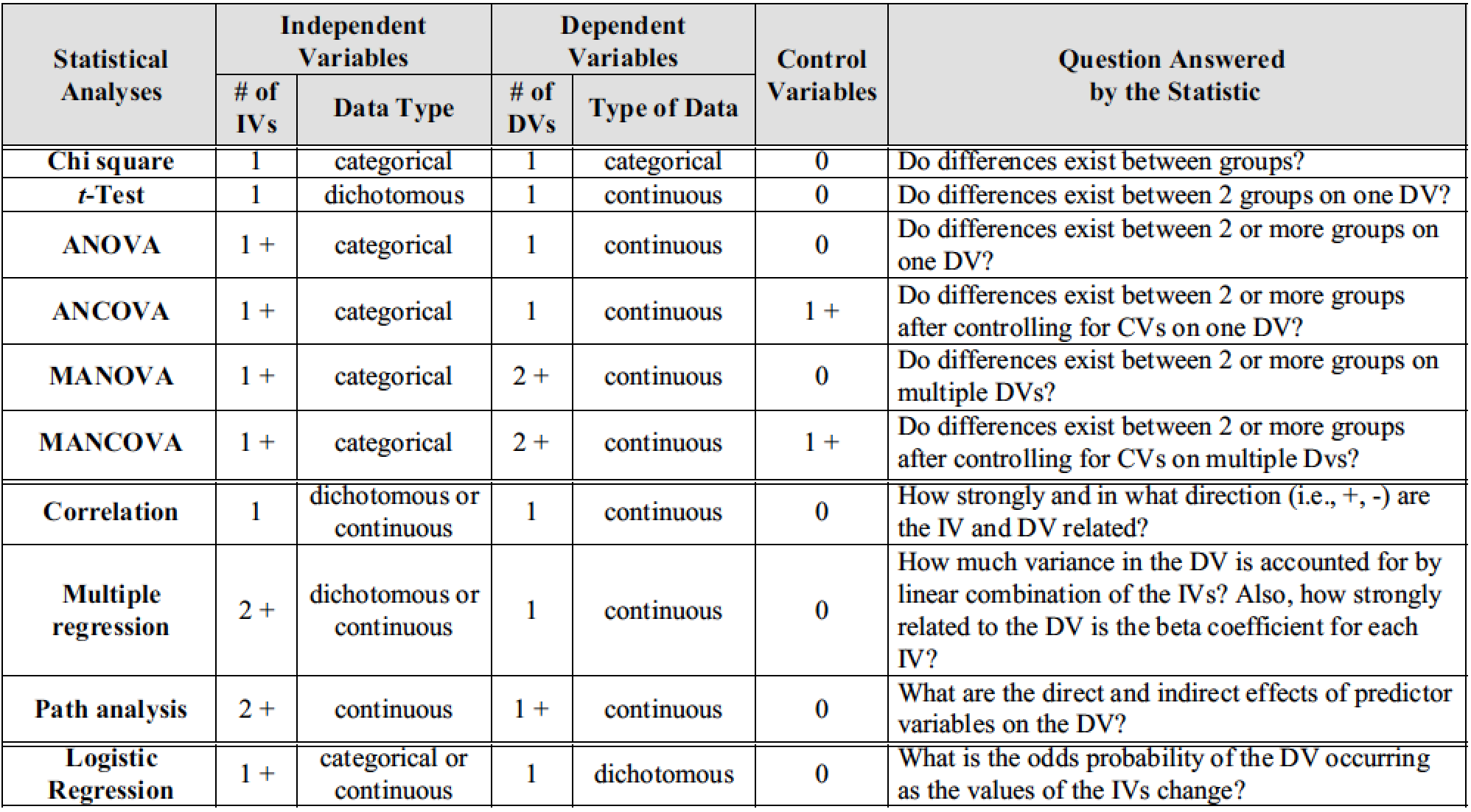
**GRE Scores**: a discrete independent variable. Collected by surveying the graduate record examination scores of the applicants. Has an average of 316.81, standard deviation of 11.47, ranges from a minimum value of 290 to a maximum value of 340. The shape of the distribution appears unimodal.

1. **Type of bi-multi-variate statistics selected**

Correlation, and multiple regression were chosen for multivariate analysis, and ANOVA were chosen for bivariate analysis.

**c. Illustration for the selected bi-multi-variate statistics**

Figure(c). *Flowchart of multivariate statistics 2015*



*Note.* Data for multivariate statistics

From the flowchart of statistical analysis, we could see that since we have continuous and categorical independent variables and continuous dependent variable, we tended to choose correlation, and multiple regression to for our multivariate analysis. For bivariate analysis, we were looking for the relationship between Letter of Recommendation Strength and chances of admission, which had one categorical independent variable and one continuous dependent variable, ANOVA were chosen for bivariate analysis based on the flowchart.

In addition, it is necessary to test the distribution of the selected variables through dot plots and histograms. Since the type of statistics to be performed are based on the distribution of variables. For instance, a histogram could reveal the distribution of variables is normal or non-normal, leading to the further discussion on whether the transformation of variable is necessary. In addition, dot plots could help to indicate the necessity of removing outliers to increase the normality of variable distribution.

**Figure/Charts/Graphs for univariate analysis**

**i. Dot Plot**

Most of the dot plots for continuous variables contain no obvious outliers, and indicate distributions that are close to normal distribution.

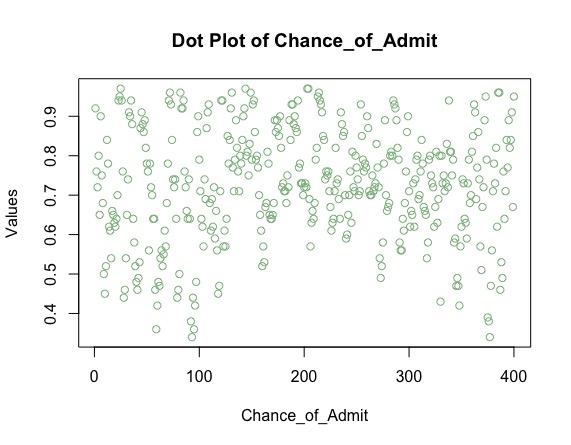


Figure (ia). The dot plot of Chance\_of\_Admit

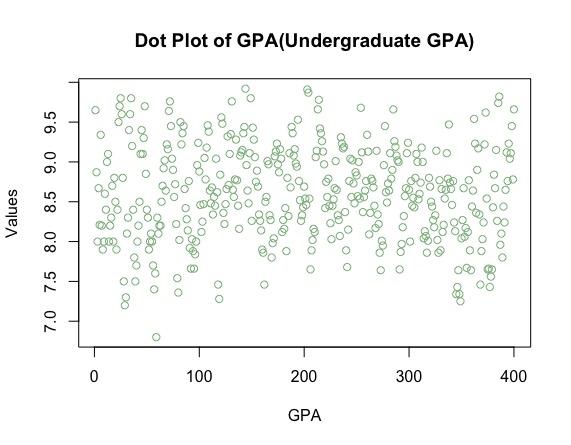


Figure (ib). The dot plot of GPA (Undergraduate GPA)

