Data Analysis Assignment 3

COMM 7370 | Spring 2023 | Due: see syllabus

> **!** Remember to...
>
> 1. Set up your R script as you did in your previous assignment.
> 2. Include pseudocode in your R script.

Instructions

1) Load the packages we have been using in this course. Then, familiarize yourself with the CPS codebook (Table 1) and read the CPS dataset into R.
2) Determine the mean wage earned per hour among males and females. State your findings in your R script as comments. Visualize the mean wage earned per hour by sex in a figure.
3) Implement the appropriate statistical test[1] to determine whether males have significantly higher wages earned per hour than females. Start by stating your hypothesis based on the previous step. Conduct the test and report the test statistic and p-value. Does your test support or refute your hypothesis?
4) Visualize the relationship between wages earned per hour and number of years of work experience. Formulate a hypothesis about the relationship between these variables and test that hypothesis using the appropriate statistical test. Report the test statistic and p-value. Does your test support or refute your hypothesis?
5) Plot the mean of job satisfaction by job sector. Then, formulate a hypothesis about whether there are differences in satisfaction by job section and conduct and report the appropriate statistical test. Does your test support or refute your hypothesis?
6) Visualize mean wage per hour by job sector. Formulate a hypothesis about the relationship between these variables, conduct an appropriate statistical test and report your results. Does your test support or refute your hypothesis?

[1] If you need help selecting a statistical test, refer to the flowchart in Data Analysis Assignment 2.

Codebook

Data for this assignment comes from the 1985 Current Population Survey (CPS). The CPS is used to supplement census information between census years. These data consist of a random sample of persons from the survey, with information on wages and other characteristics of the workers, including sex, number of years of education, years of work experience, occupational status, region of residence and union membership. This data frame has 534 observations.

Table 1: Codebook for the CPS dataset.

Variable	Description
`wage`	Wage (US dollar per hour)
`educ`	Number of years of education
`race`	Factor with levels: NW (non-White), W (White)
`sex`	Factor with levels: F, M
`hispanic`	Factor with levels: Hisp, NH (non-Hispanic)
`south`	Factor with levels: NS (non-south), S (south)
`married`	Factor with levels: Married, Single
`exper`	Number of years of work experience
`union`	Factor with levels: Not (not part of union), Union
`age`	Age in years
`sector`	Factor with levels: clerical, const, manag, manuf, other prof, sales, service
`satisfaction`	1 = satisfied with current employment, 0 = otherwise

Submission

Submit your R script (which should have a `.R` extension) to Canvas. Your R script should:

1) Include code to install and load the packages.
2) Contain comments and/or pseudocode.
3) Run in its entirety without errors.

To ensure that your R script runs without errors, you should:

- Save your script.
- Completely shut down RStudio or restart your R session.
- Reopen RStudio and your `.R` script.
- Run the entire script by clicking the "Run" button in the top right of the R script.

> **!** Important
>
> **These standards apply to all submissions in this course that require R scripts. You should follow these instructions for preparation, naming, and saving of your R script for *all* of your data analysis assignments.**