LA-8: 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 full COVID dataset (`covid-full.csv`) and codebook (`covid-full-codebook.csv`). Use the codebook to familiarize yourself with the variables in the data and read the data into R.

2) Create a new variable called `asia_europe`. Have the variable equal `Asia` if the country is China, India, Japan, Singapore, or South Korea. Have the variable equal `Europe` is the country is Denmark, France, Germany, Italy, or Spain.

   Obtain descriptive statistics for the number of time respondents washed their hands on the previous day for Asia and Europe.

   a) What is the mean and standard deviation of handwashes in Asia?
   b) What is the mean and standard deviation of handwashes in Europe?

3) Conduct a statistical test to determine whether there is a significant difference in the number of times people living in Asia compared to Europe washed their hands.

   a) State your hypothesis.
   b) Select and conduct a statistical test to address your hypothesis.
   c) Report the test statistic and p-value.
   d) Does the statistical test support or refute your hypothesis?

4) Implement the appropriate statistical test to determine whether there is a significant linear relationship between how many times people left their home and the total number of contacts they had with other people.

   a) State your hypothesis.
   b) Select and conduct a statistical test to address your hypothesis.
   c) Report the test statistic and p-value.
   d) Does the statistical test support or refute your hypothesis? If there is a significant relationship, describe it. Does this finding make sense? Why or why not?

5) Is there a difference in whether the majority of respondents in Asia compared to those in Europe said their lives were impacted by COVID?

   a) State your hypothesis.
   b) Select and conduct a statistical test to address your hypothesis.

c) Report the test statistic and p-value.
d) Does the statistical test support or refute your hypothesis? If there is a significant difference, describe it (e.g., which region, Asia or Europe, had more respondents who said their lives were impacted by COVID). You can use the `filter()` and `freq()` functions to help you describe the differences.

6) **BONUS:** Use the data and a statistical test to determine whether there is a relationship between the number of days since the COVID outbreak began and how often respondents wore masks outside. Conduct the statistical test, report the results of the test (i.e., test statistic and p-value) and explain your conclusion.

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.