Econometrics A: Problemset 1

LAURA MAYORAL
Institute for Economic Analysis and Barcelona GSE

January 2021

Deadline: January 28rd before 15:00 (Paris time). Please submit your answers electronically –scanned or typed, as you prefer–, to your TA, sanghyun.park@insead.edu.

Note: Problems 1 and 2 are designed for you to review basic concepts that we will use in the next classes (large sample theory, hypothesis testing, p-values, etc).

Exercises

1. Read carefully Handout 0 (Large sample theory) and answer the following questions.

The production process of a good is considered to work satisfactorily if less than 1% of the produced units are defective. To determine whether this is the case, a team has gathered data on the quality of the products. More specifically, 10,000 products were examined and it was found that 106 of them were defective. Define the random variables X_i=1 if product i is defective, and 0 otherwise, for product $i = 1 \dots N$, and assume that these variables are i.i.d.

(1) Provide an estimator of the probability that one product is defective. (Hint: Notice X_i is a Bernoulli distribution and that in these distributions $E(X_i) = p(X_i = 1)$. Then, use the method of moments (i.e., replace population moments by sample moments to obtain an estimator of this quantity.)
(2) Using the collected data provide an estimate of that probability.
(3) Define an estimator of the variance of the random variable X_i. (Hint: remember that X_i is a Bernouilli random variable and therefore its variance is $p(1-p)$ where $p = P(X = 1)$). Using your estimate of p, provide an estimate for the variance of X_i.
(4) Use the Law of Large Numbers to describe the limit of the estimator provided in (1). Is this estimator consistent?
(5) Use the Central Limit theorem to describe the asymptotic distribution of the estimator in (1)

2. Read carefully Handout 0 (hypothesis testing section). Using the data in the previous exercise construct an (asymptotic) test of hypotheses at the 5% level (i.e., $\alpha = .05$) to determine whether the production process works well or not (Remember that it works well if $p < 0.1$).

While doing this, answer the following questions:

(1) Carefully define the meaning of α (type I error); also define the meaning of the type II error.
(2) Describe the null and the alternative hypotheses

(3) Describe the test statistic that you can use to test those hypotheses
(4) describe the critical region (i.e., the values of the test-statistic for which you reject the null hypothesis)
(5) compute the value of the test using the data the problem.
(6) Can you reject H_0? Clearly justify your answer.
(7) Define the concept of p-value.
(8) Compute the p-value associated to the test-statistic you computed in c)

3. Let y and z be random scalars, and let $\mathbf{x}$ be a 1×k random vector, such that $\mathbf{x}_1 = 1$. Consider the population model:

$$E(y|\mathbf{x}, z) = \mathbf{x}\gamma + \delta z$$

$$Var(y|\mathbf{x}, z) = \sigma^2$$

i) Write a probabilistic model of y as a function of the conditional expectation specified above and a random disturbance u.

ii) Under the assumptions above, compute the conditional mean and the conditional variance of u. Is u conditionally homoscedastic? (that is, is the (conditional) variance unrelated the values of $\mathbf{x}$ or z). Justify your answer.

iii) Compute the unconditional mean and the unconditional variance of u. (Hint: use the law of iterated expectations).

iv) The main assumption that we need for identification of γ and δ is that $E((\mathbf{x}\ z)'u) = 0$. Under the assumptions above, is this condition met? clearly justify your answer.

v) Assume now that the variable z cannot be observed, and that you are considering this model instead.

$$y = \mathbf{x}\gamma + u^*$$

where now the error u^* contains the old error and the omitted variable, i.e.: $u^* = u + \delta z$. Assume further than $\mathbf{x}$ and z are uncorrelated. Is it still possible to identify γ? Clearly justify your answer.

vi) Will your answer in v) change if $\mathbf{x}$ and z were correlated? Justify your answer.

Computer Practise

4. Binscatter is a useful STATA command for data visualization, that provides a non-parametric estimation of the conditional expectation. Read this document to understand what it does.

i) install the binscatter command in STATA.

ii) Type "help binscatter" to know more about the options that this command provides.

iii) Load the dataset "nlsw88.dta" in the stata memory (hint: type "help sysuse" to learn how you can do this.)

iv) Plot a scatter plot relating wages and ttl_exp (total work experience). Add a line to the plot that describes the best linear fit between these variables. Describe the plot. (Hint: you can check this page for help https://stats.idre.ucla.edu/stata/modules/graph8/intro/introduction-to-graphs-in-stata/

v) Now plot a binscatter plot with the same variables. Compare the two graphs.

v) Change the default number of bins in your binscatter plot to 40

vi) Produce a binscatter that connects the different bins (hint: use the linetype option)

vii) Compute two different binscatters (drawn in the same figure) relating the variables mentioned above, one for the people that are married and another for single people.