**ACDEMIC PERFORMANCES AND CHANCES OF GRADUATE ADMISSION**

Submitted to:

Lynn A. Agre, MPH, PhD

Statistics 467

July 30, 2018

**I. ABSTRACT**

Introduction: This study investigates the association between chances of admission into graduate programs and Indian students’ performances in colleges. Chances of admission is 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.

Method: The value of students’ performances in colleges is the defined by varies aspects, including GPA, GRE scores, TOEFL scores, and strength of statement of purpose and letter of recommendation.

Results: The bivariate relationship between chances of admission into graduate programs and Indian students’ performances in colleges are investigated with univariate analysis like histograms and normal probability plots, and bivariate analysis like variance-covariance matrix, scatter plots and correlation matrix.

Discussion:  While a typical feature of an Indian student in the lower chances of admission appears to have relatively lower GPA, GRE scores, TOEFL scores, and strength of statement of purpose and letter of recommendation. This research suggests that the GPA is an important factor that affects the value of the car.

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

**II. INTRODUCTION**

The primary purpose of this study was to compare the association between the chance of graduate school admissions of a college student and the academic performance of the college Indian students who have records of GRE scores, Toefl score, college rating, strength of their statement of purpose and letter of recommendation. 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, the chance of graduate school admissions of a college student, based on the preponderance of evidence.

To understand why GPA or grade point average is important in the perspective of admission committees, Tara Kuther [5] 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”, which could also imply your performance of graduate study. Prior study has identified a significant relationship between the chance of graduate school admissions of a college student and the performance of this student in the university subject tests. According to The Economist, “depending on the type of graduate program you’re pursuing, that GPA left over from your college days may play a large part in your chances of admission. In fact, many graduate schools have a GPA cutoff for applicants. Though these cutoffs vary across programs and fields, Master’s programs tend to have a lower cutoff than PhD programs. The most common cutoff for Master’s is 3.0” [3] .

And to be more specifically express the influence of GPA, the research done by Ann L. Mullen [4] has 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”. 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” [4] . Therefore, to illustrate the importance of GPA, “generally, most master’s programs require minimum of GPAs of 3.0 or 3.3, and most doctoral programs require minimum GPAs of 3.3 or 3.5” [5] .

What’s more, GPA not only implies a student’s academic performance in the college, but also reflects the “parents' education, revealing a link between parents' education and academic performance. We also found strong direct influences of college GPA and major on entry into graduate school. Thus, a final mechanism that helps to transmit social background effects is academic performance, which supports the earlier findings of Ethington and Smart (1986). We also note, however, that those who achieve at the undergraduate level are more likely to continue their education, regardless of their family background” [4].

**III. METHODS**

1. **Measures of Variables**

In order to help students shortlist universities with their profiles, this dataset [1] 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 Admit (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 Admit (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 skew to the left. (Figure(iva)).

**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. Theses 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 (Figure(ivb)).

**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. (Figure(ivc)).

**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. (Figure(ivd)).

**Serial No.:** a discrete independent variable. 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. (Figure(ive)).

**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. (Figure(ivf)).

**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. (Figure(ivg)).

**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. (Figure(ivh)).

1. **Variable Selection**

We selected variable GPA as a single characteristic to investigate its correlation with Chance of Admit. The hypothesis delineated for this given dataset would be performing t-test on the coefficient of GPA independent variable in the model, and see the corresponding p-value. If p-value <0.05, the coefficient of GPA independent variable would be insignificant.

Since both the distribution of GPA and Chance of Admit are fairly symmetric, we did not involve transformation of these two variables. And statistical univariate analysis is presented as following.

**IV. RESULTS**

1. **Summary Table of Variables**

|  |  |  |  |
| --- | --- | --- | --- |
| *Serial No.* |  | *GRE Score* |  |
| Mean | 200.5 | Mean | 316.8075 |
| Standard Error | 5.780715065 | Standard Error | 0.573682306 |
| Median | 200.5 | Median | 317 |
| Mode | #N/A | Mode | 324 |
| Standard Deviation | 115.6143013 | Standard Deviation | 11.47364611 |
| Sample Variance | 13366.66667 | Sample Variance | 131.6445551 |
| Kurtosis | -1.2 | Kurtosis | -0.700489457 |
| Skewness | 2.2372E-17 | Skewness | -0.062893459 |
| Range | 399 | Range | 50 |
| Minimum | 1 | Minimum | 290 |
| Maximum | 400 | Maximum | 340 |
| Sum | 80200 | Sum | 126723 |
| Count | 400 | Count | 400 |
| Confidence Level(95.0%) | 11.36446554 | Confidence Level(95.0%) | 1.127817704 |

|  |  |  |  |
| --- | --- | --- | --- |
| *TOEFL Score* |  | *University Rating* |  |
| Mean | 107.41 | Mean | 3.0875 |
| Standard Error | 0.303475689 | Standard Error | 0.057186406 |
| Median | 107 | Median | 3 |
| Mode | 110 | Mode | 3 |
| Standard Deviation | 6.069513777 | Standard Deviation | 1.14372813 |
| Sample Variance | 36.83899749 | Sample Variance | 1.308114035 |
| Kurtosis | -0.578778411 | Kurtosis | -0.796286965 |
| Skewness | 0.057215914 | Skewness | 0.171260277 |
| Range | 28 | Range | 4 |
| Minimum | 92 | Minimum | 1 |
| Maximum | 120 | Maximum | 5 |
| Sum | 42964 | Sum | 1235 |
| Count | 400 | Count | 400 |
| Confidence Level(95.0%) | 0.59661114 | Confidence Level(95.0%) | 0.112424317 |

|  |  |  |  |
| --- | --- | --- | --- |
| *SOP (Statement of Purpose Strength)* |  | *LOR (Letter of Recommendation Strength)* |  |
| Mean | 3.4 | Mean | 3.4525 |
| Standard Error | 0.050343432 | Standard Error | 0.044923877 |
| Median | 3.5 | Median | 3.5 |
| Mode | 4 | Mode | 3 |
| Standard Deviation | 1.006868641 | Standard Deviation | 0.898477548 |
| Sample Variance | 1.013784461 | Sample Variance | 0.807261905 |
| Kurtosis | -0.675610343 | Kurtosis | -0.662484117 |
| Skewness | -0.275761168 | Skewness | -0.106991479 |
| Range | 4 | Range | 4 |
| Minimum | 1 | Minimum | 1 |
| Maximum | 5 | Maximum | 5 |
| Sum | 1360 | Sum | 1381 |
| Count | 400 | Count | 400 |
| Confidence Level(95.0%) | 0.098971527 | Confidence Level(95.0%) | 0.088317077 |

