Final Report: Emalee Schuler

## Background/Motivation for the Study

The phrase “money can’t buy happiness” has always intrigued me. To me, it has always seemed to disregard the significance of wealth in American society. When I was looking through the variables in the NHANES data set, I saw that depression was one of the reported variables along with median income. This piqued my interest and reminded me of the aforementioned phrase, and I wanted to see if there was a statistically significant difference in reported depression and median income.

When researching about previous studies on incidences of higher rates of depression in those with less wealth, I found a 2018 study titled, “Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms.” This study reported, the “pooled risk ratio was 1.19 (95% CI: 1.07‐1.31), [which demonstrated] greater risk of depression in populations with higher income inequality relative to populations with lower inequality (Patel et al., 2018).” I wanted to see if the data I gathered in this assignment would support their conclusion.

Patel, V., Burns, J. K., Dhingra, M., Tarver, L., Kohrt, B. A., & Lund, C. (2018). Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *17*(1), 76–89. https://doi.org/10.1002/wps.20492

## Research Question and Hypothesis

Question: Is there a notable difference in the prevalence of depression between individuals with lower income and those with higher income in the NHANES dataset?

Hypothesis: The prevalence of depression is greater in individuals with lower income than those with higher income in the NHANES dataset.

Null Hypothesis: There is no significant difference in the prevalence of depression between individuals with lower income and those with higher income in the NHANES dataset.

Alternative Hypothesis: The prevalence of depression does not differ between individuals with lower income and those with higher income in the NHANES dataset.

I think my research will show that the prevalence of depression is higher in individuals with lower income compared to those with higher incomes. I going to try to show that their is a statistically significant difference between the two groups.

## Data Description and Exploratory Data Analysis

**Targeted Variables:**

* **HHIncomeMid**: This column represents income categories (e.g., $2,500, $7,500, $12,500, etc.), which are grouped by median income ranges.
* **depressed\_atall**: This column represents depression status, with two categories, “no” and “yes,” indicating whether an individual reported depression. Mutated from Depressed variable to indicate whether an individual indicated feeling depressed either “Most” or “Several” days.

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':  
  
 filter, lag

The following objects are masked from 'package:base':  
  
 intersect, setdiff, setequal, union

* **HHIncomeMid**: This column represents income categories (e.g., $2,500, $7,500, $12,500, etc.), which are grouped by median income ranges.
* **depressed\_atall**: This column represents depression status, with two categories, “no” and “yes,” indicating whether an individual reported depression. Mutated from Depressed variable to indicate whether an individual indicated feeling depressed either “Most” or “Several” days.

Registered S3 method overwritten by 'mosaic':  
 method from   
 fortify.SpatialPolygonsDataFrame ggplot2

The 'mosaic' package masks several functions from core packages in order to add   
additional features. The original behavior of these functions should not be affected by this.

Attaching package: 'mosaic'

The following object is masked from 'package:Matrix':  
  
 mean

The following objects are masked from 'package:dplyr':  
  
 count, do, tally

The following object is masked from 'package:ggplot2':  
  
 stat

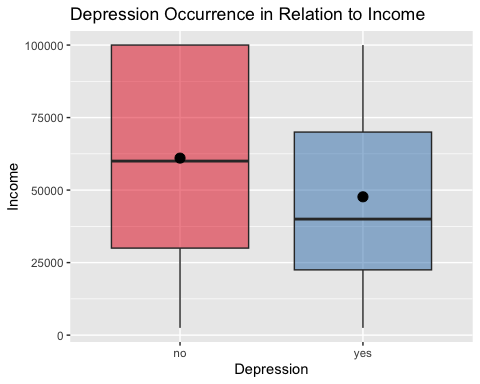
The following objects are masked from 'package:stats':  
  
 binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,  
 quantile, sd, t.test, var

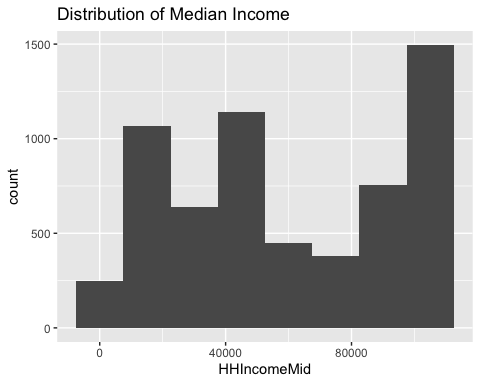
The following objects are masked from 'package:base':  
  
 max, mean, min, prod, range, sample, sum

depressed\_atall min Q1 median Q3 max mean sd n missing  
1 no 2500 30000 60000 1e+05 1e+05 61027.48 32102.69 4876 0  
2 yes 2500 22500 40000 7e+04 1e+05 47683.82 31952.35 1292 0

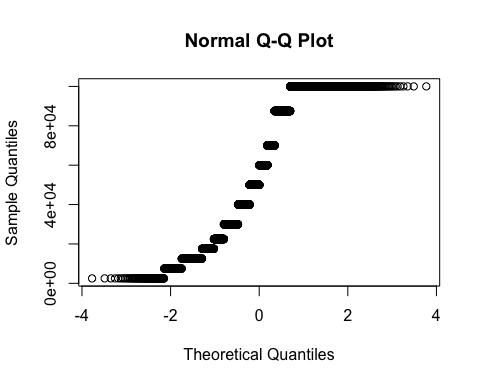
The contingency table breaks down how many individuals within each income category reported any depression or no depression.

4876 respondents denied any depression, and 1292 respondents stated they experience(d) depression. The average income for individuals reporting “yes” was $47,683.35, and individuals who reported “no” had an average income of $61,027.48.





[1] -0.001532872



Shapiro-Wilk normality test  
  
data: my.ctable  
W = 0.45926, p-value = 8.11e-11

Since the distribution of Income is not strongly skewed, the population sample is over 30 (the Central Limit Theorem is satisfied), and the normality plot indicates the data is not distributed normally (as indicated by Shapiro Test with p-value less than 0.05), a confidence interval based on the t-distribution is appropriate.

The box plot shows that the average income for those that reported “no” depression was higher than those who did report depression.

## Analysis

Welch Two Sample t-test  
  
data: HHIncomeMid by depressed\_atall  
t = 13.333, df = 2035.4, p-value < 2.2e-16  
alternative hypothesis: true difference in means between group no and group yes is not equal to 0  
95 percent confidence interval:  
 11380.99 15306.33  
sample estimates:  
 mean in group no mean in group yes   
 61027.48 47683.82

## Conclusions

The 95% confidence interval, [11380.99,15306.33], does not include 0, and the p-value is extremely small. This suggests strong evidence that there is a significant difference in average income between individuals who had reported depression and those who did not. The average income in the group without depression is approximately $61,028, while the average income in group where individuals experienced any depression is approximately $47,684. My hypothesis was accurate. It can be concluded that there is a difference in the prevalence in depression between those with lower and higher incomes. Specifically, depression is more prevalent in those with lower incomes and less prevalent in those with higher incomes.