B.E. International Program

Faculty of Economics, Thammasat University

EE325 INTRODUCTORY ECONOMETRICS

SEMESTER 2/2011

PROBLEM SETS 1&2

Instructions

1) The Problem Sets 1&2 count towards your final score accumulation of EE325: Introductory Econometrics. ATTEMPT ALL QUESTIONS.
2) To complete the assignment, you are allowed to discuss the problems with your classmates but have to write up solutions completely on your own.

   Copying is plagiarism and will be treated as an "honor code violation".
3) The date of submission is on Tuesday 21st February 2012 by 11.15 hrs. (in class)
4) If you have questions, please send me an e-mail at pwrasai@gmail.com

---

1. Shown are the probability distributions of job satisfaction scores for a sample of information systems (IS) senior executives and IS middle managers (Computerworld, May 26, 1997). The scores range from a low of 1 (very dissatisfied) to a high of 5 (very satisfied).

Job Satisfaction Score	Probability	
	IS Senior Executives	IS Middle Executives
1	0.05	0.04
2	0.09	0.10
3	0.03	0.12
4	0.42	0.46
5	0.41	0.28
Total	1.00	1.00

   a) What are the expected values of the job satisfaction scores for senior and middle executives?
   b) Compute the variances and standard deviations for both groups
   c) Compare the overall job satisfaction of both groups (Note: take into accounts the s.d.)

2. Referring to the Table 1 in the excel file,
   a) What are random variables do we have? What are their types?
   b) Suppose Nominal GDP (NGDP) and Nominal Consumption (NCONS) is one group, and real GDP (RGDP) and Real Consumption (NCONS) is another group. In each group which variable should be independent variable, and which should be dependent variable? Why?
   c) Plot scatter diagrams for both groups (Note: an independent variable should be on the horizontal axis). What can you say about the relationships? Are they consistent with your expectation?
   d) Find the expected values, variances for all variables. Find covariance and correlation for both groups.

3. "We usually observe the high correlation between the number of police officers and the crime rate across cities in a country"
   a) What does the statement above mean in a plain word?
   b) What does it mean by "high correlation does not imply causation"?

4.
   a) What is the difference between discrete and continuous probability density function?
   b) In a throw of two dice, the random variable X, the sum of numbers shown on two dice. Can you list all possible outcomes of the random variables, and the PDF associated with each outcome?
   c) Is Poisson distribution a discrete or continuous PDF? Why?

5. Suppose that a machine is used for a particular task in the morning and different task in the afternoon. Let X and Y represent the number of times the machine breaks down in the morning and in the afternoon, respectively. Table below gives the joint probability distribution of (X,Y).

Y / X	0	1	2
0	0.1	0.2	0.18
1	0.04	0.09	0.08
2	0.06	0.11	0.14

   a) Find the marginal probability distribution of X
   b) Are X and Y independent? Show your proof.
   c) Find E(X+Y). What does it mean?
   (Note: Cov(X,Y) = E(XY)-E(X)E(Y); $E(XY) = \sum_i \sum_j X_i Y_j p(X_i, Y_j)$)
   d) What is the probability that the machine will be broken more than one time in the morning AND less than one time in the afternoon?

6. Suppose we have population regression function (PRF): $Y_i = \beta_1 + \beta_2 X_i + u_i$, and sample regression function (SRF): $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$,
   a) What are PRF and SRF? How are they different?
   b) Can we find β_1 and β_2 from the sample? Why?
   c) If we have only sample, what can we find?
   d) Is it necessary that $\hat{\beta}_1$ and $\hat{\beta}_2$ are always equal to β_1 and β_2 ? Why?
   e) What are properties of $\hat{\beta}_1$ and $\hat{\beta}_2$ that make them a good estimators of β_1 and β_2 ?
   f) What are the names for u_i and $\hat{u}_i$? Are they the same? Why?

7. Are the following models linear regression models? Why? (Note: An intrinsically linear model is counted as a linear model. If you conclude that a model is intrinsically linear, show your manipulation.)
   a) $Y_i = e^{\beta_1 + \beta_2 X_i + u_i}$
   b) $Y_i = \dfrac{1}{1 + e^{\beta_1 + \beta_2 X_i + u_i}}$
   c) $\ln Y_i = \ln \beta_1 + \beta_2 \ln\left(\dfrac{1}{X_i}\right) + u_i$
   d) $\ln Y_i = \beta_1 + (0.75 - \beta_1) e^{-\beta_2 (X_i - 2)} + u_i$
   e) $Y_i = \beta_1 + \beta_2^3 X_i + u_i$

8. Referring to the Table 2 (in the excel file),
   a) Plot Y against X
   b) Suppose the estimated equation is $\hat{Y} = 28.9015 + 5.3836X$, compute $E(Y)$ and $E(Y \mid X = 100)$.

9. Table 3 (in the excel file) gives data on mean SAT scores for college-bound seniors for 1972-2007. These data represent the critical reading and mathematics test scores for both male and female students.
   a) Use the horizontal axis for years and the vertical axis for SAT scores to plot the critical reading and math scores for males and females separately.
   b) What general conclusions do you draw from these graphs?
   c) Knowing the critical reading scores of males and females, how would you go about predicting their math scores?
   d) Plot the female math scores against the male math scores. What do you observe?