Deliverable 4

Nathaniel Jones

# **Table of Contents**

EXECUTIVE SUMMARY  
- **Hypothesis**  
- **Summary of Findings**  
- **Actionable Items**

# **Appendix**

1. DATA MANIPULATION
2. DO THE RESIDUALS FOLLOW A NORMAL DISTRIBUTION?
3. NONPARAMETRIC TEST: Permutation Test of the Slope/Correlation.
4. VISUALS with categorical variables

# **Executive Summary**

**Summary of Findings:**

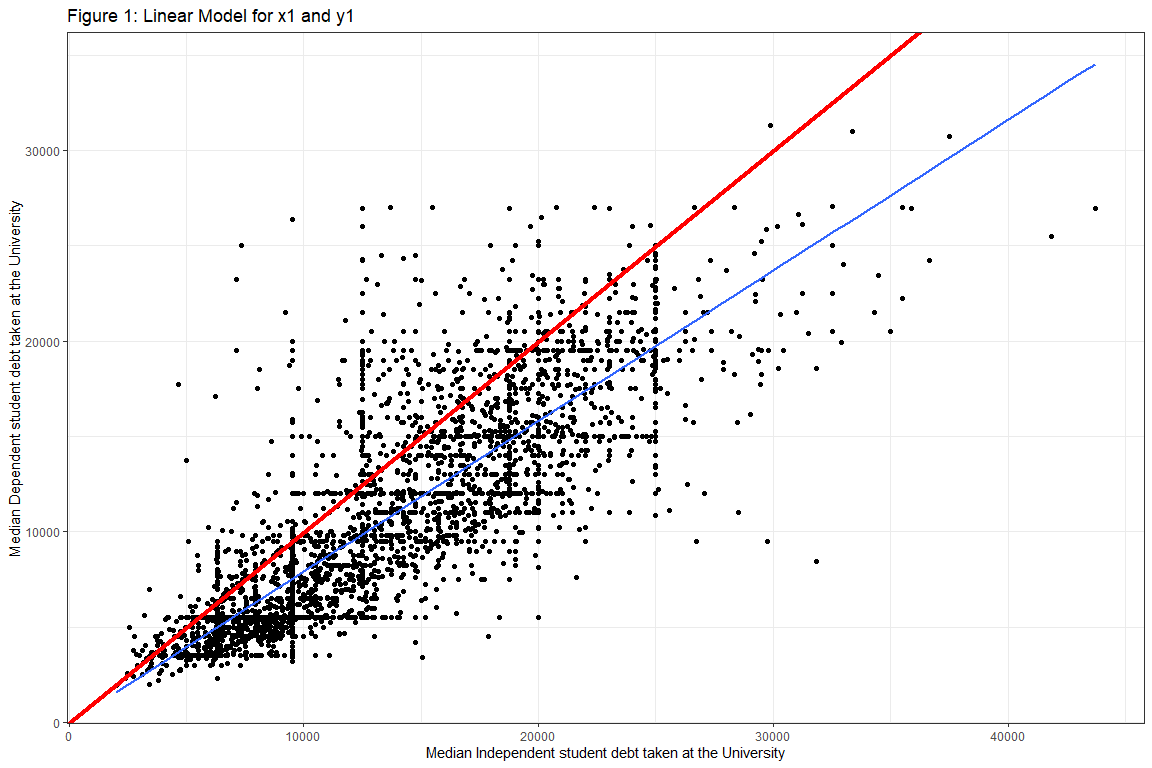
**Hypothesis:** Do students who are independent from their parents take on the same amount of debt as students who are dependent?

Students that are financially dependent on their parents are offered less in federal loans by the government than their independent peers. The “average” student debt at a university is calculated for both independent and dependent students.

The data shows that the median debt accumulated by an independent student at a university is higher than the median debt accumulated by a dependent student at the same university. In fact, for every unit increase in the median independent student debt accumulated at a university, the median dependent student debt at the same university will increase by 0.7892. A Nonparametric test was chosen as the test procedure due to the skewness present in the residuals of the model.

