Intro to Econometric Theory
Heinz School, Carnegie Mellon University
90-906, Spring 2004-5

Homework #1, due Friday, February 11, 2005

1. Consider a sample of people who were asked to report their sex (Everyone answers, and everyone answers either male or female). From their answers, we construct the following matrix:

$$X = \left[\begin{array}{ccc} \mathbf{1} & M & F \end{array} \right] \tag{1}$$

   In the matrix, each row corresponds to one person's response; **1** is a column of 1s, M is a column with a 1 in the ith row if the ith person answered male and a zero otherwise, and F is a column with a 1 in the ith row if person i answered female and a zero otherwise.

   Is $X'X$ invertible. Why or why not?

2. We discussed in class that variance matrixes are always positive semi-definite but not always positive definite. In the case of scalar random variables, a singular variance matrix means a zero variance — that is, that the random variable is not really random at all, but fixed.

   Now, let's think about what a positive semi-definite but not positive definite (hereafter PSDNPD) variance matrix means for the random vector it corresponds to. Let x be a random vector with PSDNPD variance matrix V.

   (a) Must there be a fixed non-zero vector α, such that $\alpha' V \alpha$=0? Why or why not?
   (b) Suppose there is such an α, what would the variance of $\alpha' x$ be?
   (c) What does this tell us about $\alpha' x$?
   (d) Does this mean that $\alpha' x = 0$?
   (e) So, what does it mean for a random vector to have a PSDNPD variance matrix?

3. Suppose we have an estimator of per-capita income in the US, $\hat{I}$, and an estimator of the share of income saved (NOT the average over people of share of income saved but rather the average over dollars of share of income saved!), $\hat{s}$. The true values of these parameters are μ_I and μ_s.

These two estimators are derived somehow from a survey of size n, and we have proved the following result for these estimators:

$$\sqrt{n}\left(\left(\begin{array}{c}\hat{I}\\ \hat{s}\end{array}\right)-\left(\begin{array}{c}\mu_I\\ \mu_s\end{array}\right)\right)\sim N\left(\left(\begin{array}{c}0\\ 0\end{array}\right),\left[\begin{array}{cc}400 & 20\\ 20 & 9\end{array}\right]\right) \tag{2}$$

Please construct an approximate (why approximate?) 95% confidence interval for the per-capita savings, $\mu_I \mu_s$.

4. You are interested in the effect of going to the emergency room on a person's life expectancy. There is a sample of people collected on a particular day, some of whom went to an emergency room and some of whom did not. They are followed over time until they die. From each person there are two items in the dataset: T_i the length of time from the day of the survey until death and ER_i a variable which equals one if they went to the emergency room and zero if they did not. You wish to run a regression, but the people who conducted the study will not let you have the data, only summary statistics. What should you ask for and how will you use it?

5. Considering the example in problem 4, consider two linear models:

   (a) $T_i = \beta_1 + \beta_2 ER_i + \epsilon_i$

   (b) $ER_i = \beta_3 + \beta_4 T_i + \epsilon_i$

   Let's call the OLS estimates you would get from regressions run on these equations $\widehat{\beta}_{1,OLS}, \widehat{\beta}_{2,OLS}, \ldots$. What relationship will there be between $\widehat{\beta}_{2,OLS}, \widehat{\beta}_{4,OLS}$? What sign ($\pm$) do you expect for $\widehat{\beta}_{2,OLS}, \widehat{\beta}_{4,OLS}$?

6. We usually interpret regression models as embodying some causal story (the RHS causing the LHS); however, OLS does not "know" this: it produces estimates regardless of whether our implicit causal story is right. Discuss this fact in the context of the models in question 5 and your expectations about the signs of the relevant coefficients.