RECSM: Quantitative Methods in Social Research

Day 1 - 02 07 2025

Burak Sonmez

Please skim through all of the instructions before getting started.

Note: This document also walks you through installing R and RStudio.

1. Install R

- Go to https://cran.r-project.org/
- Select "Download R for Windows"
- Click "install R for the first time"
- Click "Download R 4.0.3 for Windows"
- Select "Download R for (Mac) OS X"
- Click "R-4.0.3.pkg"

2. Install RStudio

- Go to https://www.rstudio.com/products/rstudio/download/
- Select "RStudio 1.3.1093 - Windows 7+ (64-bit)" for Windows users
- Select "RStudio 1.3.1093 - macOS 10.13+ (64-bit)" for Mac users

3. Problem set 1

Step 1: What is the result of summing all numbers from 1 to 250?

Step 2: Calculate the square root of 1500 and round it to the nearest possibility.

Step 3: For example, suppose the yearly number of MSc students at UPF during the period 2010 to 2020 is 75, 152, 230, 141, 292, 79, 211, 137, 156, and 111, respectively. What are the mean, the variance, and the standard deviation?

$$Hint: \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

4. Problem set 2

Step 1: First download the folder (RECSM workshop) on your computer

Step 2: Create a new file called "day1.R" in your RECSM workshop folder and write all the solutions in it.

Step 3: Clear the workspace and set the working directory to your RECSM workshop folder.

Step 4: Load the Asylum dataset into your R environment and inspect the name of variables. The asylum data is about the perception of asylum seeker size in the UK.

Table 1: The Codebook

Variable	Description
asylum	How many do you think are asylum seekers from Syria out of every 100 asylees in Britain?
sex	1 = male, 2 = female
age	Age of respondent
news	Do you normally read any daily morning newspaper at least three times a week?
religious	Do you regard yourself as belonging to any particular religion?
urban	Population density, 4 categories (highest density is 4, lowest is 1)
hhincome	Income bands for household, high number = high HH income
partyid	1 = Tories, 2 = Labour, 3 = SNP, 4 = Greens, 5 = Ukip, 6 = BNP, 7 = other

Step 5: What is the level of measurement for each variable in the Asylum dataset?

Step 6: Calculate the correct measure of central tendency for asylum, hhincome, news.

Step 7: Calculate the correct measure of dispersion for asylum, hhincome, news.

Step 8: How many respondents identify with the Labour?

Step 9: Calculate the variance and standard deviation of asylum for each party identification.

Step 10: Find the party identification of the oldest and youngest respondents.

Step 11: Find the 20th, 40th, 60th and 80th percentiles of age.

Step 12: Create a box plot for asylum grouped by the religious variable to show the difference between asylum for people who believe in any religion and people who do not.

Step 13:What is the mean of asylum for men and for women?

5. Problem set 3

Step 1: Using the Asylum dataset, please try to answer to what extent political differences explain perceptional biases regarding Syrian asylum seekers.

Step 2: Report summary statistics of the variables (asylum, age, news, sex). Make sure that you define the variables of news and sex as a factor.

Step 3: Explore the relationship between sex and asylum perception through a boxplot, using `ggplot`. You can customise your ggplot boxplot by adding more layers (e.g. labels, title, aesthetic specifications). Please state your interprations/observations from plotting.

Step 4: Please conduct hypothesis testing to confirm whether the mean differences in asylum perception are statistically significant between male and female respondents. Hint: (T-test)

- **Null Hypothesis**: The average perception of Syrian asylum seeker size is the same between male and female respondents
- **Alternative Hypothesis**: The average perception of Syrian asylum seeker size differs between different sex categories.

Solutions for problem sets