# Predictive Analytics

**PostGraduation Admission Analysis for International students**

#### **Part One—Identify a Focal Point and a Dependent Variable**

**Target of analysis:** This analysis will predict the chances of admission and shortlisting type of universities that match students' profiles.

**Description of the context:** After a career break last year, I planned to pursue my higher studies(Masters) in the USA**.** Since I was an international student, I had to study and analyze the admission process for more than a year. I prefer to select post-graduation analysis as my course work as I had spent most of my time researching and understanding this field.

Graduate programs require applicants to submit various information to provide insight into their academic preparedness for graduate study interest in the program they are applying to. There are several admission requirements for the International Graduate Program in Universities. This analysis will help students in shortlisting universities with their profiles. The predicted output gives them a fair idea about their chances for a particular university.

This dataset contains information or Criteria of Post Graduate Admissions in the USA for international students. The dataset includes several parameters considered necessary during the application for Masters' Programs.

The UCLA Graduate Dataset inspires this dataset. The test scores and GPA are in the older format. The dataset is obtained by “**https://www.kaggle.com/mohansacharya/graduate-admissions**”

**Understanding Data:**

* There are 400 Observations of 9 Variables. The attributes included are

1. Serial No.
2. GRE Scores ( out of 340 )
3. TOEFL Scores ( out of 120 )
4. University Rating ( out of 5 )
5. Statement of Purpose ( out of 5 )
6. Letter of Recommendation Strength ( out of 5 )
7. Undergraduate GPA ( out of 10 )
8. Research Experience ( either 0 or 1 )
9. Chance of Admit ( 0.3-0.9 )

* The minimum Chance of Admission is 34%, and the maximum is 97%. The average Chance of Admission is 72%.
* The maximum GRE score is 340, and the Minimum GRE Score is 290. The average GRE score is 316.80, with 11.47 from the mean.

##### **Identify the dependent variable that will guide your prediction or decision**.

**Dependent variable:** Chance of Admission (0.3-0.9)

**Independent variables** that you believe affect the dependent variable: UG CGPE, GRE Score, TOEFL Score, University Rating, SOP, Research Analysis

| **Independent Variable** | | |
| --- | --- | --- |
| **Summary of independent variable** | **Categorical or quantitative?** | **The argument for / description of the associates with the dependent variable** |
| UG CGPE | Quantitative (out of 10) | Every university has set a minimum undergraduate CGPE as a mandatory admission requirement. So the variable is associated with Chances of admission |
| GRE Score | Quantitative  (out of 340). | Every university has set a min GRE Score is a mandatory admission requirement for International students. So the score is associated with Chances of admission |
| TOEFL Score | Quantitative  (out of 120). | Each university has set a min TOEFL Score, a mandatory admission requirement for International students to demonstrate their English Language Proficiency. So the score is associated with Chances of admission |
| Research | Categorial  (0/1) | This attribute indicates whether a student has done research under graduation. This attribute might influence the chances of admission |

#### **Part Two—Map Decisions to Outcomes**

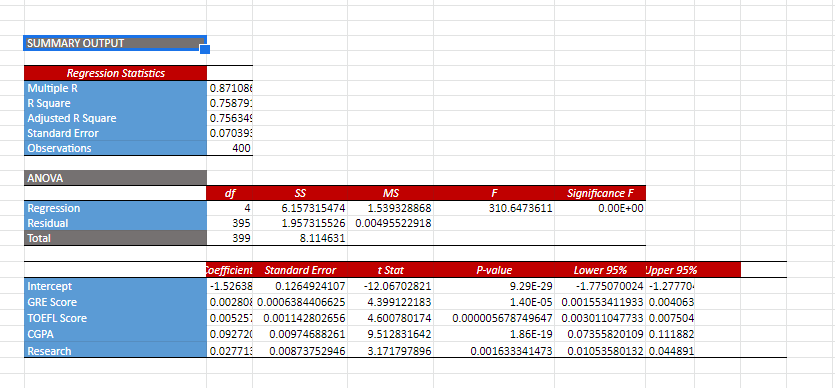
| **Candidate Independent variables** | | |
| --- | --- | --- |
| **Independent variable** | **Regression equation** | **Screenshot of scatterplot** |
| UG CGPE | y=0.2088x-1.071 |  |
| GRE Score | y = 0.01x-2.436 |  |
| TOEFL Score | y = 0.018x-1.2734 |  |
| Research | y = 0.1583x+0.6377 |  |

* CGPE: From the scatter plot where it is clear that increasing CGPA increases the Chance of Admission. People having CGPA near nine are likely to get admission.
* GRE Score: From the scatter plot, it indicates that on increasing GRE Score, the Chance of Admission is also Increasing.
* Toefl Score: From the scatter plot, it is clear that the Chance of Admission increases on increasing TOEFL Score.
* Research:
  + Out of 400 Students, 219 have done Research during UG, while 181 have not done it.
  + 55% of Students have done Research during their UG Program.
  + Students who have done research during UG are more likely to get admission than those who have not done.

