**ENV710 APPLIED STATISTICAL MODELING FOR ENVIRONMENTAL MANAGEMENT**

**Group LAB ASSIGNMENT Comparison of Means**

DUE: 7 October 2024

Upload to knitted html and .Rmd to Canvas prior to lecture.

In this lab, you will use R to conduct comparison of means tests to examine changes in energy consumption and hospitalization rates in response to changes in the adoption of an Ecuadoran clean cookstoves program (called PEC). In this assignment, you will compare electricity consumption and hospitalization rates across cantons (which are like counties in the USA) with high and low rates of program adoption. Please turn **ONE write-up per group** (maximum 1,000 words or approximately two single-spaced pages of text (not including figures)) by the due date. The lab must be submitted in **html that was knitted from your Rmd**. Use tables to report your results in a clear and structured manner. Please use in-line R to discuss your results. Your report **should not** contain any R code, warning, or error messages. In addition to your html, you will need to submit ONE .Rmd file that contains all the code that you executed in RStudio.

In this lab, we will analyze energy consumption and hospitalization data from a study conducted by (Gould et al., 2023) that examined the effects of households transitioning from gas to electric cookstoves on energy consumption, health outcomes, and greenhouse gas emissions. The paper analyzed this transition at the parish and canton level, drawing on public Ecuadoran data from 2015 to 2021. The authors used a method called difference-in-differences to track how energy consumption (measured in kilowatts per hour in a geographic area) related to enrollment in the program. After establishing that the program increased energy consumption by offsetting cooking gas consumption, the authors argue that increased electric (induction) stove uptake from the program was associated with declines in all-cause hospitalizations, respiratory hospitalizations, and likely reduced Ecuador’s total greenhouse gas emissions.

Interrogating many of these claims would require an understanding of causal methods and an ability to merge datasets that we have not yet covered in this class and so we have developed a simplified dataset to make inferences about the effect of program adoption on health effects.

Our goal is to make inferences about energy consumption and hospitalization rates across areas with low and high program adoption by conducting the following comparisons **(NOTE: If you are a group of two people, please only focus on energy consumption.**

**Energy Consumption**

1. Energy Consumption Levels in high adoption areas vs Energy Consumption Levels in low adoption areas (all years).
2. Energy Consumption Levels in high adoption areas in the first year vs. Energy Consumption high adoption areas in the last year
3. Energy Consumption Levels in **low** adoption areas in the first year vs. Energy Consumption **low** adoption areas in the last year

Based on the comparisons above, discuss the claim that program adoption is associated with increases in energy consumption. Be sure to discuss which comparison test(s) are most appropriate in this context.

**Hospitalization rates**

Compare changes in hospitalization rates (total hospitalizations per 100,000 canton residents) across the time period in high adoption areas vs. changes in hospitalization rates in low adoption areas across the time period.

* 1. Compare first year (2015) hospitalization rates in high adoption cantons with last year (2020) hospitalization rates in high adoption cantons
  2. Compare first year (2015) hospitalization rates in low adoption cantons with last year (2020) hospitalization rates in low adoption cantons

Compare these differences to engage with the argument that program adoption is associated with decreases in all-cause hospitalization rates

Note on hospitalization rates: This is the number of hospitalizations per 100,000 residents of the canton (for that month-year)

**NOTE: If you are a group of two people, please only do the energy consumption analysis.**

**Data analysis instructions**

1. Download the PEC data from Canvas (called PEC.csv). Be sure to look over the data and then save as a .csv file before reading into RStudio. The units are per canton-per month per year.

2. Summarize and visualize the data by groups as outlined above. Present descriptive statistics in a professional table or tables. Include your graphics in a clearly labelled appendix.

3. Conduct the appropriate comparison tests to determine for the comparisons above.

4. Interpret and discuss the results of each of the tests.

5. Examine and discuss the validity of the assumptions of your comparison tests. Remember to consider the transformation when interpreting the results of the transformed data set. Discuss your choice of statistical test and its validity in this context.

6. Consider and discuss the threshold picked by the authors to compare high and low adoption areas. Do you agree with this threshold?

7. Read the paragraph on “Patterns of Induction Stove Program Enrollment” on page 2 of the paper and discuss the paper results in light of the concept “selection bias”.

**Professional Report Format (no more than 1000 words)**

Your report should be structured with the following sections:

1. Introduction (10 points)

a. Broad questions

b. Data source

c. Variables in the data set

2. Data Description (20 points)

a. Descriptive statistics

b. Discuss distributions of data

c. Discuss censored observations

d. Refer to figures in an appendix

3. Statistical Analysis and Discussion (30 points)

a. Comparison of means results

b. Discussion of assumptions

c. Discuss which tests are most appropriate in this context.

4. Conclusion (10 points)

a. Scope of inference (to what population can you infer?)

b. Limitations and weaknesses of study/analysis

c. Real-world implications

5. Appendix (20 points)

a. Include relevant figures

6. Professional writing (10 points)

a. Clearly structured professional report

b. Clearly labelled and professional tables and graphics

c. Concise writing

Reference:

Gould, C. F., Bejarano, M. L., De la Cuesta, B., Jack, D. W., Schlesinger, S. B., Valarezo, A., & Burke, M. (2023). *Climate and health benefits of a transition from gas to electric cooking*.

Attribution and acknowledgement to Danny Tobin for his work on this assignment.