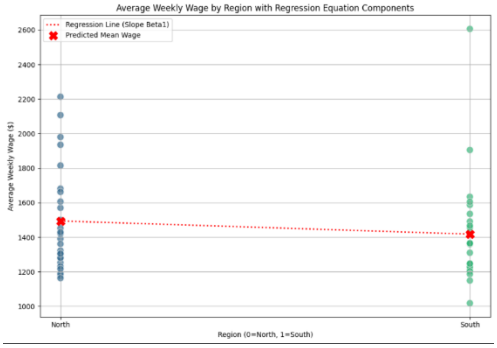
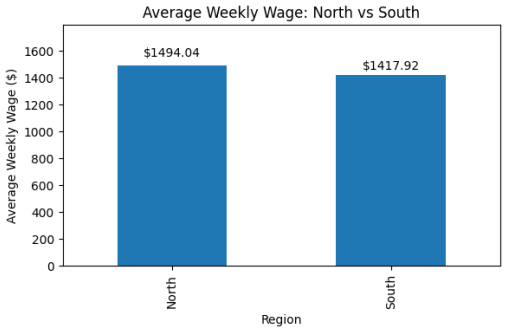
1. Introduction and Question

   The question our group is trying to answer is whether the region of the U.S. someone lives in (North or South) impacts their salary.

2. Data Description and Sources





   Our two visualizations were a bar chart comparing average weekly wages in Northern and Southern U.S., as well as the regression plot showing the relation between the two variables. We got our data from the U.S. Bureau of Labor Statistics, a very reliable source. To prepare our data, we just had to take the average weekly wage of each state and categorize each state as either North or South.

3. Methodology

   Our GLM is: Wage=B0+B1(Region)+e

   This model is a binary model, meaning the Region variable is either coded as 0 for North or 1 for South. B0 represents the average weekly wage for people living in the north. B1 represents the difference between the average weekly wage in the North and the average weekly wage in the South. ‘e’ represents the standard error. The main limitation of this

GLM is that it leaves out many other variables that could be reasons for differences in salary between the North and South.

4. Results and Analysis

   According to our Regression Results, B0=1494.0385, B1=-76.1185, and the p-value for B1 was 0.379. These results mean that the average weekly wage for people in the North is $1494.04, and the average weekly wage for people in the south is -$76.12 less than that, so $1,417.92. The p-value is very high, well above the commonly accepted threshold of .05, so that means that our evidence is not nearly significant enough to reject the null hypothesis. Put simply, it would be very plausible for us to find this data even if there is no correlation between region and wage.

5. Conclusions

   We attempted to answer whether or not living in the North or South of the United States impacts wages. To do this we first had to find data including the weekly wages in every state. We then categorized each state as either Northern or Southern, and were then able to create a GLM to find the coefficient to our region variable. Doing this, we discovered that the difference in wages between the North and the South is not statistically significant enough to reject the null hypothesis that region of U.S. does not impact wage.

6. References

   We got our data from the U.S. Bureau of Labor Statistics.

https://www.bls.gov/regions/northeast/news-release/countyemploymentandwages_vermont.htm