1. Multiple Regression

The output for the simple linear regression model, with a single feature as the “GRE Score,” explained 62.9% of the variance in Admit, which is the output Label(The meaning of that is 62.9% of the times the Admit will change with the change of GRE Score). Additionally, our predictions were within ± 0.059 Admit score.

However, in the multiple linear regression model, 75% of the variance in Admit Score was explained by the combination of features.



#### **Part ThreeGenerate a Revised Regression Equation**

1. **For each** of the independent variables listed in part 1, residual plot against the dependent variable “Chances of Admission”

##### Residual Plot

| **Residual Plot** | |
| --- | --- |
| **Independent variable** | **Screenshot of Residual plot** |
| UG CGPE | Chart |
| GRE Score | Chart |
| TOEFL Score | Chart |
| Research |  |

Based on above residual plots, the different values of attributes, the residuals are loosely distributed in and around zero. So they doesnt represent any relation between residuals and attribute values. This means that the residuals would show the linearity of the target and each predicted variable.

##### Addressing Multicollinearity

1. A correlation table for independent variables.

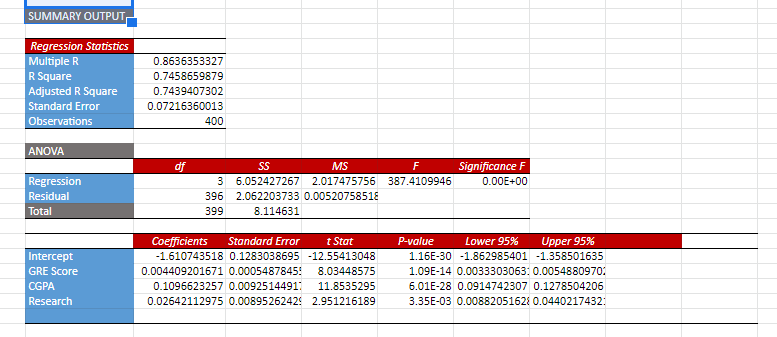
|  | *GRE Score* | *TOEFL Score* | *CGPA* | *Research* | *Chance of Admit* |
| --- | --- | --- | --- | --- | --- |
| GRE Score | 1 |  |  |  |  |
| TOEFL Score | 0.835976803 | 1 |  |  |  |
| CGPA | 0.7938843257 | 0.789385402 | 1 |  |  |
| Research | 0.5803906362 | 0.4898578507 | 0.5134070058 | 1 |  |
| Chance of Admit | 0.8026104596 | 0.7915939869 | 0.8256743101 | 0.553202137 | 1 |

* There is a Strong Positive Correlation of 82 % of CGPA with Chance of Admission. This means the Chance of admission increases with an increase in CGPA
* There is a Strong Positive Correlation of 80 % of GRE score with a Chance of Admission. This means the Chance of admission increases with an increase in GRE Score.
* There is a Strong Positive Correlation of 79 % of the TOEFL score with a Chance of Admission. This means the Chance of admission increases with an increase in TOEFL Score.

Correlation between Independent variables should be less than (correlation with independent variable)

* + GRE & TOEFL correlation>0.8- **Multicollinearity**
  + GRE & CGPE< 0.8
  + GRE & Research <0.8
  + TOEFL& CGPE<0.79
  + TOEFL& Research <0.79
  + CGPE & Research<0.82
* From above correlation table, Attributes GRE,TOEFL> 0.8. (Multicollinearity exists).So generating a revised model using attributes GRE, CGPE and Research.