|  |  |  |  |
| --- | --- | --- | --- |
| *GPA(Undergraduate GPA)* |  | *Chance of Admit* |  |
| Mean | 8.598925 | Mean | 0.72435 |
| Standard Error | 0.029815855 | Standard Error | 0.007130467 |
| Median | 8.61 | Median | 0.73 |
| Mode | 8 | Mode | 0.64 |
| Standard Deviation | 0.596317096 | Standard Deviation | 0.14260933 |
| Sample Variance | 0.35559408 | Sample Variance | 0.020337421 |
| Kurtosis | -0.458475626 | Kurtosis | -0.389125918 |
| Skewness | -0.065990544 | Skewness | -0.3534481 |
| Range | 3.12 | Range | 0.63 |
| Minimum | 6.8 | Minimum | 0.34 |
| Maximum | 9.92 | Maximum | 0.97 |
| Sum | 3439.57 | Sum | 289.74 |
| Count | 400 | Count | 400 |
| Confidence Level(95.0%) | 0.058615803 | Confidence Level(95.0%) | 0.014017979 |

1. **Skewness**

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 [7]:

* 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.

W scanned through the histograms of all eight 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 univariate analysis.

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

**i. Dot Plot**

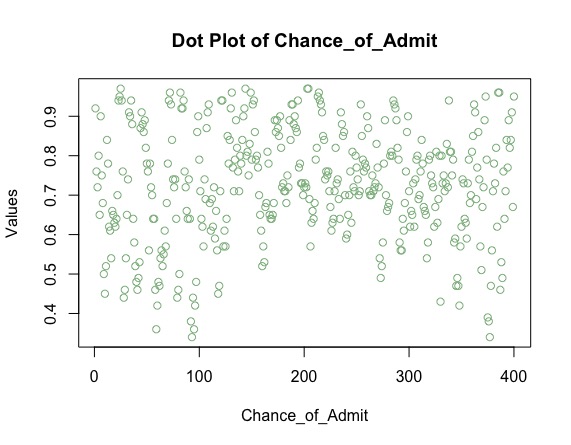


Figure (ia). The dot plot of Chance\_of\_Admit

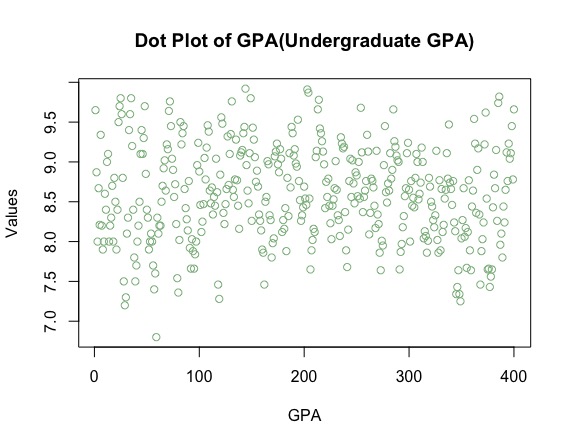


Figure (ib). The dot plot of GPA (Undergraduate GPA)

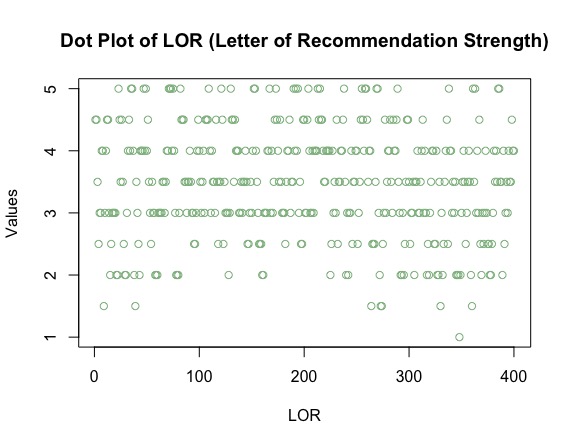


Figure (ic). The dot plot of LOR (Letter of Recommendation Strength)

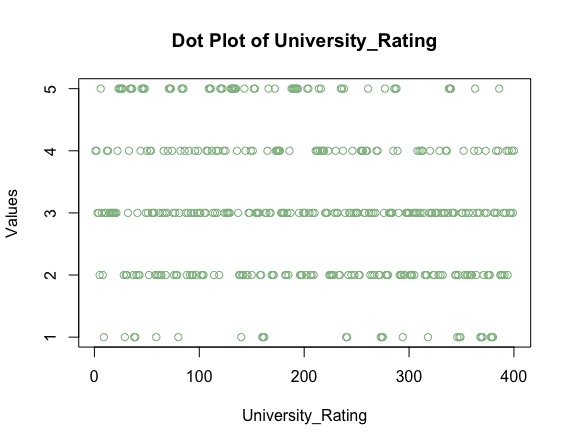


Figure (id). The dot plot of University\_Rating



Figure (ie). The dot plot of Serial No.

