

Econometrics A: Problemset 1

LAURA MAYORAL

Institute for Economic Analysis and Barcelona GSE

January-February 2019

Deadline: January 23rd. Please submit your answers in paper or electronically to your TA, Arzi Adbi, arzi.adbi@insead.edu.

1. A researcher wants to assess the impact of alcohol consumption during pregnancy on newborns' weight. To that effect, she employs survey data where women declare their weekly alcohol intake. The weight of the babies is also recorded.

- (1) What do you think about this procedure? do you think that selection can be an issue in this case? If your answer is affirmative, do you think that it will lead to an overestimation or to an underestimation of the effect of alcohol consumption on newborns' weight? justify your answer.
- (2) Alternatively, the researcher is planning to run an experiment where pregnant women are randomly assigned to the "treatment" of interest. Describe how this have to be done. Do you think an ethics committee would approve of such an experiment?

Note: You can read [here](#) a newspaper article written by an economist on the benefits of knowing econometrics when it comes to interpreting medical advice.

2. Properties of expectations, variances and covariances

Let X_1 , X_2 and X_3 be three random variables such that $E(X_1) = 2$; $E(X_2) = 3$; $E(X_3) = 0$, $Var(X_1) = 2$; $Var(X_2) = 4$; $Var(X_3) = 4$; $cov(X_1, X_2) = -1$; $cov(X_2, X_3) = 2$; $cov(X_1, X_3) = 0.4$; $E(X_2|X_1) = 2$.

i) Compute:

- (1) $E(3X_1 + 0.5X_2 + 4X_3 + 7)$
- (2) $Var(X_1 + X_2 - X_3)$
- (3) $Corr(3X_1 + 2, 1/6X_3 + 2)$
- (4) the best predictor for $Z = 3X_2 + 5$, given information on X_1 .

3. A team of researchers wants to analyse the relationship between gender and wage in the car manufacturing industry in country C.

- (1) What is the population of interest?
- (2) Briefly describe how you could gather an i.i.d. sample on this population.

- (3) The following table describes the joint distribution associated to gender and wage in that industry. More specifically, X is a dummy variable defined as X : gender (1 if women, 0 if man) and Y (salary) is a discrete variable that takes three values: 0 if low, 1 if medium, 2 if high). Compute: i) the marginal distributions of X and Y ; ii) the conditional distributions of Y given $X = 0$ and $X = 1$; iii) $E(X)$.

X, Y	0	1	2
0	0.3	0.2	0.1
1	0.2	0.1	0.1

- (4) You pick a person at random from this population and turns out to be a woman. What is the probability that she gets a medium or a high salary?
- (5) Compute the best predictor for Y if $X=1$.
- (6) are X and Y independent? justify your answer

4. Computer practise.

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. Type “help binscatter” to know more about the options that this command provides.
- ii) Load the dataset “nlsw88.dta” in the stata memory (hint: type “help sysuse” to learn how you can do this.)
- iii) Plot a scatter plot relating wages and ttl_exp (total work experience). Describe the plot.
- iv) Now plot a binscatter plot with the same variables. Compare the two graphs.
- v) Change the default number of bins to 40
- vi) Produce a binscatter that connects the different bins (hint: use the linetype option)