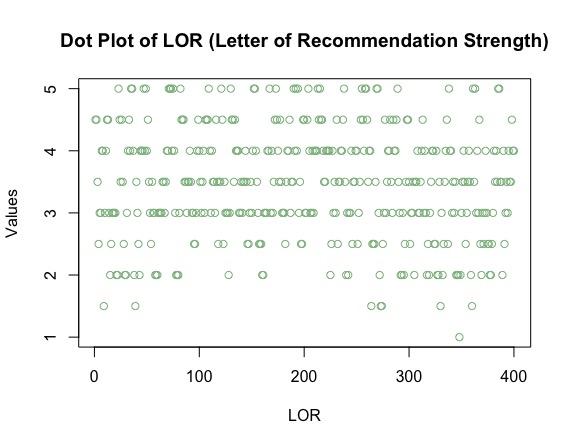


Figure (ic). The dot plot of LOR (Letter of Recommendation Strength)

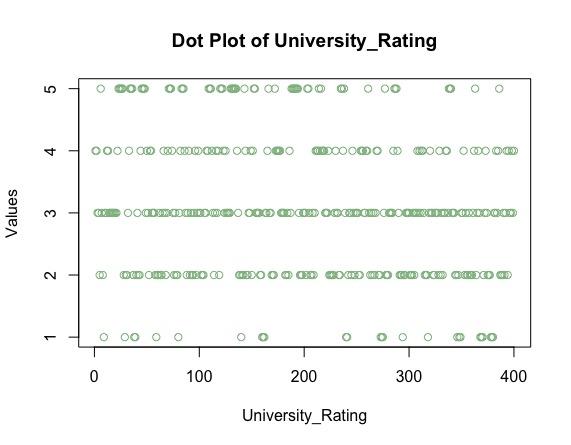


Figure (id). The dot plot of University\_Rating

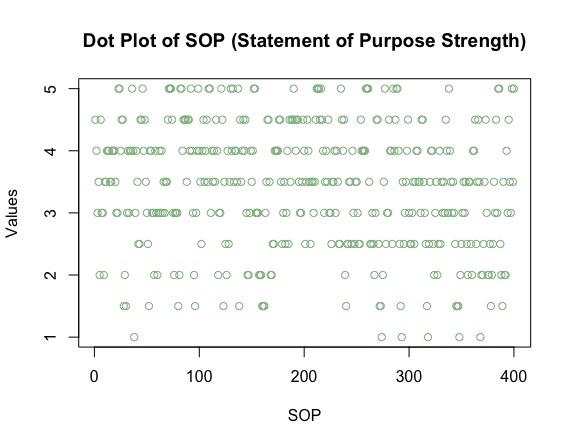


Figure (ie). The dot plot of SOP (Statement of Purpose Strength)

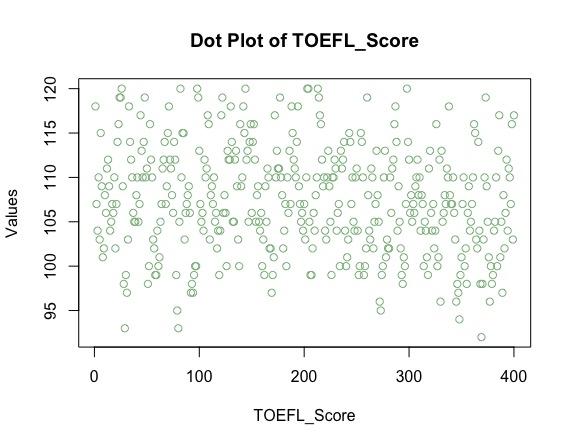


Figure (if). The dot plot of TOEFL\_Score

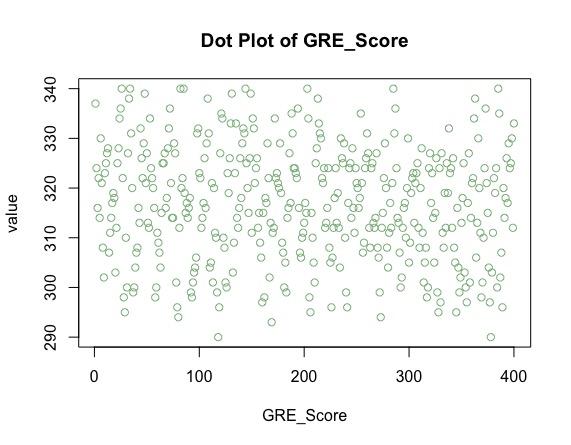
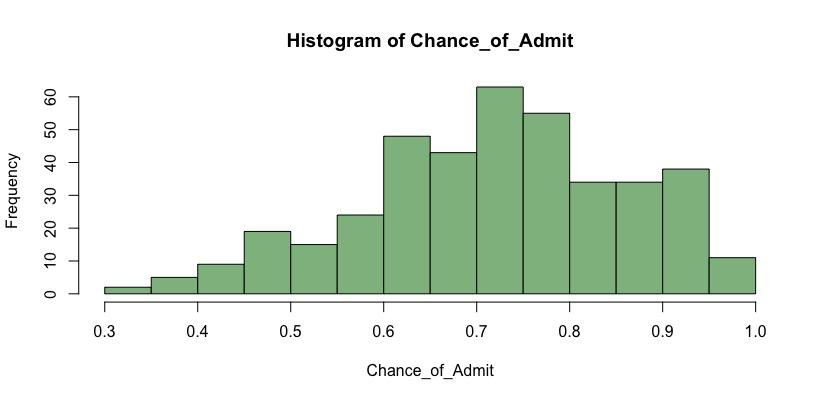


