

Homework 6: Hypothesis Tests

Answer the following questions in a .pdf or .docx, explaining all of your answers and putting any tables and figures in the document as necessary. When data is called for to answer applied questions, I will provide it in bblearn. Turn in your R code that created all of the tables and figures separately, and be sure that it runs from source in such a way that it loads the data and performs all the tests without me fiddling with it. Make sure to document your R source code using `#` comments if you want partial credit.

Because on one question we are using simulation/random draw methods in this assignment, please be sure to set your seed in your homework. Please set the seed using `set.seed(4534)` in your homework assignment. That means we will get the same answers when we perform random sampling, bootstrapping, and other simulation methods.

For hypothesis tests, we will be using a dataset on wage compensation of freelance and contract workers with different characteristics and employment sectors (`freelancers.csv`). The codebook for these variables and what their values mean is provided. All t-tests will assume unequal variances (the default in R) and have a 2-sided alternative hypothesis.

1. First, we will perform several theory-based one sample t-tests on the mean of the income variable for all workers in the dataset. Can we conclude that the population income level this is drawn from is different from 50,000 at a criteria of .10? What about the a criteria of .01? What is the p-value of the test?
2. Now, perform the same hypothesis tests as above instead using the resampling method with 1000 resamples, answering the same questions. Remember to set your seed as requested.
3. Now, we will perform difference of means t-tests comparing several different groups. First, can we conclude that women and men receive the same income in our sample at a .10 level of statistical significance? What is the p-value of this test?
4. Now we turn to the difference between industries. First, create an indicator variable for each sector of the categorical variable `sector` by installing and loading the library `fastDummies` and then executing the following code:

```
freelancers <- dummy_cols(freelancers, select_columns = "sector")
```


With the resulting the data, is there evidence at the .10 level that information technology workers make the same income as all other workers? What is the p-value of that test? What about workers in finance relative to all other workers in the sample at the .01 level? What is the p-value of that test?