Matthew Ye

October 17, 2018

<p style="text-align:center">Relationship between sex and income in the United States</p>

It has long been observed and claimed that males make more income than females doing the same job. This phenomenon is known as the gender wage gap. In the long movement to create equity for men and women, the wage gap is one of many issues we face in our society. It is important that we analyze pertinent data that can effectively determine if such issues are present, and if so, the level of importance of the issue.

The target population are people working and living in the United States. I hypothesize that males, on average, make more money than women. This memo will show that males do have a higher income on average than females, and it is statistically and substantively significant.

To investigate this problem, I used the nationally representative data in the 2016 General Social Survey (GSS). Specifically, I looked at respondent's sex and their annual income. Sex/gender was used as the independent variable and income was used as the dependent variable. 2876 people were polled about their sex, responses only include "Male" or "Female." 1632

Table 1: Sex Summary

	Freq.	Percent
Male	1276	44%
Female	1591	56%
Total	2876	100%

people were polled about their income level. Responses range from "Under $1000," to "$25,000 to $30,000," up to "$170,000 or over." There were 26 total possible responses. Income's had a minimum of <$1000, a max of >$170,000, a median of $32,500, a mean of $45,125, and a standard deviation of 39,226. There was some recoding I had to do. I needed to change the ordinal nature of the income variable into a quantitative variable. As such, I employed the

Table 2: Income Summary

Freq	Min	Max	Med.	Mean	SD
1,632	1,000	170,000	32,500	45,125	39,226

recoding to the midpoint technique, which took the ranges of each response, and assigned it to a numeric value precisely in the middle of the two numbers. For example, the factor level of "$6000 to $6999" was recoded to a numeric value of "6500". This was to get a more precise value when taking the mean.

There were a lot of missing values for the income response category. This may have impacted our results and analysis as it is possible for the data to lack the representation needed

for fair analysis of all males and females. It's possible that within the missing data, the gap between male and female wages could've closed.

We found evidence that males have a higher income on average than females. The difference between the average incomes was nearly $16,000. Along with this substantive significance, our Independent Samples t-test revealed the difference in incomes to be statistically significant at the 0.05 alpha level, $t(1630) = 8.41$, $p<0.05$. The Cohen's d effect size is 0.42, which is close to a medium effect size. When conducting a variance test at alpha = 0.05, we failed to reject the null hypothesis that the variances were equal.

Mean Income by Sex

$53,297.31

$37,305.46

Male   Female

Sex
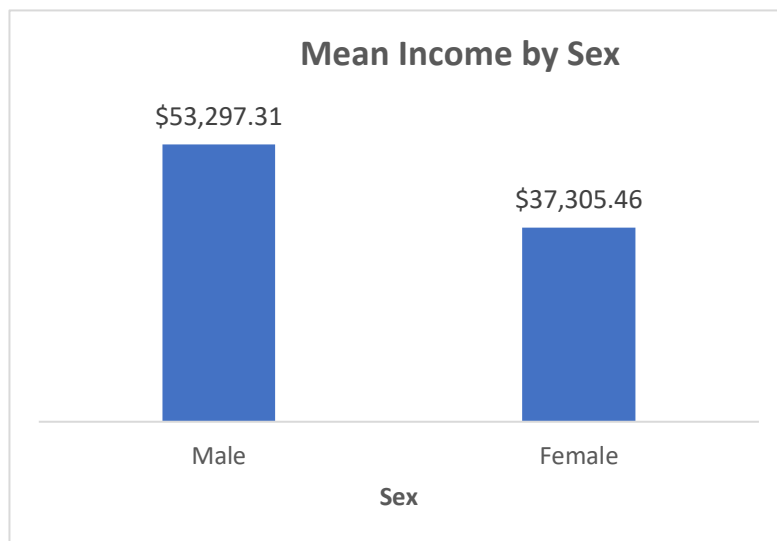
The difference in means was large, which contributed to the medium sized Cohen's d. Since our results were both statistically and substantively significant, we can safely reject the null hypothesis that there is no difference in the means of income between men and women. Thus, I can recommend generalizing our finding to the population.

This study does have its limitations. Almost half of the population sampled did not answer to the income variable, thus losing a large amount of our results and having not entirely representative data. Since there were so many missing values, it's possible that within those missing values, lies data that actually supports the null hypothesis. In the future, we may only accept data that has a greater than 75% response rate to ensure accurate and representative data.