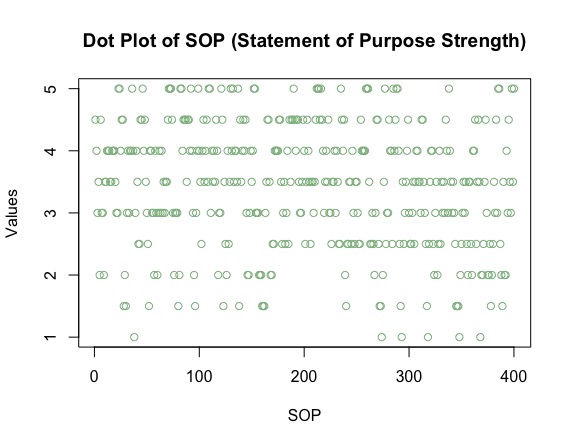


Figure (if). The dot plot of SOP (Statement of Purpose Strength)

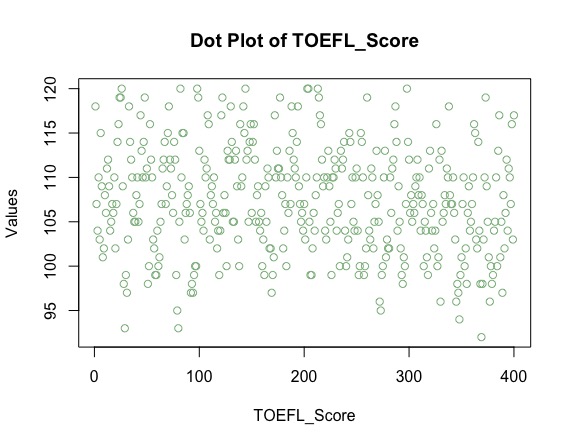


Figure (ig). The dot plot of TOEFL\_Score

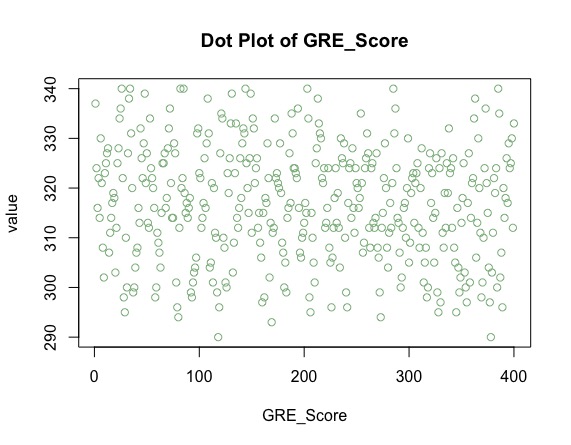


Figure (ih). The dot plot of GRE\_Score

**ii. Box Plot**

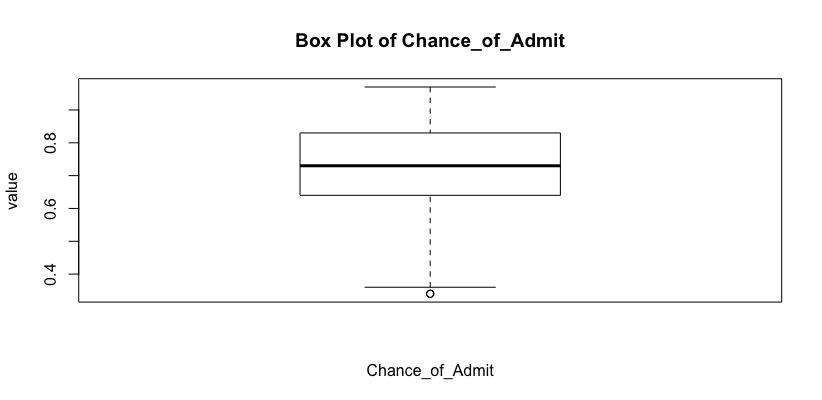
****

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

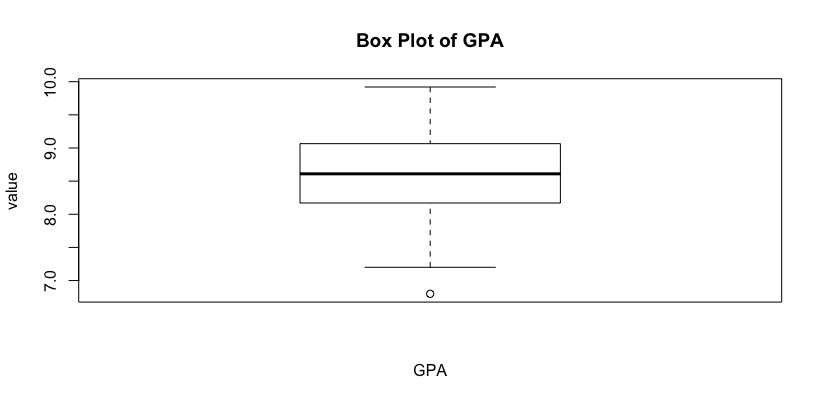


Figure (iib). The dot plot of GPA

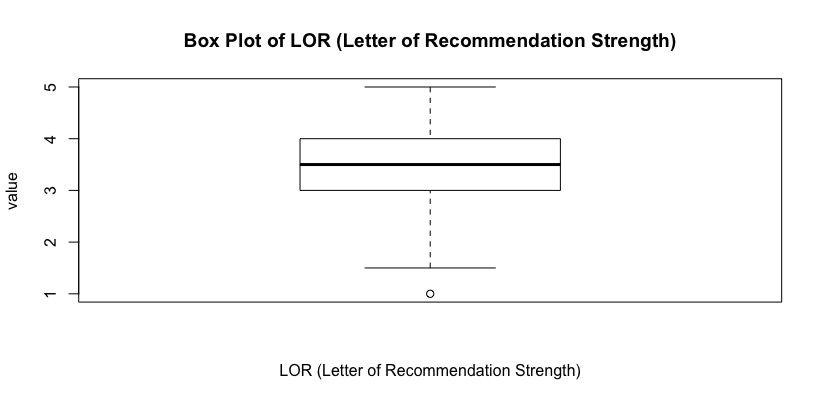
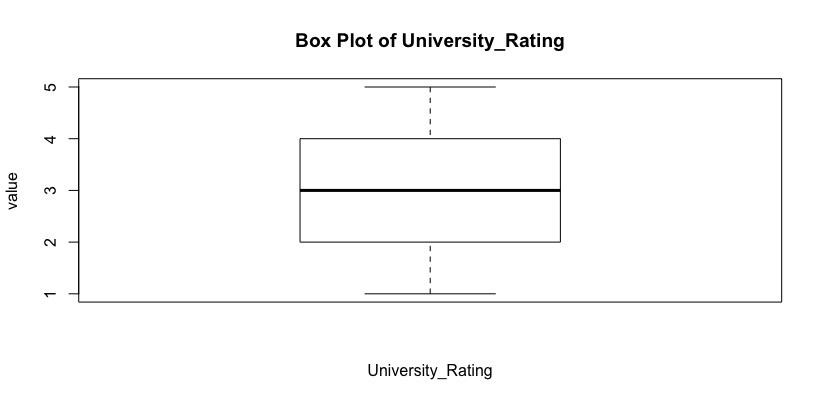


