A Regression Analysis of the Gender Pay Gap

Maria-Cristiana Gîrjău

Revised on 2019-09-15

This document should include all the code used to wrangle and analze your data. You should comment your code within each code chunk, and you should document your process as you go. This means describing what each code chunk is doing if it's not immediately obvious (e.g., "After reading in the data, we noticed that this categorical variable was being treated as a numeric variable, so we forced it to be a factor instead."), every decision you made (e.g., "In the code below, we dropped every person over 90 years old and collapsed such-and-such categorical variable down to three categories instead of 10."), and what you are gathering from any output (e.g., "The histogram above shows us that the data are skewed right with one extreme outlier around 300."). Only set `echo=F` for code you do not want to receive feedback on. Use `eval=F` if there is code that is not working and that you want help with!

Data wrangling

Get the data into the format needed for analysis. Any time in other sections that you realize your data should actually look a different way (e.g, maybe you decide you really want to collapse a categorical variable or something), make sure to come back to this section to implement that change from the beginning. This will be an iterative process!

```r
# Read in dataset
acs_sample_raw <- read_csv("data/acs230_3k.csv")
```

Data exploration

Explore distributions and associations graphically and numerically.

Data analysis

Perform analyses.

Assessment

Current questions

If you are running into issues or have wrangling problems outside of your skillset, let me know here how I can help! You can also ask for pointers on how to implement a method that maybe we haven't covered in course but you would like to include.