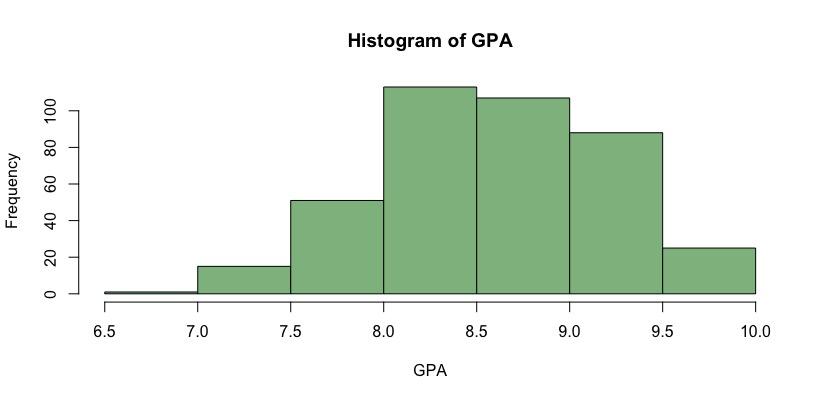
Figure (ig). The dot plot of GRE\_Score

**ii. Histogram**

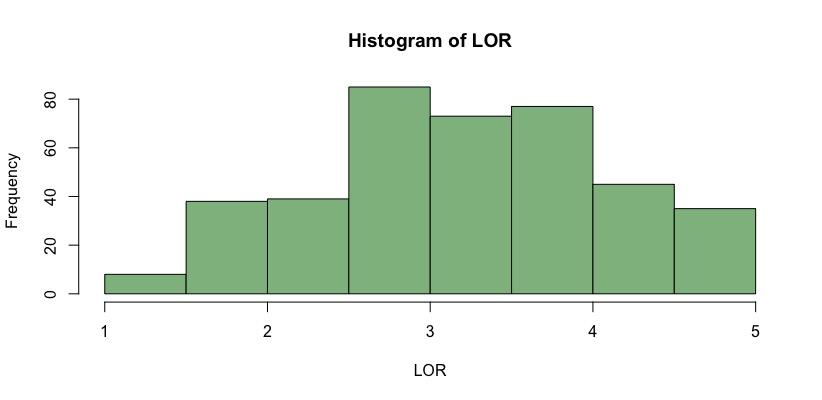
Since most of the histogram distribution are close to normal with a slight skewness to the left, transformation of variables is not necessary in this case.

****

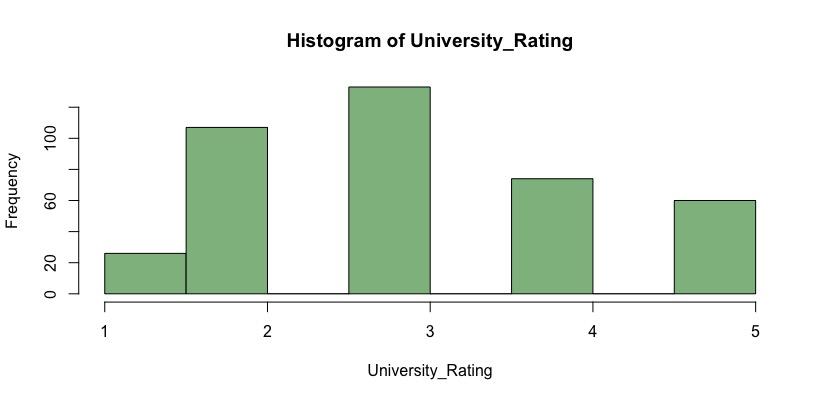
Figure(iva). Histogram for Chance\_of\_Admit



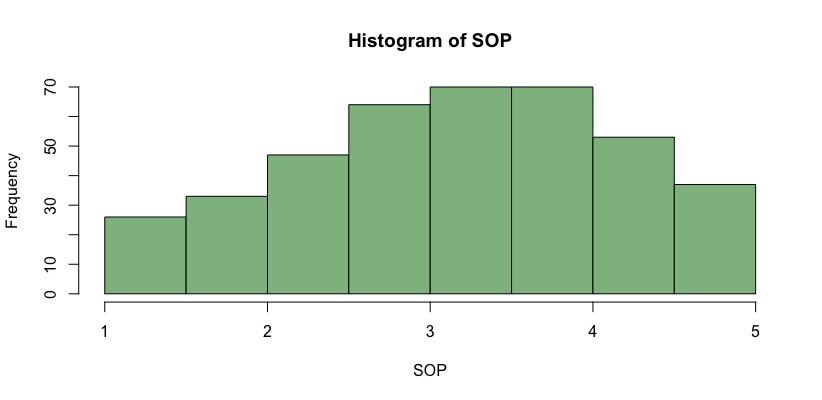
Figure(ivb). Histogram for GPA



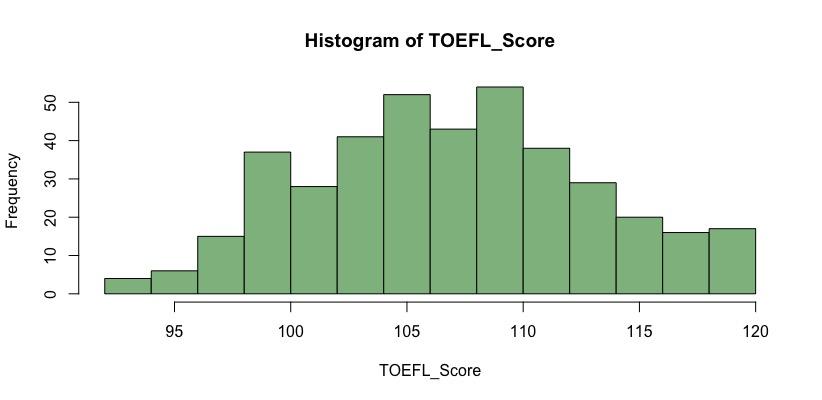
Figure(ivc). Histogram for LOR



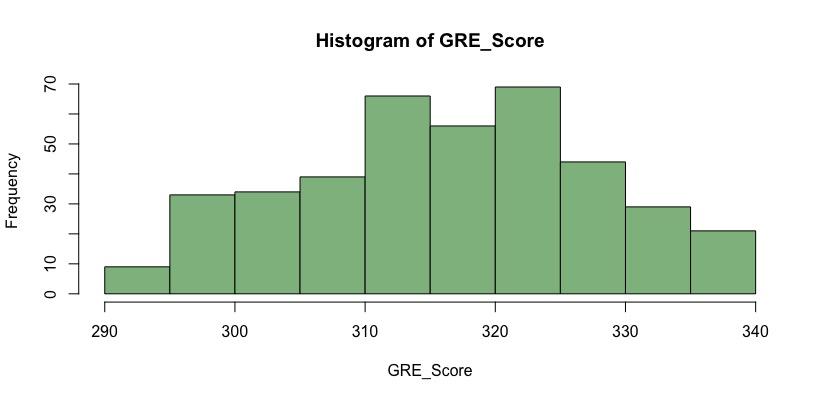
Figure(ivd). Histogram for University Rating



Figure(ive). Histogram for SOP



Figure(ivf). Histogram for Toefl Score



Figure(ivg). Histogram for GRE

**d. Use of transformed variables in bivariate and multivariate analysis**

Skewness is usually described as a measurement of a dataset’s symmetry or lack of symmetry. A perfect symmetrical dataset has a skewness of 0. Therefore, the normal distribution has a skewness of 0. A positive skewness indicates that the size of the right-handed tail is larger than the left-handed tail, meaning the distribution skewed to the right. On the contrary, if the size of the right-handed tail is smaller than the left-handed tail, the distribution will be skewed to the left with its skewness being negative.

