Introduction à l'économétrie sur STATA, Univ.Bdx (2017-18)

Pierre Levasseur

Université de Bordeaux

M1

Introductory Econometrics

Économie du Développement & Intelligence Économique

2017-18

Problem set 1

Statistics (Rappel); Simple Regression Model

A. STATISTICS

Exercise 1 (STATA)*

a) Suppose you have a population characterized by the normal distribution with mean 500 and standard deviation 75.
    1. Keep a sample of size 2000 and compute the Interval Confidence at 95% confidence level around the population mean.
    2. Keep a sample of size 2000 and compute the Interval Confidence at 95% confidence level. Suppose you do not know the population variance.
b) Keep a new random sample, but of size 100 instead of 2000. Compute the CI at the 95% confidence level. Do you expect a larger or smaller CI?
c) Compute the CI now at the 99% confidence level. Do you expect a larger or smaller CI?

Exercise 2 (STATA)*

A manufacturer of 3D printing machine considers that the manufacturing process is working properly if the mean weight of the machines is 8.6 Kg. The population standard deviation of these printing machines is unknown. Suppose a random sample of size n = 36 yields an average weight of 8.7 kilograms and a standard deviation of about 0.3 kilograms. Should the

manufacturer conclude the process is working properly or improperly? Use the .05 level of significance.

Exercise 3 (STATA)

We want to test the efficacy of a medical treatment. Suppose that we have 200 patients (N=200) which are randomly assigned to two groups: the first one receives the *treatment* (N_1=100) and the other receives the *placebo* (N_2=100). After 8 weeks, 19 of the placebo treated patients showed improvements, whereas 27 of those treated improved. Is there any reason to believe that the treatment is necessary? Use an alpha =0.05

Exercise 4 (STATA)*

Use data in Dossier1_Wage_female.dta. The data (randomly chosen from a population of French workers) contain 526 individuals. Let *wage* denote the hourly salary in Euro; *female* is a dummy variable equals to one when the observation corresponds to a female worker and zero in case of man.

Test whether you observe a significant difference in terms of wage between man and women. Assume equal variances in the two groups.

B. SIMPLE REGRESSION MODEL

Exercise 5

Let *kids* denote the number of children ever born to a woman, and let *educ* denote years of education for the woman. A simple model relating fertility to years of education is:

$$kids = \beta_0 + \beta_1 educ + u$$

a) What kinds of factors are contained in u? Are these likely to be correlated with level of education?
b) Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain.

Exercise 6

In the simple linear regression model

$$y = \beta_0 + \beta_1 x + u$$

Suppose that $E(u) \neq 0$. Letting $\alpha_0 = E(u)$, show that the model can always be rewritten with the same slope, but a new intercept and error, where the new error has a zero expected value.

Exercise 7

Consider the simple linear regression model:

$$y = \beta_0 + \beta_1 x + u$$

Show that $var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Exercise 8

a) Show that the sample average of the OLS residuals is zero
b) Show that the sample covariance between the regressors and the OLS residuals is zero
c) Show that the sample cross product between the fitted values and the residuals is zero
d) Show that the total sum of squares (TSS) is the sum of the explained sum of squares (ESS) and the residual sum of squares (RSS), that is:

$$\sum [y_i - \bar{y}]^2 = \sum (\hat{y}_i - \bar{\hat{y}})^2 + \sum \hat{u}_i^2$$

Exercise 9*

How does the correlation coefficient differ from regression slope?

Exercise 10 (STATA)*

The following table (Dossier1_GPA) contains the ACT scores and the GPA (grade point average) for 8 college students. Grade point average is based on a four-point scale and has been rounded to one digit after the decimal.

Student	GPA	ACT
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

a) Estimate the relationship between GPA and ACT using OLS; that is, obtain the intercept and slope estimates in the equation

$$\widehat{GPA} = \hat{\beta}_0 + \hat{\beta}_1 ACT$$

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much higher is the GPA predicted to be, if the ACT score is increased by 5 points?

b) Compute the fitted values and residuals for each observation and verify that the residuals (approximately) sum to zero.

c) What is the predicted value of GPA when ACT = 20?

d) How much of the variation in GPA for these 8 students is explained by ACT? Explain.

Exercise 11

Consider the savings function:

$$sav = \beta_0 + \beta_1 inc + u$$

$$u = \sqrt{inc} \cdot e$$

where *e* is a random variable with $E(e) = 0\ var(e) = \sigma_e^2$. Assume that *e* is independent of *inc*.

a) Show that $E(u|inc) = 0$, so that the key zero conditional mean assumption is satisfied. [*Hint:* If *e* is independent of *inc*, then $E(e|inc) = E(e)$.]

b) Show that $var(u|x) = \sigma_e^2 inc$, so that the homoskedasticity Assumption is violated. In particular, the variance of *sav* increases with *inc*. [*Hint:* $var(e|inc) = var(e)$, if *e* and *inc* are independent.]

c) Provide a discussion that supports the assumption that the variance of savings increases with family income.

Exercise 12 (STATA)*

Use the data stored in Dossier1_IQ.dta

a) Estimate the relationship between IQ and education by supposing a linear relationship:

b) Which is the average increase of IQ in case of an 1 and 2 years increase in education?

c) Compute the predicted valued of IQ and draw a graph with the IQ estimated.

d) Compute the residual sum of squares (SSR)

e) Compute the Explained Sum of Squares (SSE)

f) Compute R squares = SSE/SST = 1- SSR/SST

g) Show the Algebraic properties of OLS

h) Compute the IQ for an individual with 10 years of education

i) Compute the standard error for $\hat{\beta}_1$

j) Draw a graph with IQ values and the estimated model.

Exercise 13

Use again the data stored in Dossier1_IQ.dta

a) Estimate the (linear) relationship between monthly salary (wage) and IQ score (IQ)

b) Which is the estimated increase in wage in case of additional 15 points in IQ?

c) Suppose that you are interested in to know the effect of 1 additional point of IQ on the percentage increase on wage. Which model you suggest?

d) Which is the percentage increase of wage associated with an increase of 15 points of IQ?