Figure (iic). The dot plot of LOR (Letter of Recommendation Strength)

 Figure (iid). The dot plot of University Rating

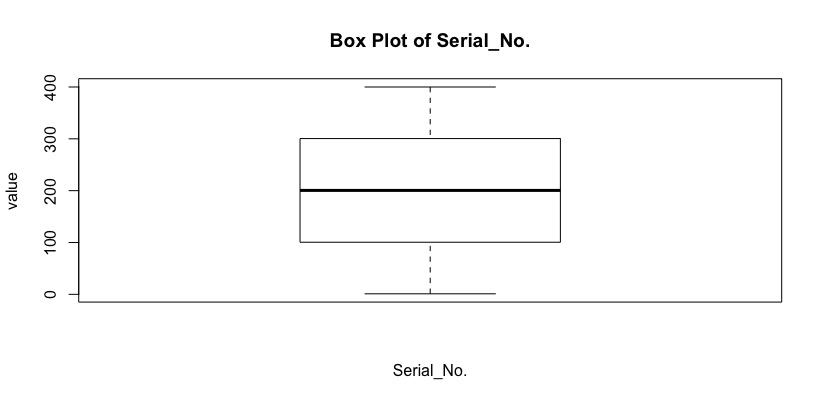


Figure (iie). The dot plot of Serial\_No.

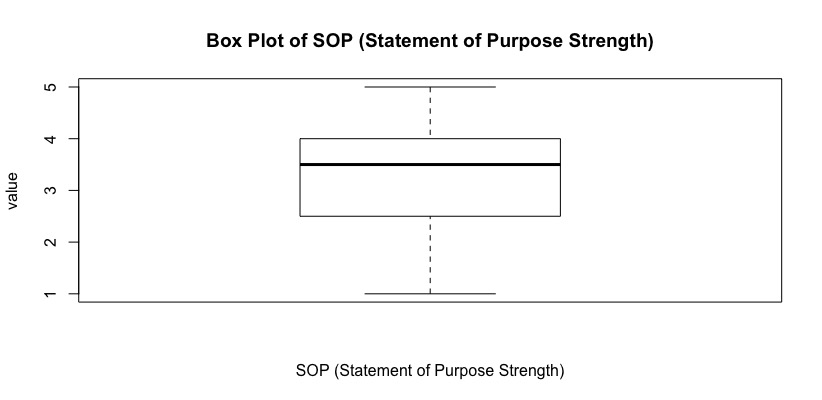
****

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

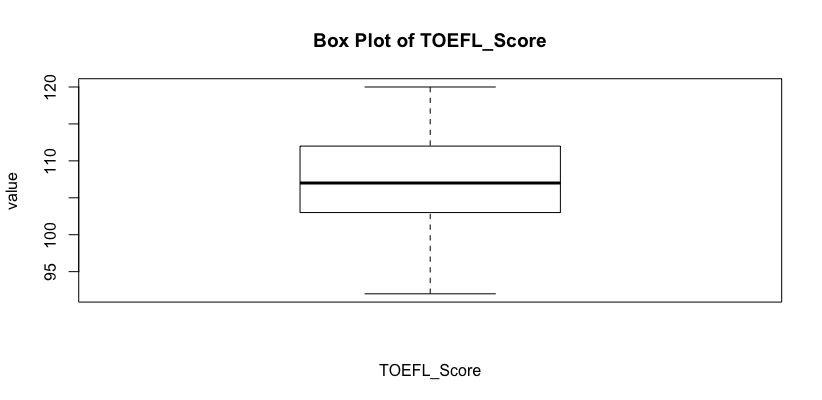
****

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

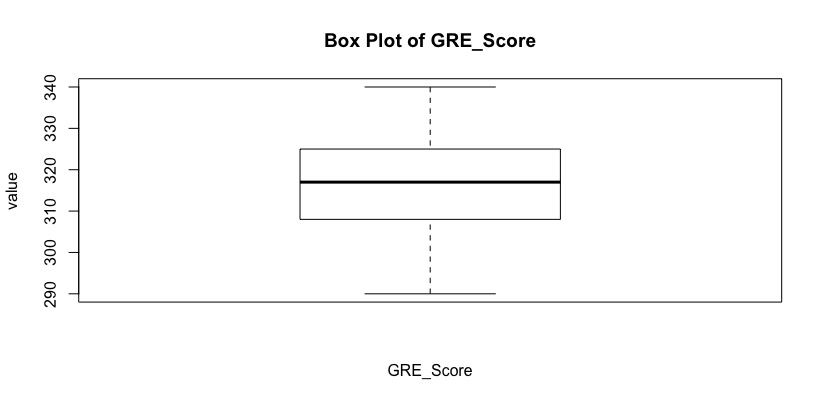
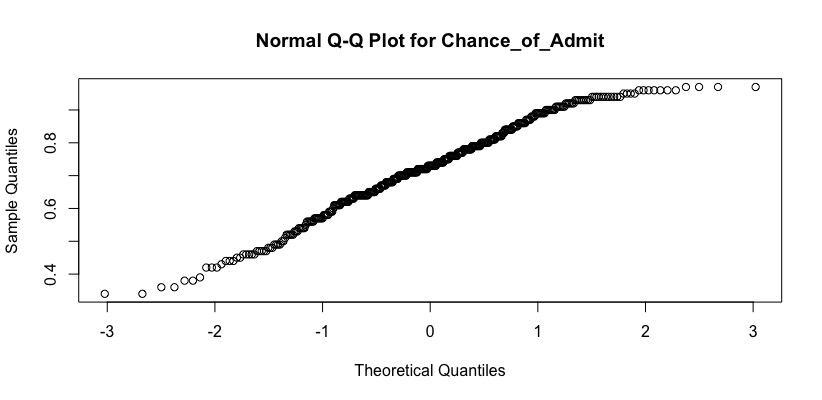
****