To decide whether the we need to transform a variable based on its skewness, we have the following criteria:

· If the skewness is between -0.5 and 0.5, the variable is fairly symmetrical;

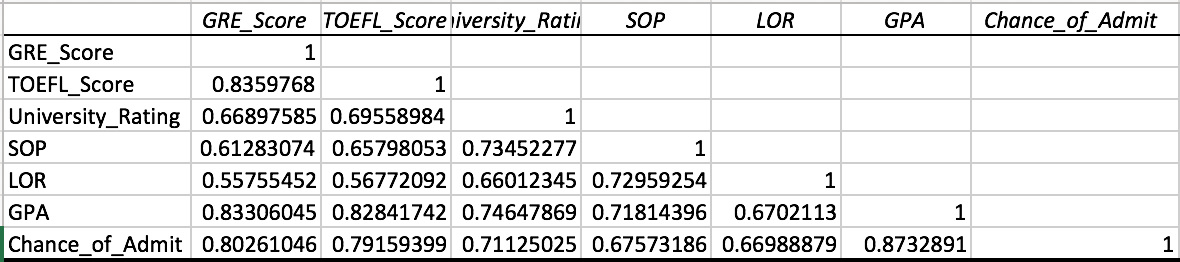
· If the skewness is between -1 and -0.5 or between 0.5 and 1, the variable is moderately skewed;

· If the skewness is less than -1 or greater than 1, the variable is highly skewed.

We scanned through the histograms of all six variables, including independent variables and dependent variables, and found out that all the skewness of all variables was within the range of -0.5 to 0.5, which meant that they were fairly symmetrical. Therefore, there is no need to introduce transformation of variables to this analysis.

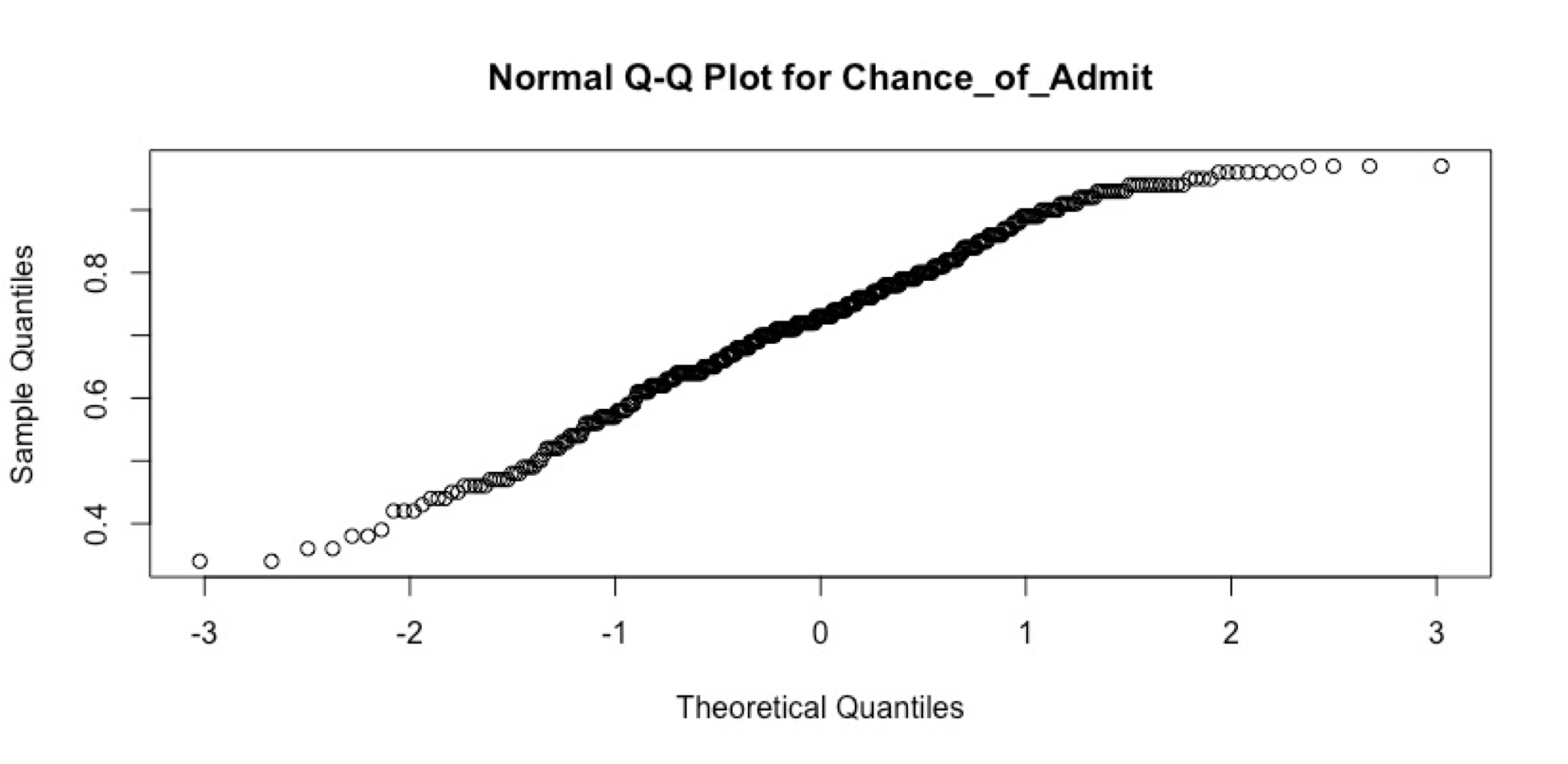
If we want to transform categorical variables to avoid redundant levels or to avoid rare levels, we can simple combine the different levels. If data has heavy skewness, common transformations of the data include square root, cube root, and log.

