Geneva Graduate Institute (IHEID)

Econometrics I (EI035), Fall 2024

Marko Mlikota

Problem Set 2

Due: Sunday, 13 October, 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 $\mathbb{E}[x_i|\theta] = \theta$. In addition, you assume that this noise u_i is Normally distributed with some known variance σ^2: $u_i \sim N(0, \sigma^2)$. Note that this – combined with the equation for x_i above – is just another way of writing $x_i \sim N(\theta, \sigma^2)$.

(a) Define and derive the Maximum Likelihood (ML) estimator of θ, $\hat{\theta}_{ML}$.
Hint: You first need to derive the likelihood, i.e. the distribution of your data $\{X_i\}_{i=1}^n$ conditional on θ, $p(x|\theta)$, based on the distribution of a single observation X_i, $p(x_i|\theta)$.

(b) Set up a Likelihood Ratio (LR) test of size $\alpha = 0.05$ for testing $\mathscr{H}_0 : \theta = \theta_0$ against $\mathscr{H}_1 : \theta \neq \theta_0$, i.e. determine the test-statistic $T(X)$ and the corresponding critical value c_α.
Hint: Note that if $Y \sim N(\mu, v)$, then $(Y - \mu)/\sqrt{v} \sim N(0, 1)$ and $(Y - \mu)^2/v \sim \chi^2_1$.

(c) Suppose $\sigma^2 = 6$ and you observe $n = 4$ observations, $x_1 = 178$, $x_2 = 161$, $x_3 = 168$ and $x_4 = 172$. Based on this data, can you reject $\mathscr{H}_0 : \theta = \theta_0 = 175$ (i.e. that the average height of female adults in Switzerland is 175cm)?

(d) Now let's suppose you could only find the test-statistic for the LR test, $T(X)$, but not the critical value c_α. Do so numerically, i.e.

(1) For $m = 1 : M$, with $M = 1000$,

- draw a sample $\{x_i^m\}_{i=1}^n \sim N(\theta_0, \sigma^2)$, setting $\theta_0 = 175$, $\sigma^2 = 6$ and $n = 4$,
- compute $T(x^m)$.

Plot a histogram of $\{T(x^m)\}_{m=1}^M$. This is your numerical approximation of the distribution of $T(X)$ under $\mathscr{H}_0$.

(2) Sort your draws $\{T(x^m)\}_{m=1}^M$ from lowest to largest and take the $M(1-\alpha)$th draw. This is your numerical approximation of c_α, the $100(1 - \alpha)$th quantile of the distribution of $T(X)$.

Is the value you get close to the true, analytically obtained c_α? What do you expect to happen if you take a larger value for M? Does your conclusion from the previous exercise change if you set up your test numerically as opposed to analytically?

(e) Based on your LR-test, find a (general) expression for the 95% confidence interval for θ_0, $C(X)$. How does that interval look like if you apply it to your particular data? Is $\theta_0 = 175$ in that interval? Explain why it should (not) be.

(f) Let's again suppose you were not able to analytically set up the LR test and, based on it, find $C(X)$. Find the confidence interval $C(x)$ for your sample numerically as follows. First, fix a grid $\mathcal{T}$ of values for θ_0, $\mathcal{T} = 160 : 0.1 : 180$, and create a vector vc of the same dimension as $\mathcal{T}$. Then, for each $\theta_0 \in \mathcal{T}$,

(1) repeat the numerical procedure from above to find $c_\alpha(\theta_0)$, the (numerical approximation of the) critical value for a size $\alpha = 0.05$ test for testing $\mathcal{H}_0 : \theta = \theta_0$.

(2) compute the LR-test-statistic $T(x; \theta_0)$ for your sample x. If $T(x; \theta_0) < c_\alpha(\theta_0)$, then $\theta_0 \in C(x)$ and you record a 1 in the corresponding entry in vc, otherwise $\theta_0 \notin C(x)$ and you record a 0.

Illustrate your $C(x)$ using a scatter plot: put $\mathcal{T}$ on the x-axis and, for each value $\theta \in \mathcal{T}$, have a one on the y-axis if θ is in $C(x)$ and a zero otherwise. How does your $C(x)$ compare to the one obtained analytically?