Geneva Graduate Institute (IHEID)
Econometrics I (EI035), Fall 2024
Marko Mlikota

Problem Set 1

Due: Sunday, 29 September, 23:59

- Prepare concise answers.
- State clearly any additional assumptions, if needed.
- Submit your solutions, along with any code (if applicable), in a **single pdf file** through **Moodle**. If you choose to write your solutions by hand, please make sure your scanned answers are legible.
- Grading scale:

5.5	default grade
6	absolutely no mistakes and particularly appealing write-up (clear and concise answers, decent formatting, etc.)
5	more than a few mistakes, or single mistake and particularly long, wordy answers
4	numerous mistakes, or clear lack of effort (e.g. parts not solved or not really attempted)
1	no submission by due date

Problem 1

Suppose you have data on the height of n female adults living in Switzerland – $\{x_i\}_{i=1}^n$ – whereby the observations in your sample are independent. Based on that, you want to estimate the average height of female adults in the whole population (i.e. the whole of Switzerland). Let this parameter of interest be denoted by θ. You can write your observations as

$$x_i = \theta + u_i \ , \quad \text{with} \quad \mathbb{E}[u_i|\theta] = 0 \ ,$$

i.e. the height of an individual i, x_i, is given by the true average height θ plus some noise u_i around it. Note that this is just another way of writing $\mathbb{E}[x_i|\theta] = \theta$.

(a) Find a point estimator for θ using the Least Squares (LS) estimation method, $\hat{\theta}$.

(b) What is the mean of $\hat{\theta}$? Is your $\hat{\theta}$ unbiased? Besides assuming $\mathbb{E}[x_i|\theta] = \theta$, is there any other assumption on the pdf of $x_i|\theta$ involved in finding this quantity? Are any assumptions regarding your sample $\{x_i\}_{i=1}^n$ involved?

(c) What is the variance of $\hat{\theta}$? Besides assuming $\mathbb{E}[x_i|\theta] = \theta$, is there any other assumption on the pdf of $x_i|\theta$ involved in finding this quantity? Are any assumptions regarding your sample $\{x_i\}_{i=1}^n$ involved?

(d) Suppose $u_i \sim U(-5, 5)$, i.e. u_i is distributed Uniformly between -5 and 5. Using a statistical software of your choice, write a program that, given a choice of n and θ simulates a dataset $\{x_i\}_{i=1}^n$. Fix $\theta = 175$ and $n = 10$. Compute $\hat{\theta}$ using this simulated data. Is your estimate close to the true value of $\theta = 175$? What happens under a dataset with $n = 100$ observations? What happens if you take $n = 1000$?

(e) Now let's use the program you wrote to analyze the behavior of $\hat{\theta}$ in repeated sampling.

  (a) simulate $M = 100$ different datasets of size $n = 10$: $\{\{x_i^m\}_{i=1}^n\}_{m=1}^M$

  (b) for each dataset $\{x_i^m\}_{i=1}^n$, compue the LS-point estimator $\hat{\theta}^m$

  (c) plot a histogram of $\{\hat{\theta}^m\}_{m=1}^M$

  Comment on the histogram (distribution) of $\hat{\theta}$-values. Is it in line with your expectations, based on the calculations you did for the questions above?

(f) Redo the previous exercise using $n = 100$ as well as $n = 1000$. How does the histogram (distribution) of $\hat{\theta}$ change? Relate this to the theoretical analysis we conducted in class.