**f. Bivariate Testing**

****

Figure(f1). Correlation Matrix

**i. Residual Diagnostic Plot**

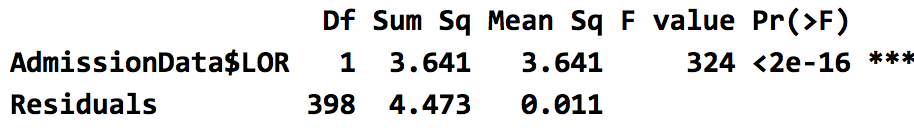


Figure(fi). Residual Diagnostic Plot

The Normal QQ plot helps us to assess whether the residuals are roughly normally distributed. If the residuals look far from normal we may be in trouble. In particular, if the residual tend to be larger in magnitude than what we would expect from the normal distribution, then our p-values and confidence intervals may be too optimistic. i.e., we may fail to adequately account for the full variability of the data. If the residuals here are a perfect match to the diagonal line, these residuals look to be normally distributed. (Chouldechova)

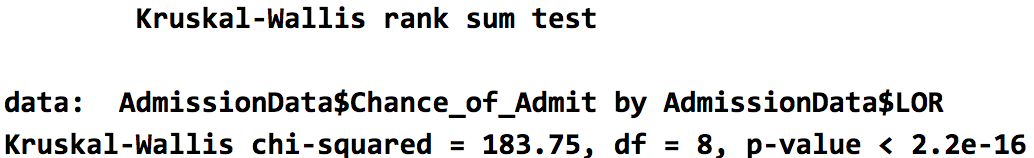
However, in our case, we saw a QQ plot where the residuals deviate from the diagonal line in both upper and lower tail, showing to be ‘lighter’ (have smaller values) than what we would expect under the standard modeling assumptions.

**ii. ANOVA /MANOVA – F-Distribution - Kruskall-Wallis Test**



Figure(fii1). One-way ANOVA and F-distribution

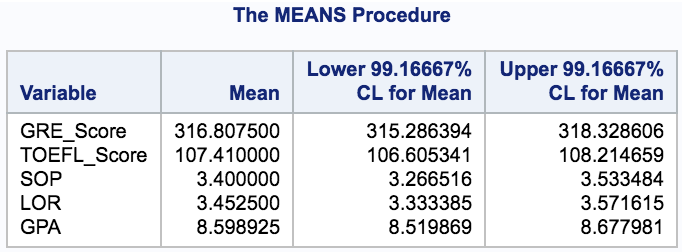
As the p-value Pr(>F) is less than the significance level 0.05, we could conclude that there are significant differences between groups of Letter of Recommendation Strength, which means that Letter of Recommendation Strength does affect Indian American students’ chance of admission by graduate programs.



Figure(fii2). Kruskal-wallis Test

The null hypothesis for Kruskal-wallis Test is that the Letter of Recommendation Strength are identical populations. To test the hypothesis, we apply the Kruskal-wallis Test to compare the independent Letter of Recommendation Strength data. The p-value turns out to be nearly zero. Hence, we reject the null hypothesis. At 0.05 significance level, we conclude that the Letter of Recommendation Strength from 1 to 5 are non-identical populations.

**iii. Simultaneous Confidence Intervals**

****

Figure(fiii). Simultaneous Confidence Intervals

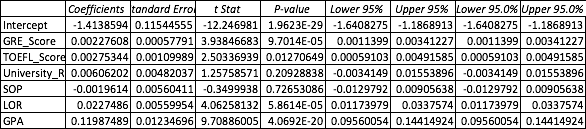
The percentage of times that a group of confidence intervals will all include the true population parameters or true differences between factor levels if the study were repeated multiple times. The simultaneous confidence level is based on both the individual confidence level and the number of confidence intervals. For a single comparison, the simultaneous confidence level is equal to the individual confidence level. However, each additional confidence interval causes the simultaneous confidence level to decrease in a cumulative way.(Looney et al., 1995)

If we want the overall coverage probability to be at most 95%, we can construct 6 univariate confidence intervals, each with significance level 0.05/6. For 95% of the random samples drawn from the population of Indian American students who are applying for graduate programs, the population means will be contained in the regions that are computed in this way.

**g. Multivariate Regression**

**a. Classical Linear Regression Analysis**

**i. Likelihood Ratio Tests for the Regression Parameters (Predictors)**

****

Figure(ai). Likelihood Ratio Tests

The likelihood ratio is a statistic that reject the null hypothesis if the value of this statistic is too small.

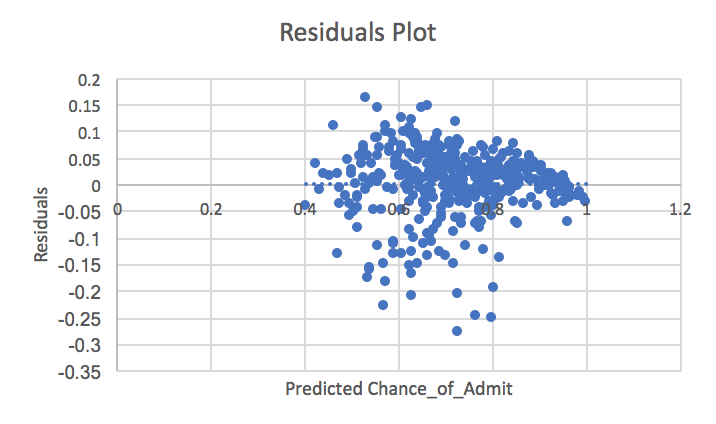
For testing H0: =0, for GRE\_Score, Toefl\_Score, LOR, and GPA, we find the p-values of t-stat are much smaller than 0.05. We reject these values if they are too large. Thus, we conclude that GRE\_Score, Toefl\_Score, LOR, and GPA are significant in the first order model under 0.05 significance level.

For testing H0: =0, for University\_Rating and SOP, we find the p-values of t-stat are much larger than 0.05. We reject these values if they are too large. Thus, we conclude that University\_Rating and SOP are not significant in the first order model under 0.05 significance level.

**ii. Report R square and significance as p value or confidence interval**

Examining the Excel result for this model, we find: r= 0.89, about 79.87% of the variability in the dependent variable scores around the mean of dependent variable is explained by multivariate regression between dependent variable and independent variables. The std. dev. of the dependent variable scores about a multivariate linear regression using independent variable is approximately 0.45 times the std. dev. of the dependent variable scores about the mean of the dependent variable. As independent variable increases, the estimated dependent variable increases in this fitted multivariate regression model.