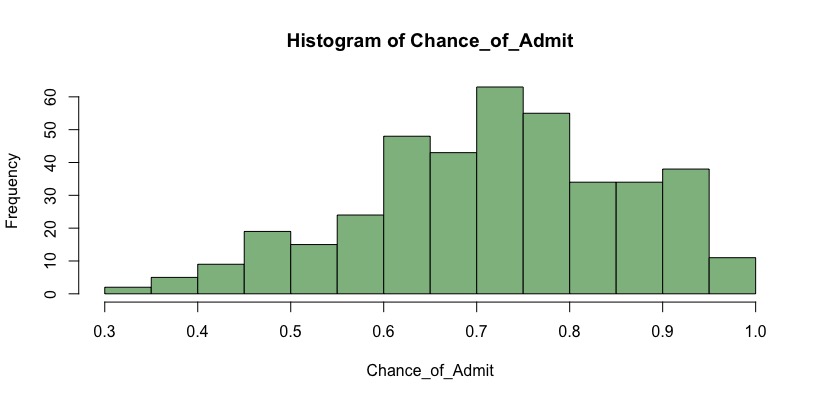
Figure (iih). The dot plot of GRE\_Score

**iii. Normal Probability Plot**

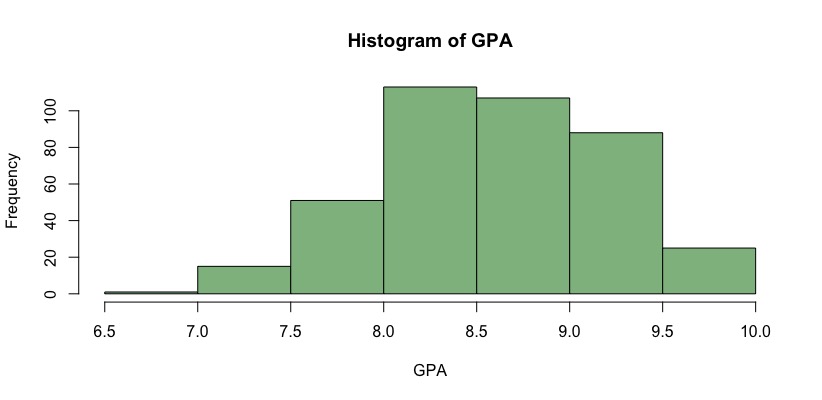


Figure(iiia). Normal Q-Q Plot for Chance\_of\_Admit

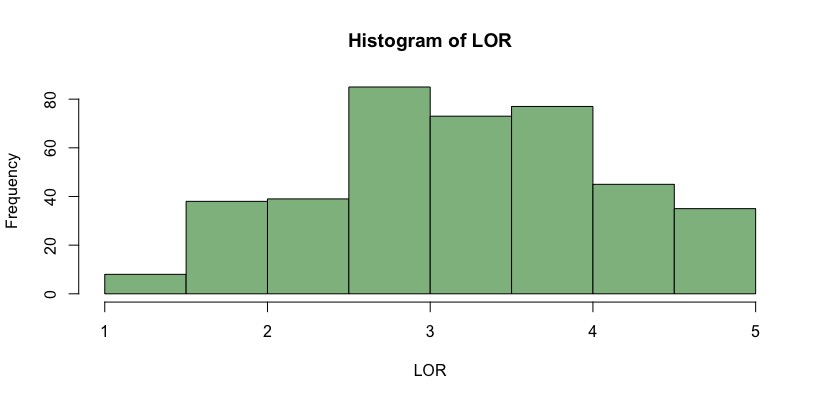
**iv. Histogram**

****

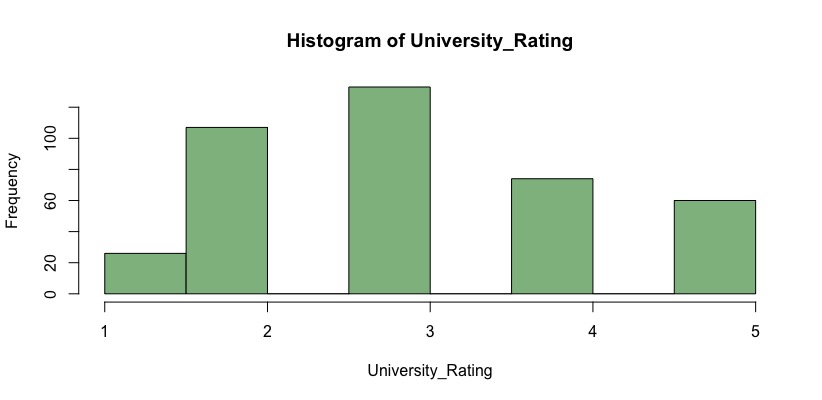
Figure(iva). Histogram for Chance\_of\_Admit



