LA-9: Practice Data Analysis (15 points)

Learning Outcomes

In this assignment, you will practice the data analysis skills that you have learned this semester.

> **Tip**
>
> Read all the instructions carefully before starting the assignment.

Instructions

1) Download the CPS dataset (`CPS.csv`) and codebook (`CPS-codebook.csv`). Use the codebook to familiarize yourself with the variables in the data and read the data into R. Install the packages that we have been using this semester.

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.

2) Find the mean wage earned per hour for males and females.
    a) What is the mean wage earned per hour for females?
    b) What is the mean wage earned per hour for males?

3) Implement the appropriate statistical test to determine whether males have significantly higher wages earned per hour than females.
    a) State your hypothesis about whether males earn higher wages per hour than females.
    b) Which statistical test will you use to address your hypothesis? Conduct the statistical test you have chosen.
    c) What is the test statistic and associated p-value?
    d) Does the test support or refute your hypothesis? Do males or females earn higher wages on average?

4) Implement the appropriate statistical test to determine whether there is a significant linear relationship between wages earned per hour and number of years of work experience.
    a) State your hypothesis about the relationship between wages earned per hour and number of years of experience.
    b) Which statistical test will you use to address your hypothesis? Conduct the statistical test you have chosen.
    c) What is the test statistic and associated p-value?
    d) Does the test support or refute your hypothesis? What is the relationship, if any, between wages earned per hour and number of years of experience?

5) Implement the appropriate statistical test to determine whether job satisfaction varies significantly among the different job sectors.
    a) State your hypothesis about whether job satisfaction varies significantly among job sectors.
    b) Which statistical test will you use to address your hypothesis? Conduct the statistical test you have chosen.
    c) What is the test statistic and associated p-value?
    d) Does the test support or refute your hypothesis? Does job satisfaction vary among different job sectors?
6) Implement the appropriate statistical test to determine whether wages vary by job sector.
    a) State your hypothesis about whether wages and job sector are related.
    b) Which statistical test will you use to address your hypothesis? Conduct the statistical test you have chosen.
    c) What is the test statistic and associated p-value?
    d) Does the test support or refute your hypothesis? Do wages vary among different job sectors?

Submission

Submit your R script (named `LA-#_FirstName-LastName.R`) to Canvas.

Your R script should:

1) Include commands and functions that are necessary to address all the questions in the assignment.
2) Contain comments that answer the questions in the assignment.
3) Run in its entirety without errors.

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

- Save your script.
- Navigate back to Your Workspace on Posit Cloud.
- Reopen your project.
- Run the entire script line-by-line without editing it to ensure there are no errors.

> **!** 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 individual lab assignments.