##### REVISED MODEL

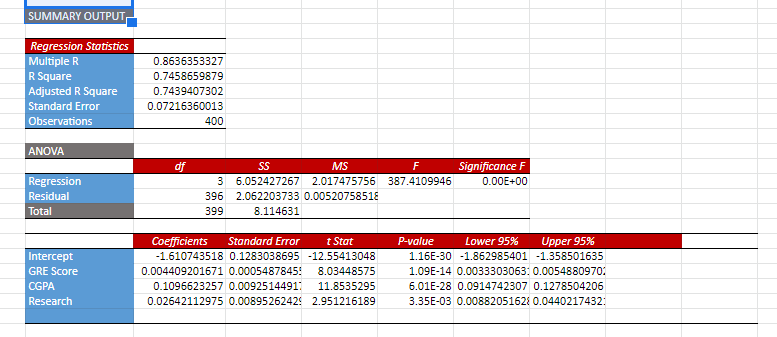


**NEW MODEL INTERPRETATION**

* P VALUE OF GRE,CGPA AND RESEARCH IS LESS THAN 0.05 SO THEY ARE SIGNIFICANT VARIABLE.
* SLOPE OF SIGNIFICANT IND. VAR. :- A. GRE SCORE : 0.0044 C. CGPA : 0.109 D. RESEARCH : 0.026
  + A) INCREASE IN GRE SCORE WILL INCREASE CHANCE OF ADMISSION. C) INCREASE IN CGPA WILL INCREASE CHANCE OF ADMISSION D) INCREASE IN RESEARCH WILL INCREASE CHANCE OF ADMISSION
* R SQUARED IS 0.745 WHICH MEANS 78% OF DEPENDENT VARIABLE IS EXPLAINED BY INDEPENDENT VARIABLE.
* ADJUSTED R SQUARED IS 0.743 WHICH IS LESS THAN R SQUARED.

#### **Part Four—Validate Model**

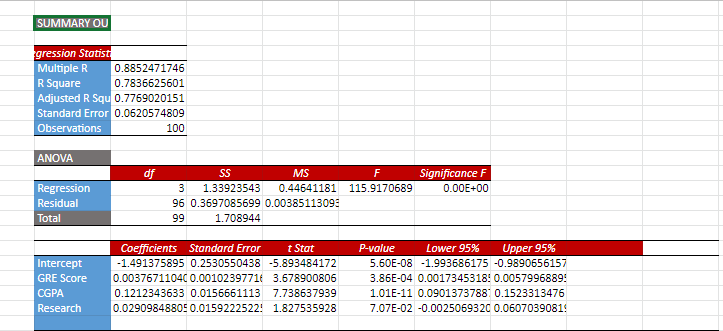
Final Model :



Testing Model with a Holdout Sample

* Tested the model using a holdout sample, providing extra confidence.

Performed Holdout sampling test, by randomly selecting 100 observations, below is the regression output.



* P VALUE OF GRE,CGPA AND RESEARCH IS LESS THAN 0.05 SO THEY ARE SIGNIFICANT VARIABLE.
* SLOPE OF SIGNIFICANT IND. VAR. :- A. GRE SCORE : 0.003 C. CGPA : 0.1219 D. RESEARCH : 0.029
  + A) INCREASE IN GRE SCORE WILL INCREASE CHANCE OF ADMISSION. C) INCREASE IN CGPA WILL INCREASE CHANCE OF ADMISSION D) INCREASE IN RESEARCH WILL INCREASE CHANCE OF ADMISSION
* R SQUARED IS 0.78 WHICH MEANS 78% OF DEPENDENT VARIABLE IS EXPLAINED BY INDEPENDENT VARIABLE.
* ADJUSTED R SQUARED IS 0.77 WHICH IS LESS THAN R SQUARED.
* Residual between predicted Y(Chances of Admit) and Actual Y is very low

| Chance of Admit | Predicted Y | Residula |
| --- | --- | --- |
| 0.74 | 0.705278449 | 0.03472155105 |
| 0.77 | 0.6647485502 | 0.1052514498 |
| 0.9 | 0.8566496955 | 0.04335030454 |
| 0.75 | 0.6748375378 | 0.07516246223 |
| 0.66 | 0.6726730095 | -0.01267300952 |
| 0.71 | 0.6825018381 | 0.02749816191 |
| 0.7 | 0.6915085615 | 0.008491438454 |
| 0.52 | 0.5852073651 | -0.06520736508 |

However, the SOP, LOR, and University ranking attributes were not considered in the model due to lower predicting capacity.

**Conclusion: Analysed graduation admission dataset with 400 observations to predict Chances of Admissions.During analysis found the GRE Score , CGPA and research attributes were important to predict chances of admissions. Generated a multiniear model using these attributes. Tested the model with random holdout sample.The residuals between chances of admission and predicted Chnaces of admission was very low. Also the model had very less standard error and higher co-efficent values. Hence Linear regression can be used to predict Graduation Admissions for International Students.**