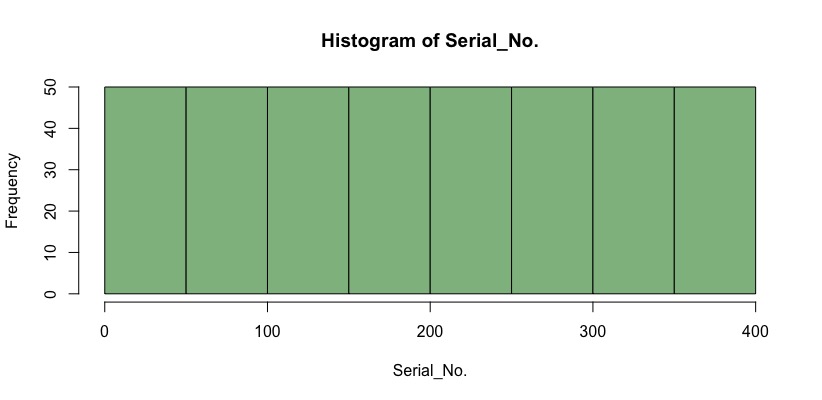
Figure(ivb). Histogram for GPA



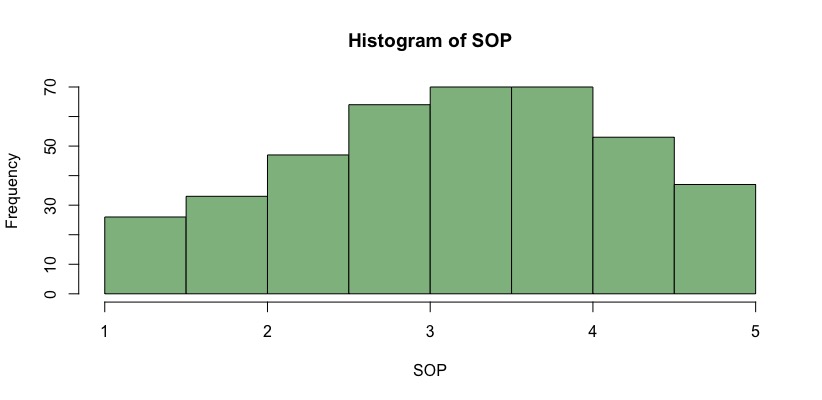
Figure(ivc). Histogram for LOR



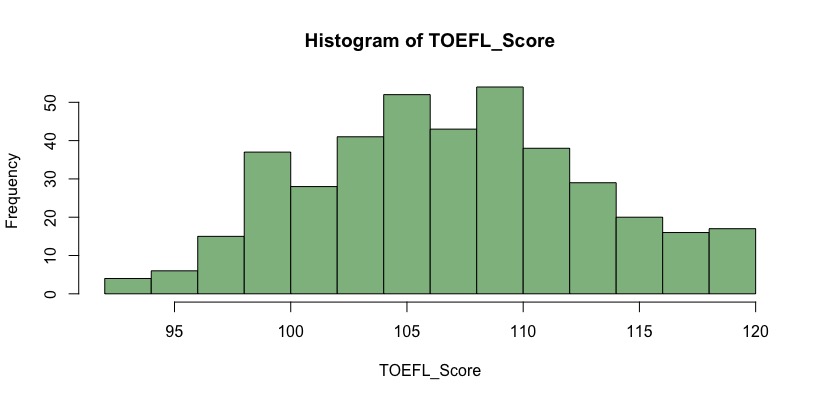
Figure(ivd). Histogram for University Rating



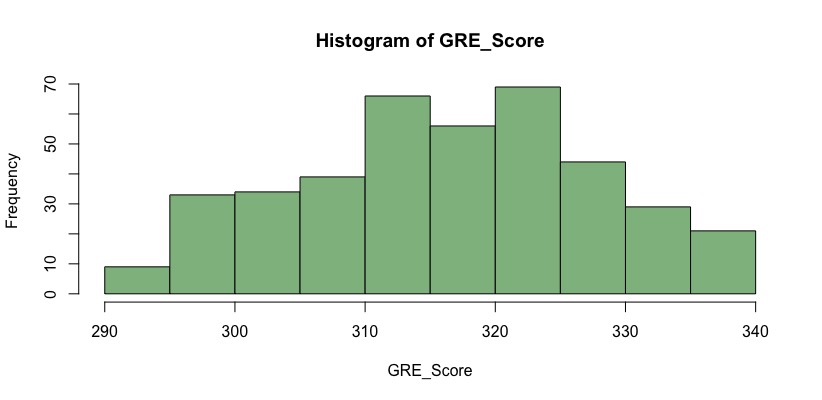
Figure(ive). Histogram for Serial No.



