Ashley Torres

Assignment 2 Logistic Regression

---

1. Are the classes (admit/ don't admit) balanced in the data set?

The classes (admit/don't admit) are not balanced in the data set. Below is how it is divided.

0	**1**
0.6825	0.3175
273	127

2. How would you describe the distribution of GRE scores? Is it skewed or approximately normal?

The calculation of skewness is -0.143275741178172. This number is very close to zero meaning it is nearly symmetrical(are nearly balanced).

3. Which variable in the data is the most important for predicting admission status?

Using this data I would say the **rank** is the most important because it has the smallest p value meaning it is very strong evidence against the null hypothesis.

```
Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.091615   1.296665  -2.384 0.017112 *
gre          0.000809   0.001253   0.646 0.518578
gpa          0.903378   0.384148   2.352 0.018691 *
rank        -0.502850   0.151783  -3.313 0.000923 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

OWN Exploratory Data Analysis (EDA)

From this data I can conclude that the variables gre, gpa, and rank have some sort of relationship whether someone is admitted or not. The Rank seems to be the most important variable whether someone is admitted or not. After that someone's GPA seems to be the second most important variable in the admission process.