**Actionable Items:**

* A positive relationship between the median dependent student debt and the median independent student debt was found (as seen in Figure 1 represented by a blue line).
* While a perfect relationship (represented by a red line) appears as a 1:1 function starting at zero, the slope of the actual relationship between the variables is more gradual.
* Thus, the median independent student debt accumulates faster than the median dependent student debt.



## **A. Data Manipulation**

            There were 4,597 observations used in this report. These observations contained no missing values for the variables representing the median dependent student debt, the median independent student debt, and the Majority or Minority Pell category.

## **B. DO THE RESIDUALS FOLLOW A NORMAL DISTRIBUTION?**

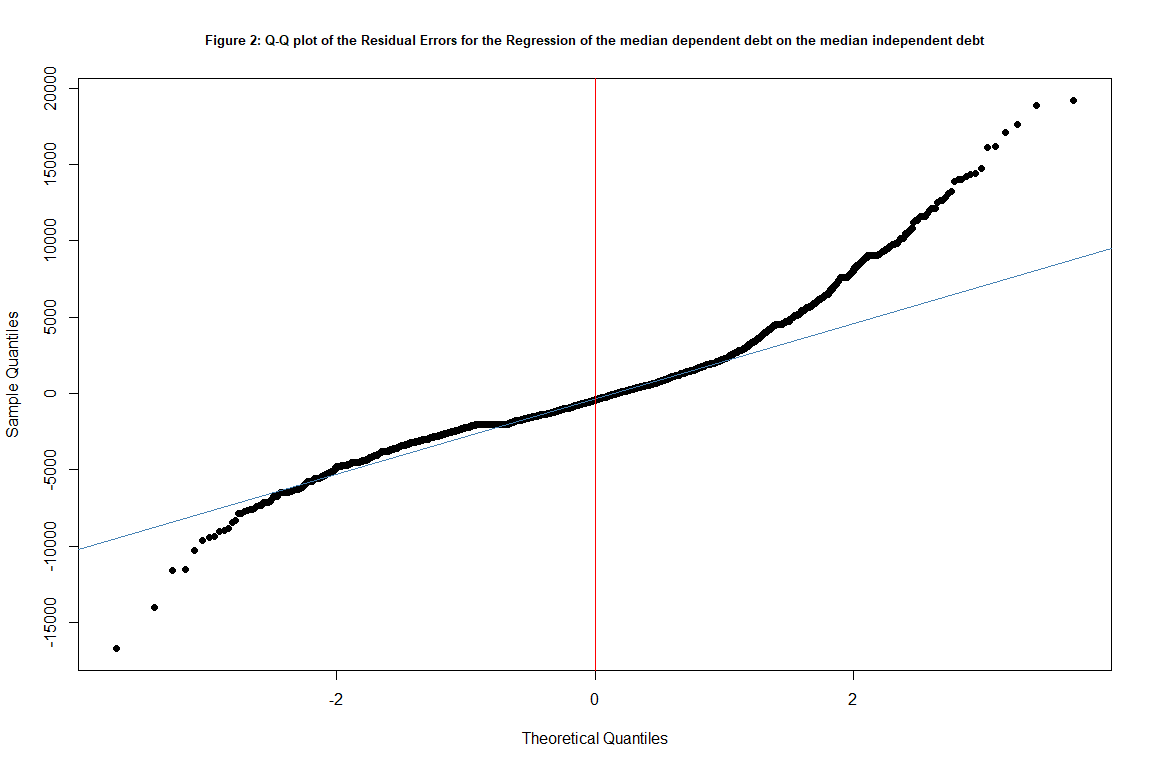


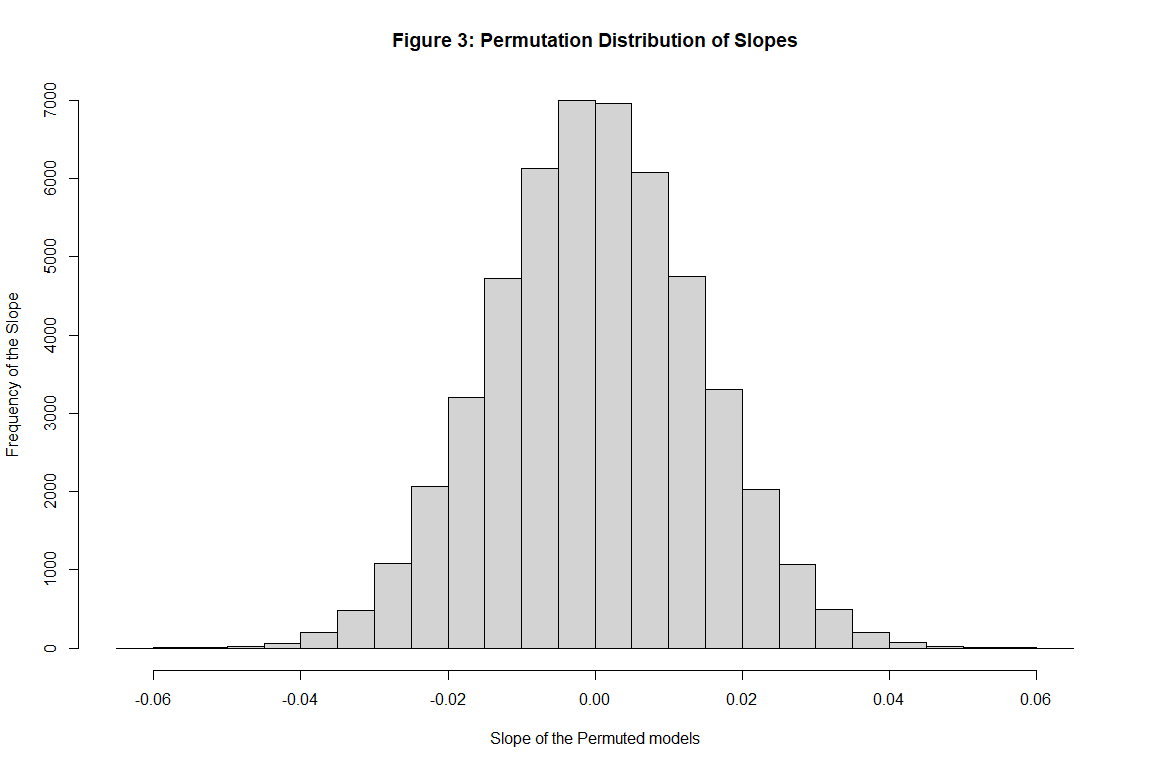
Figure 2 displays the Q-Q plot of the Residual Errors for the model predicting the median dependent student debt at a university. This plot shows a strong curve at the extreme ends of the distribution suggesting that the tails are skewed. A KS test are performed on the residuals and found a p-value of 0. Since this p-value is less than 0.05, the results of the KS test is that the shape of the distribution is different from a normal distribution. Therefore, the residuals of the model are not normally distributed.