Figure(ivf). Histogram for SOP

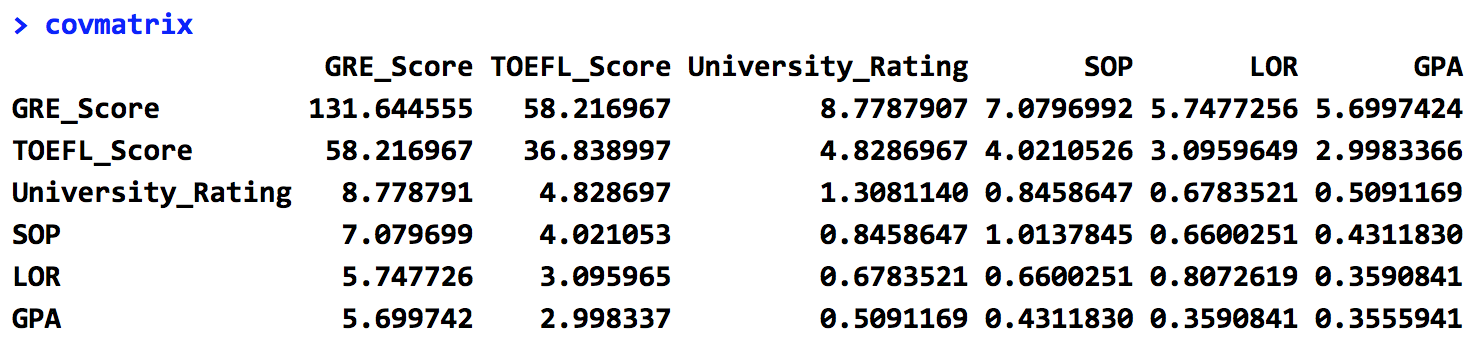


Figure(ivg). Histogram for Toefl Score



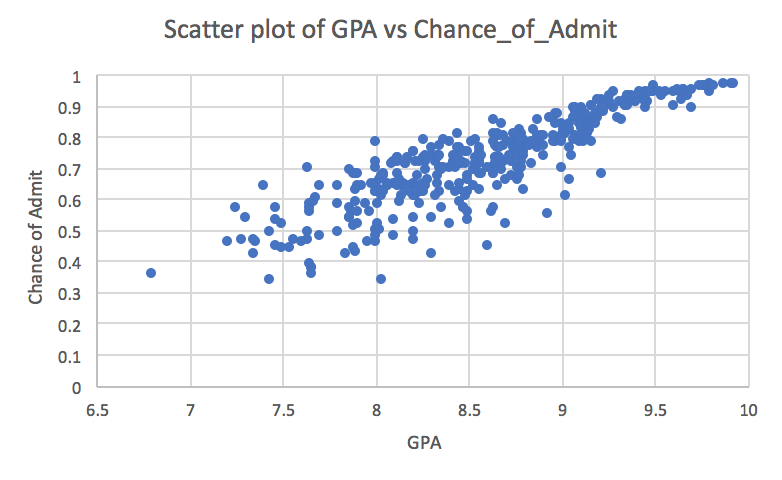
Figure(ivh). Histogram for GRE

1. **Bivariate Analysis/Graphs**
2. **Variance-Covariance Matrix**

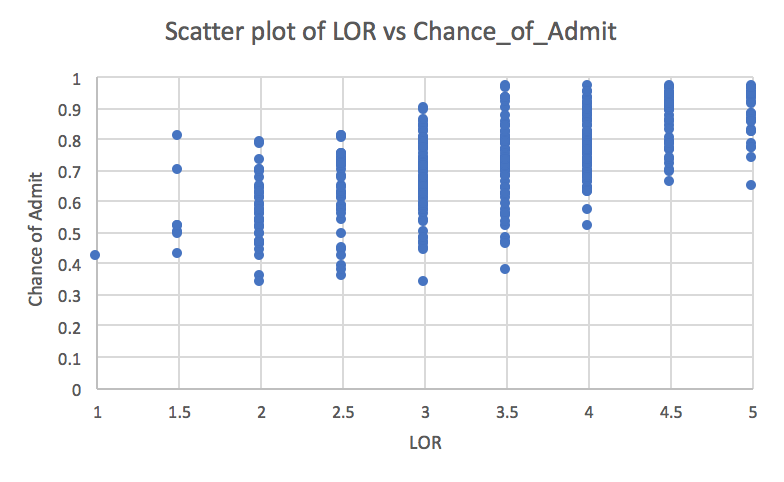
****

Figure(di). The variance-covariance matrix table [9]

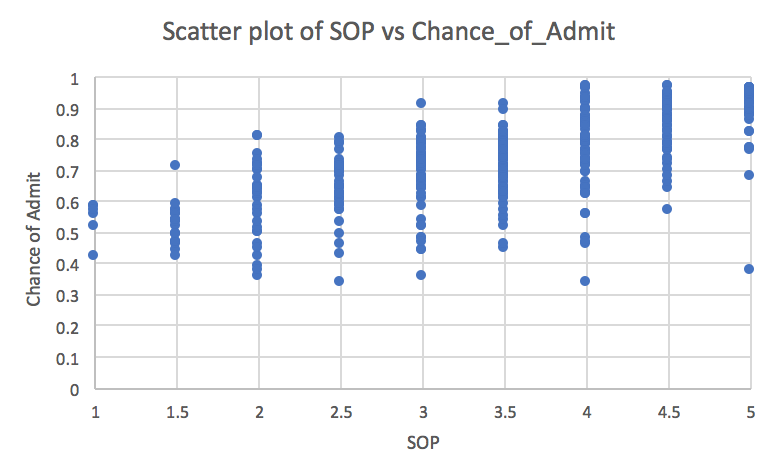
1. **Scatter Plot**

****

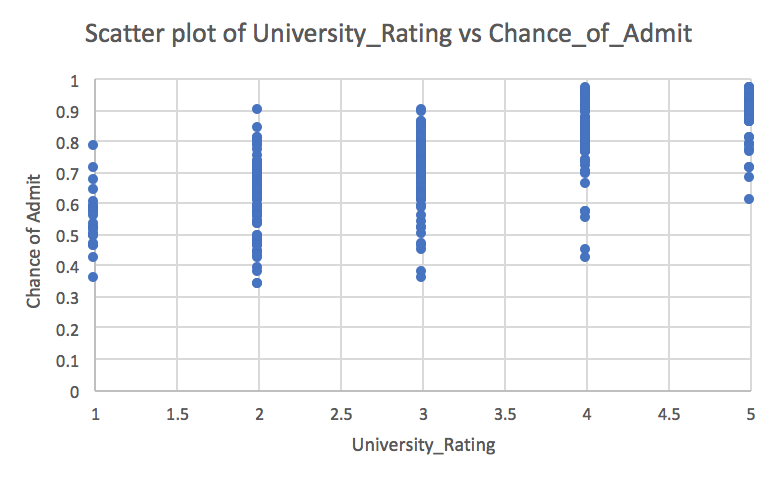
Figure(diia). The scatter plot of GPA vs Chance\_of\_Admit



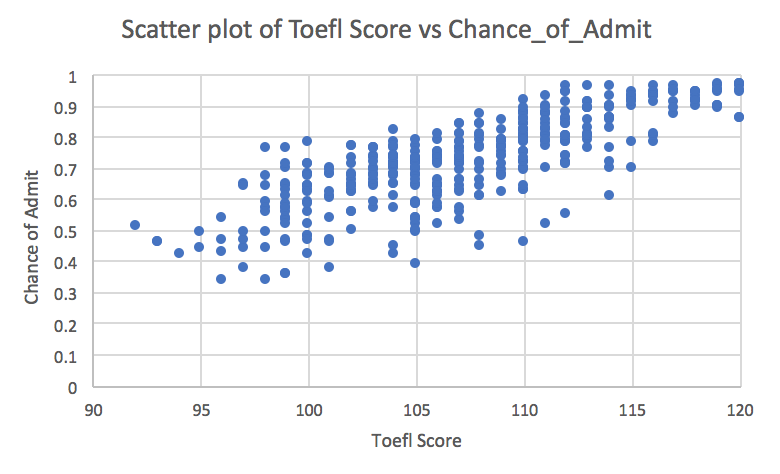
Figure(diib). The scatter plot of LOR vs Chance\_of\_Admit



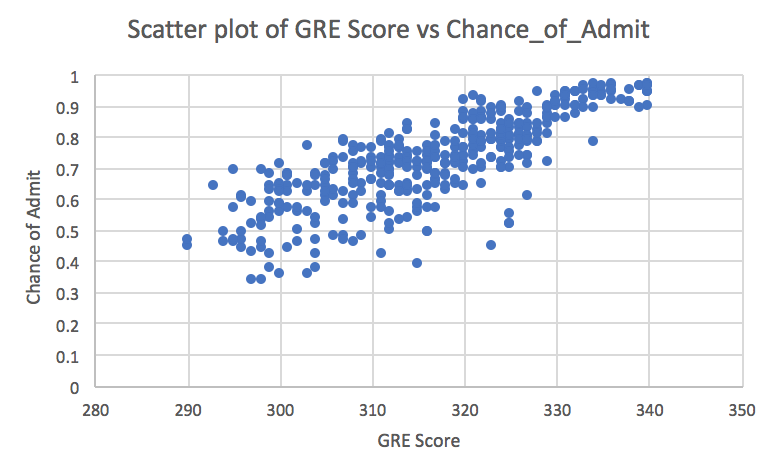
Figure(diic). The scatter plot of SOP vs Chance\_of\_Admit



