Homework 2: R Practice

Prof. Tzu-Ting Yang

November 19, 2024

INSTRUCTION

- You should submit the Homework 2 before 12/8 (Sunday midnight).

- Total grade of Howework 2 is 100. Question 1 accounts for 5, Question 2 accounts for 10, Question 3 accounts for 20, Question 4 accounts for 10, , and Question 5 accounts for 50

- This homework will help you practice R commands and let you work on your term paper.

- Your answers should include the R commands you use, explain their meanings, and your findings.

- Please upload your **answer sheet** and **log file** using this link: **https://www.dropbox.com/request/pUezbaMeolnK3rFxfO4L**.

- Format of the file name: StudentID_YourName. For example, r09323010_TzuTingYang

1 READ DATA

1. Please describe how you load the data into R. *Hint:* You can use commands such as **read.csv()**, **read.table()**, **read_excel()**, or **read_dta()** to load your data. Remember you might need to install and load packages (e.g., **readxl** or **haven**) first.

2 Examine Data

1. Check if there are any missing values for any variables in the dataset. Describe your findings. *Hint:* You can use functions such as **is.na()**, **complete.cases()**, or **summary()** to identify missing values.

2. Examine whether there are any exact duplicate observations, considering the values of all variables. Describe your findings. *Hint:* You can use **duplicated()** or **distinct()** from the **dplyr** package.

3 Create Sample for Analysis

1. Briefly describe three R commands you used during the data cleaning process. *Hint:* The following commands from the **tidyverse** packages could be helpful:

   - **mutate()**: Create new variables or modify existing ones
   - **summarise()**/**summarize()**: Aggregate data and create summary statistics
   - **group_by()**: Group data for subsequent operations
   - **filter()**: Select observations based on conditions
   - **select()**: Choose specific variables to keep or remove
   - **arrange()**: Sort data based on variables
   - **bind_rows()**: Combine datasets by stacking (similar to append)
   - ***_join()**: Merge datasets (**left_join**, **right_join**, **inner_join**, **full_join**)
   - **pivot_wider()**/**pivot_longer()**: Reshape data between wide and long formats
   - **as.factor()**/**as.numeric()**: Convert between data types

4 Visualize Data

1. Create a graph that can represent one of the findings in your term paper. *Hint:* You can use **ggplot2** with **geom_point()**, **geom_line()**, or other appropriate geometries.

2. Please explain the main findings of this graph

5 Term Paper Writing

1. Write down a paragraph to introduce your research topic. Include:
   - Brief description of your research question
   - Why you chose to study this topic

2. Write a paragraph describing your sample construction process. Include:

- Data cleaning steps
- Sample selection criteria
- Final sample size and time period

3. Create a table to display descriptive statistics of your sample and provide a brief explanation. Include:
   - Present means and standard deviations for outcome and control variables
   - Can be compared between treatment and control groups

4. Briefly describe your empirical methodology for estimating causal relationships. Include:
   - Description of your identification strategy
   - Mathematical expressions of your empirical model
   - Key assumptions for this method

5. If you have preliminary results, create a table or figure to present your findings and write a paragraph explaining your current findings. Include:
   - Interpretation of main estimates

6. If you don't have preliminary results yet, describe how you used generative AI (e.g., ChatGPT, Claude) to assist your research during the term paper process. Include:
   - What kind of tasks generative AI helped you complete, please provide an example
   - Write down the prompts you gave to the AI