ECOM20001: Econometrics 1

Tutorial 2: Visualising and Describing Data in R, Summation Practice Problems

---

Overview

This tutorial consists of two parts. In part 1, you study visualising data with R. Part 2 provides practice problems with the summation operator.

Part 1: Visualising and Describing Data in R

A. Getting Started

Please create a Tutorial2 folder on your computer, and then go to the LMS site for ECOM 20001 and download the following files into the Tutorial2 folder:

- tute2.R
- tute2_crime.csv

The first file is the R code for tutorial 2, the second file is the .csv file that contains the dataset for the tutorial.[1] The dataset has the following 5 variables:

- stateid: identifier for a US state
- vio: violent crime rate — incidents per 100,000 people
- rob: robbery rate — incidents per 100,000 people
- dens: population per square mile of land
- avginc: real per capita personal income in the state

With the R file and data downloaded into your Tutorial2 folder, you are ready to proceed with the tutorial. Please go to the tute2.R file to continue with the next 2 sections of the tutorial, which are Visualising Data and Descriptive Statistics.

[1] The reference research article for these data is: Donohue, J., Ayres, I. (2003): "Shooting Down the 'More Guns Less Crime' Hypothesis," *Stanford Law Review*, 55, pp. 1193-1312.

B. Questions

Having worked through the tute2.R code and graphs, please answer the following questions.

1. Discuss the sample means, standard deviations, min and max for each of the four main variables in the dataset: vio, rob, density, avginc.
    - What does a "typical" state look like in the dataset? Focus on sample means in describing a typical state. Be sure to state the units of a variable to accurately describe what a typical state looks like.
    - Discuss how the min and max of each variable, highlighting the range of values that each variable takes on. Just how varied is the degree of violent crimes and robbery rates, and population densities and per capita incomes in the sample? How violent and robbery-filled is the worst state compared to the best state?
2. How do the respective probability densities of vio, rob, density, avginc look?
    - Focus on their mean, standard deviation, and skewness
3. Comment on the 3 scatter plots listed below; these should be saved in your working directory for tutorial 2. Visually, does a relationship appear exist in each graph? If so, offer an **economic explanation**[2] for why the relationship might exist. There may be multiple explanations, so you may offer various explanations if you wish. But just one explanation is fine.
    - Robbery vs Violence: fig_nice_scatter_rob_vio.pdf
    - Robbery vs Per Capita Income: fig_nice_scatter_avginc_vio.pdf
    - Robbery vs People per Square Mile: fig_nice_scatter_dens_rob.pdf

[2] **Economic explanations focus on the costs and benefits of a particular behaviour for explaining empirical patterns.** Here's an example economic explanation for a positive relationship between robbery rate and person per square mile: a robber might find it easier (or less costly) to rob people in a dense city than in a sparse regional town because there is less distance between the robber and people (e.g., the robber's targets) in cities. This could be an economic explanation for a positive relationship between the robbery rate and people per square mile (or urban density). Other possible economic explanations to consider: what if people are richer in urban places (affects robbers benefit from robbing people)? What is police are less effective in identifying robbers in dense places than in regional places (affects the robbers expected cost of robbing people)? Are there other economic explanations you could think of that affect robbery cost and benefits? Similarly for violent crime offender costs and benefits?

Part 2: Summation Practice Problems

1. Show the following equality is true:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

2. Show the following equality is true:

$$n\bar{x} = \sum_{i=1}^{n} x_i$$

3. Show the following equality is true:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

4. Show the following equality is true:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$