Figure(diid). The scatter plot of University\_Rating vs Chance\_of\_Admit



Figure(diie). The scatter plot of Toefl Scores vs Chance\_of\_Admit

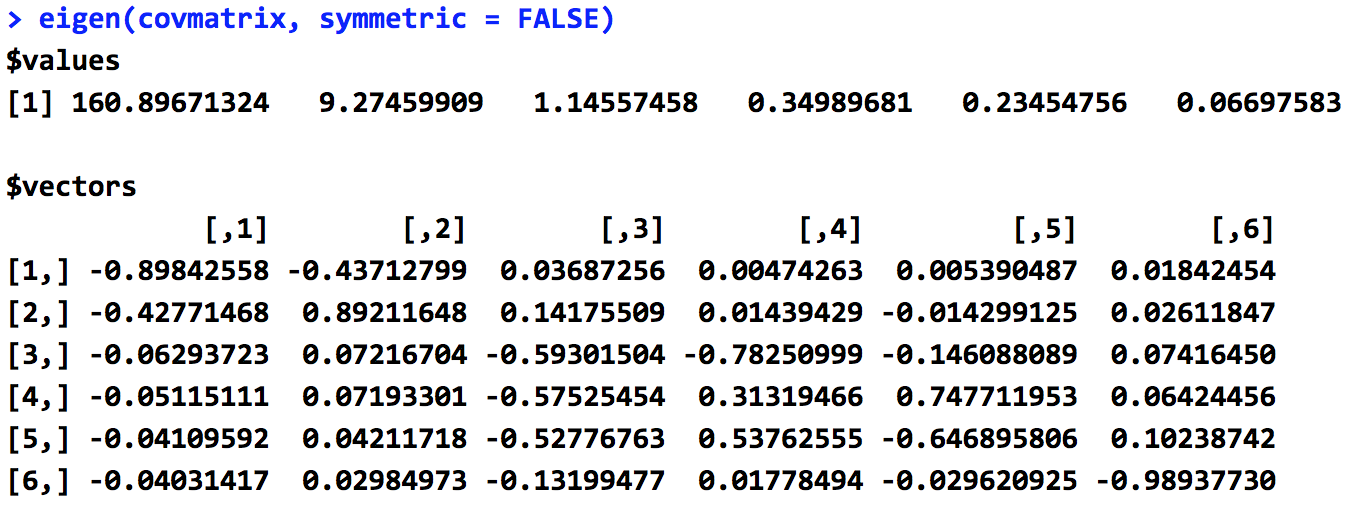


Figure(diif). The scatter plot of GRE Scores vs Chance\_of\_Admit

1. **Correlation Matrix**

|  |  |  |
| --- | --- | --- |
|  | *GPA* | *Chance\_of\_Admit* |
| GPA | 1 |  |
| Chance\_of\_Admit | 0.873289099 | 1 |

1. **Eigenvalues and associated eigenvectors**



Figure(div). The eigenvalues and the associated eigenvectors [10]

**IV. DISCUSSION**

From the research of Stephen W. Looney, he concludes that even though the power of these univariate/bivariate analysis has not been fully concluded, his experience “has been that the results of the tests, when taken as a whole, can lead to a reason- ably well-informed decision about the MVN (Multivariate analysis) assumption” [11]. First of all, the univariate analysis, like dot plots, box plots, and histograms, could improve the researchers’ understanding on the distributions of variables to decide whether variable transformation should be introduced to achieve a more symmetric distribution. What’s more, bivariate analysis, like scatter plots between all independent variables and the dependent variables and the correlation table between all pairs of the variables, could give researchers’ an idea of correlation and between all pairs of variables, which could help decide which variables are better to be selected into the models, and which should be eliminated from the model. In addition, bivariate analysis, like variance-covariance matrix, eigenvalues and eigenvectors, and maximum likelihood estimates, would reflect the level of independency of the all variables, reducing the potential errors in the further stages of researches.

Based on the type and distribution of the variables, variables, like GRE\_Score, TOEFL\_Score, SOP, LOR, GPA, tend to be included into the multivariate tests in order to predict Chance\_of\_Admit.

In order to illustrate the independency and dependency of variables, the determinant of variance-covariance matrix needs to be calculated. If the determinant of variance-covariance matrix equals to 0, then “either some of the values of the variable or the entire variable needs to be removed from the analysis. The corresponding reduced data matrix will then lead to a covariance matrix of full rank and nonzero generalized variance” [8]. Since the determinant of the previously calculated variance-covariance matrix does not equal to 0, the variables are relatively independent.

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. In addition, since this dataset is from the perspective of Indian students, the results of this study could be related and compare 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.

**V. REFERENCE**

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

[2] AOFoundation. https://www.aofoundation.org/Structure/research/clinical-research/case/doing/therapeutic-rct-prognostic-cohort/Pages/part-4.aspx. Accessed July 31, 2018.

[3] How Much Does GPA Matter in Graduate School Admissions? The Economist. https://gre.economist.com/gre-advice/graduate-school-admissions/application-advice/how-much-does-gpa-matter-graduate-school-admissions. Accessed July 31, 2018.

[4] Mullen A.L., Goyette KA, Soares JA. 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.

[5] Kuther T. What Role Does Your GPA Really Play in Graduate School Admissions? ThoughtCo. https://www.thoughtco.com/gpa-role-in-graduate-school-admissions-1685863. Accessed July 31, 2018.

[6] Ethington CA, Smart JC. Persistence to graduate education. *Research in Higher Education*. 1986;24(3):287-303. doi:10.1007/bf00992076.

[7] 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.

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

[9] 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.

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

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