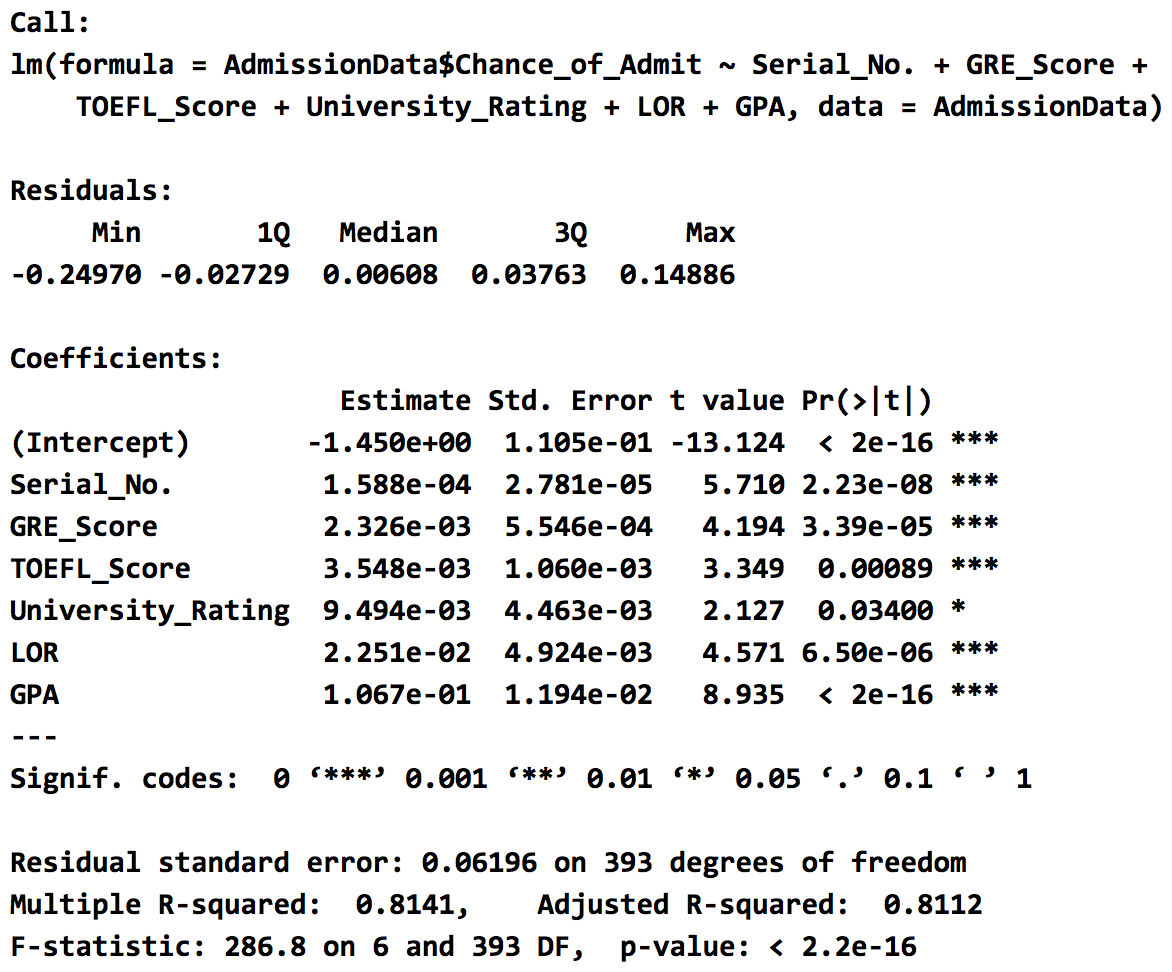
**iii. Model Checking and Fit - Residual Plots**

****

Figure(iii1). Residual plot for the multivariate model fitted

Examining the residual plot for the multivariate model (Figure(iii1)), we note: the mean of the residuals seems to be zero regardless of the value if the predicted. Since the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data.

**iv. Forward or Backward Stepwise Regression**



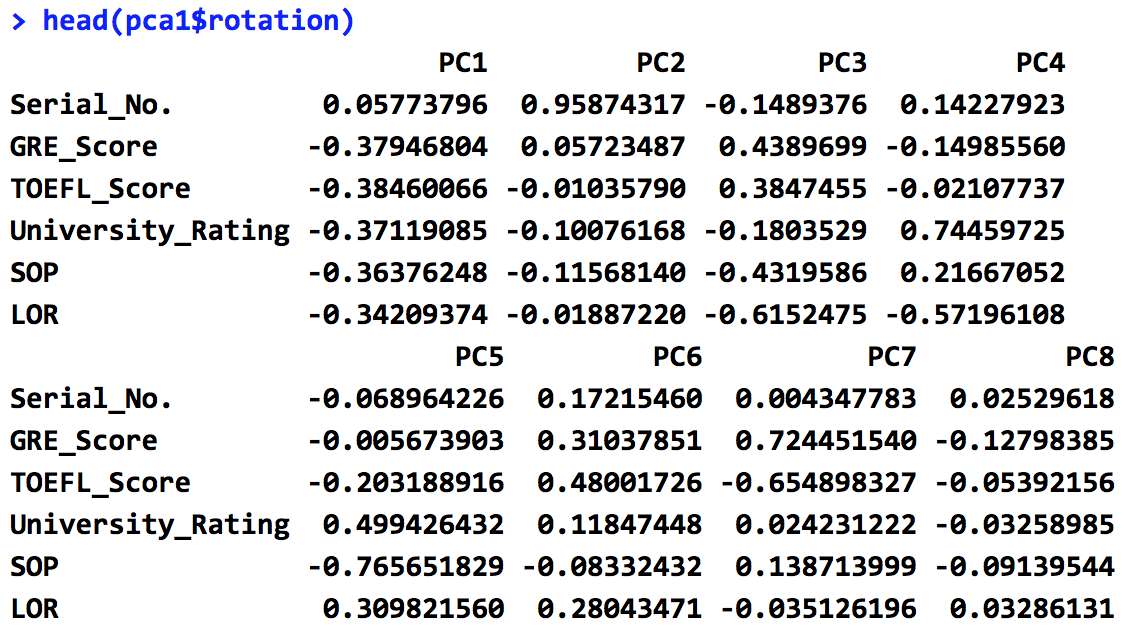
Figure(iv). Stepwise Regression

**h. Multivariate Methods**

**a. Classification Methods**

**i. Factor Analysis – Principal Components Analysis**

A typical PCA results consist of a set of eigenvalues, a table with the scores of Principal Components (PCs), and a table of loadings (or correlations between variables and PCs). The eigenvalues provide information of the variability in the data. The scores provide information about the structure of the observations. The loadings (or correlations) allow us to get a sense of the relationships between variables, as well as their associations with the extracted PCs.