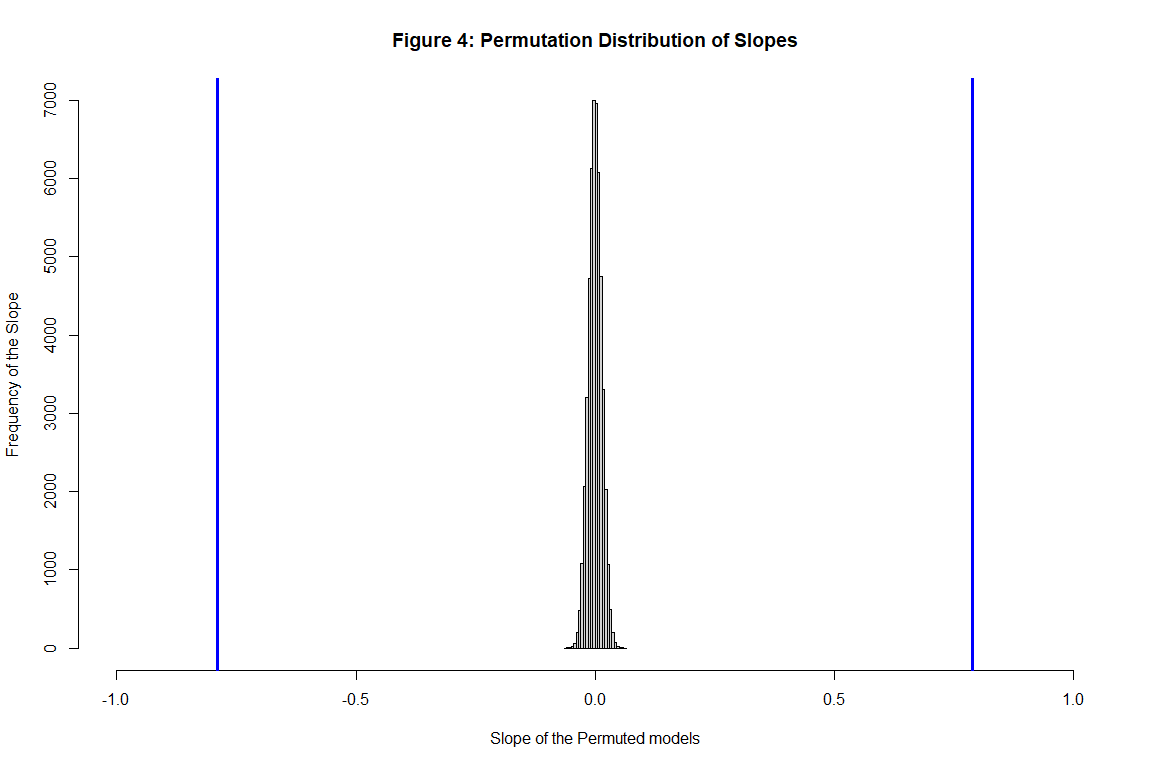
## **C. NONPARAMETRIC TEST: Permutation Test of the Slope/Correlation.**

Since we found that the residuals were not normally distributed, utilizing the Nonparametric Permutation test is suggested over the Parametric Simple Linear Regression test. The observed test statistic for the slope is 0.7892. There is a 0% chance that I will observe 0 or more permutations out of 50000 permutations to be above the observed slope, 0.7892 or its negative (-0.7892), if the true slope is 0. With this information, we can see that the assumption that the true slope is 0 has been violated, indicating that the true slope is not zero. In terms of the variables, the median independent student debt at a university is important in predicting that university’s median dependent student debt. In fact, according to this model, for every unit increase in the median independent student debt the median dependent student debt will increase by 0.7892.

## **D. Histogram of the permuted test statistics**

The histogram of the permuted slopes in Figure 3 is created by randomizing y values and then computing the slope from the randomized y values and the original x values. We can see that the majority of the permuted slopes are less than 0.1 in magnitude. Since the permuted slopes were created with the assumption that there is no relationship between the variables, we would expect the observed test statistic to be randomly placed within the distribution of permuted slopes. The sample’s observed test statistic is 0.7892, indicated by vertical blue lines at -0.7892 and 0.7892 in Figure 4.





## **PARAMETRIC TEST: Simple Linear Regression and Test of Slope/Correlation.**

The Q-Q plot indicates that the residuals of the model are not normally distributed. Since the Parametric Regression test of Slope/Correlation is sensitive to skewness, the Nonparametric test of Slope/Correlation was used to test this sample.