Figure(hai1). table of loadings

Interpretation of the principal components is based on finding which variables are most strongly correlated with each component, i.e., which of these numbers are large in magnitude. Since which numbers we consider to be large or small is of course is a subjective decision, we determine at a correlation above 0.5 is deemed important. We will now interpret the principal component results with respect to the value that we have deemed significant.

The first, sixth, and eighth principal components are not strongly correlated with any the original variables.

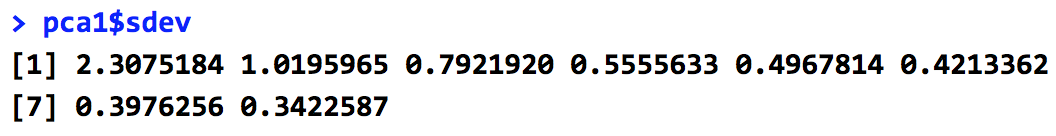
The second principal component is strongly correlated with one of the original variables. The second principal component increases with increasing Serial\_No. We could state that this principal component is primarily a measure of the Serial\_No.

The third principal component is strongly correlated with LOR. The third principal component increases with increasing LOR. We could state that this principal component is primarily a measure of the LOR.

The fourth principal component is strongly correlated with LOR and University\_Rating. The third principal component increases with increasing LOR and University\_Rating. This suggests that these two criteria vary together. If one increases, then the remaining ones tend to increase as well. This component can be viewed as a measure of the quality of LOR and University\_Rating.

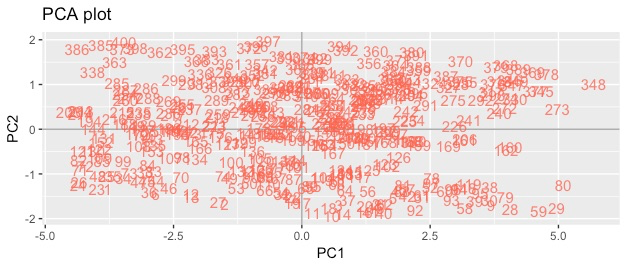
The fifth principal component is strongly correlated with SOP. The fifth principal component increases with increasing SOP. We could state that this principal component is primarily a measure of the SOP.

The seventh principal component is strongly correlated with GRE\_Score and Toefl\_Score. The seventh principal component increases with increasing GRE\_Score and Toefl\_Score. This suggests that these two criteria vary together. If one increases, then the remaining ones tend to increase as well. This component can be viewed as a measure of the quality of GRE\_Score and Toefl\_Score.



Figure(hai2). set of eigenvalues

Eigenvalues (also called characteristic values or latent roots) are the variances of the principal components. We can use the size of the eigenvalue to determine the number of principal components. Retain the principal components with the largest eigenvalues. For instance, we use only the principal components with eigenvalues that are greater than 1. Therefore, our minimum number of principal components is 2.

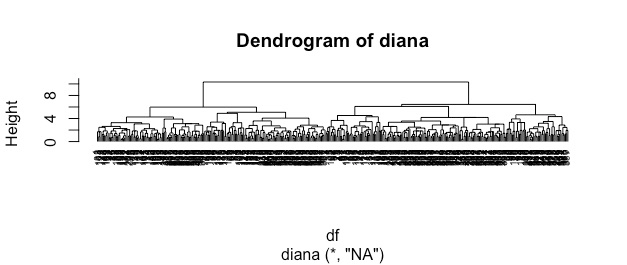


Figure(hai3). PCA Plot

Therefore, we select LOR, SOP, GRE\_Score, University\_Rating and Toefl\_Score. Creating new variables aims to establish the number of factors, that adequately captures the dimensions of the variables, required to explain relationship patterns among variables, the nature of the factors, the explanation effectiveness of the factors, and how much unique or random variance each observed variable includes.

**iii. Cluster – partition into segments (one method either: Hierarchical or k-means)**

The horizontal axis of the dendrogram represents the distance or dissimilarity between clusters. The vertical axis represents the objects and clusters. Our main interest is in similarity and clustering. Each joining (fusion) of two clusters is represented on the graph by the splitting of a horizontal line into two horizontal lines. The horizontal position of the split, shown by the short vertical bar, gives the distance (dissimilarity) between the two clusters. (“NCSS”, n.d.)

****

Figure(haiii). Hierarchical Cluster Analysis

**IV. RESULTS**

The results here suggest that the chance of being accepted into graduate programs has a strong positive correlation with an applicant’s GPA, GRE scores, TOEFL scores, Strength of recommendation letter, and the strength of statement of purpose, which means that the applicants with more satisfactory academic performance tend to have higher chance to be accepted into a graduate program than the applicants with less satisfactory academic performance. The results also suggest that acceptance into a graduate program is based on more than quantitative capabilities that are measure by GPA or TOEFL score, because strong recommendation letter and personal statement also lead to bigger chance of admission. After the GPA, official exam scores, and strength essays reaches a certain point, however, the chance of admission stopped increasing but rather stayed the same.

From the scatter plots of correlation between each aspect and the chance of admission, it is clear to see that when the scores or the strength of essays are low, the chances of acceptance are varied from almost 30 percent to 80 percent, which may mean that if an applicant’s one aspect is low, the admission committee would consider other aspects of the applicant and accept them. Contrarily, the applicants with higher exam scores and stronger essays had narrow variety on the chance of admission and they were concentrated in the percentage range of about 70 to 100 percent.

By doing multivariate regression on dependent and independent variables, we conclude a prediction model that is E (Chance of Admission) = -1.450 +0.0001588 \* GRE\_Score + 0.002326 \* TOEFL\_Score + 0.003548 \* University\_Rating - 0.009494\* SOP + 0.02251 \* LOR + 0.1067 \* GPA. R-square= 0.81. University\_Rating could be considered to be eliminated based on its large p-values at 0.01 significance level in order to a better fit. However, R-square will not be improved much by doing so.

Possible limitation of this study is that the data is only collected from the Indian-American community. Thus, the results of this study may not be the general or typical results. The results of studies done with the data from other communities with different cultures may have different outcomes than that of the Indian-American applicants.

**V. DISCUSSION**

**a. How bivariate and multivariate relationships extend the univariate analysis**

Univariate analysis is the simplest form of data analysis where the data being analyzed contains only one variable. Therefore, univariate analysis doesn’t reveal the relationship between variables. Bivariate analysis is used to find out if there is a relationship between two different variables. Something as simple as creating a scatter plot or correlation matrix could give you a picture of what the relationship between two variables. If the data seems to fit a line or curve, then there is a relationship or correlation between the two variables. Therefore, in this paper, bivariate analysis helps improve the variable selection process. Multivariate analysis is the analysis of three or more variables. Some of these methods include Canonical Correlation Analysis, Cluster Analysis, Factor Analysis, MANOVA, Multiple Regression Analysis, Partial Least Square Regression, and Principal Component Analysis. For this paper, multivariate analysis helps to improve the efficiency of model study by looking into the relationship between independent variables and reducing the collinearity between independent variables.

**b. Differences between bivariate and/or multivariate results**

SOP and University\_Rating are two independent variables that are not favored by the bivariate analysis by showing the tendency of insignificance. However, in multivariate analysis, this variability of significance on SOP and University\_Rating has been stabilized.

**c. Impact of the Study**

The results of this study provide an actual evidence of what factors there are that the admission committees actually consider during the admission process. Also, the results of this study is very helpful to the future applicants to graduate school because by looking at the results and comparing them with their own scores, they can know what range of scores and strength of their aspects like GPA, TOEFL scores, and Personal Statement need to be in more in detail, in order to be considered as possible candidates and even be accepted into graduate programs. The study of the relationship between academic performances and the chance of admission to graduate programs is also able to imply the preferences of admission committee when reviewing thousands of applications, therefore, offering future applicants a direction about which factors should they put most efforts on to improve. In addition, since this dataset is from the perspective of Indian students, the results of this study could be related and compared to the United States students, seeing which factors or criteria have been changed when it comes to international students. Therefore, this research could be used to improve the protection of equal opportunities of applicants among the world. The gap between groups could be addressed through two ways. Firstly, transformation of variables could be applied, like square root, to eliminate the gap between groups. Secondly, new ways of collecting data could be introduced to reduce the gaps between groups.

**VI. References**

Kuther, T. (2018). What Role Does Your GPA Really Play in Graduate School Admissions? ThoughtCo.

Retrieved from https://www.thoughtco.com/gpa-role-in-graduate-school-admissions-1685863. Accessed July 31, 2018.

Burmeister, J., McSpadden, E., Rakowski, J., Nalichowski, A., Yudelev, M., and Snyder, M. (2014). Correlation of Admissions Statistics to Graduate Student Success In Medical Physics

Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5711228/

Mullen A.L., Goyette KA, Soares JA. (2003) Who Goes to Graduate School? Social and Academic Correlates of Educational Continuation after College. *Sociology of Education*. 2003;76(2):143. doi:10.2307/3090274.

AOFoundation. (n.d.) https://www.aofoundation.org/Structure/research/clinical-research/case/doing/therapeutic-rct-prognostic-cohort/Pages/part-4.aspx. Accessed July 31, 2018.

Graduate School of Texas at Austin | Eligibility for Admission Consideration

<https://gradschool.utexas.edu/admissions/where-to-begin/eligibility-for-admission-consideration>

Accessed August 13, 2018.

Acharya M.S. Graduate Admissions | Kaggle. https://www.kaggle.com/mohansacharya/graduate-admissions/home. Published March 4, 2018. Accessed July 31, 2018.

Are the Skewness and Kurtosis Useful Statistics? | BPI Consulting. https://www.spcforexcel.com/knowledge/basic-statistics/are-skewness-and-kurtosis-useful-statistics#skewness. Accessed July 31, 2018.

Bagozzi RP. *Advanced Methods of Marketing Research*. Cambridge, MA: Blackwell Business; 1994.

Variance-Covariance Matrix: Stock Price Analysis in R ... http://www.bing.com/cr?IG=375A51A058A340D7A5BB6417FAB1FA08&CID=1566E21189D9678D205EEE2C882466F4&rd=1&h=IY-6UHY0EJRXBj6x3FNBCltXWHwRt0hP0tI7jiTy-Z0&v=1&r=http://www.michaeljgrogan.com/variance-covariance-matrix-calculation-r/&p=DevEx.LB.1,5540.1. Accessed July 30, 2018.

base. function | R Documentation. https://www.rdocumentation.org/packages/base/versions/3.5.1/topics/eigen. Accessed July 31, 2018.

Looney SW (1995). How to Use Tests for Univariate Normality to Assess Multivariate Normality. *The American Statistician*. 1995;49(1):64. doi:10.2307/2684816.

Rahman, K (2015). When to Use a Particular Statistical Test

Retrieved from https://www.researchgate.net/post/Whats\_the\_best\_way\_to\_analyze\_the\_relationship\_between\_continuous\_dependent\_variable\_and\_discrete\_polychotomous\_variables\_in\_spss

Chouldechova, (n.d.). Regression Diagnostic Plots

Retrieved from <http://www.contrib.andrew.cmu.edu/~achoulde/94842/homework/regression_diagnostics.html>

Accessed Date August 13, 2018

Minitab Express, (n.d.). Understanding Individual and Simultaneous Confidence Levels in Multiple Comparisons

Retrieved from https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/anova/supporting-topics/multiple-comparisons/understanding-individual-and-simultaneous-confidence-levels/

Accessed Date August 13, 2018

Sokolove, P. (n.d.). Writing Personal Statements for Graduate School

Retrieved from https://webcache.googleusercontent.com/search?q=cache:vfOWMwMp63wJ:https://www.training.nih.gov/\_assets/slides\_7-17-12+&cd=3&hl=en&ct=clnk&gl=us

Accessed August 13, 2018

Harvard, (n.d.) School of Public Health. Writing a Graduate School Application Essay

Retrieved from https://cdn1.sph.harvard.edu/wp-content/uploads/sites/36/2016/06/Writing-a-Graduate-School-Application-Essay-Guide\_Nov-20151.pdf

Accessed August 13, 2018

Cohen, K. (2013). The Truth About Applying to a U.S. University as an International Student

Retrieved from https://www.huffingtonpost.com/kat-cohen/the-truth-about-applying-\_b\_3654246.html

NCSS (n.d.) Hierarchical Clustering/Dendrograms

Retrieved from (<https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Hierarchical_Clustering-Dendrograms.pdf>)

U.S. Mission India, (2017). 12% Increase in Indians Studying in the US. in 2017

Retrieved from https://in.usembassy.gov/u-s-hosts-million-international-students-second-consecutive-year/