Economics 475: Econometrics
Homework #1

This homework is due on January 11th.

This homework has two purposes: one is to remind (or introduce to) you the rules of mathematic expectation. The other is to have you remember how to use Stata (or whatever software you choose).

If x is discrete, the expected value of x (denoted E[x]) is: $E[x] = \Sigma(x \times f(x))$ where x is the value of a random variable and f(x) is the probability of achieving that value of the random variable. Notice, that you have been doing this for a long time already. For instance, the expected value of a fair, 6 sided die is 3.5 (one-sixth of the time you get a one, one-sixth of the time you get a two, …).

The following definitions are a summary of the rules of mathematical expectation more fully explained at: http://www.cbe.wwu.edu/krieg/Econ475/Mathematical%20Expectation.pdf

1) The expectation of a random variable is the random variable's mean. In other words $E[X] = \mu_X$.
2) The expectation of a constant is that constant. In other words E[a] = a where a is a constant.
3) I write the variance of a random number X as V[X] or Var[X]. By definition $V[X] = \sigma_X^2 = E[(X - E[X])^2]$. Remember, the variance formula is $\frac{\sum(X - \overline{X})^2}{n}$ which is really taking the average of the squared deviations of X from its mean—exactly what $E[(X - E[X])^2]$ says!
4) Like rule #3, the covariance between two random variables X and Y are typically written as Cov[X,Y] = E[(X-E[X])(Y-E[Y])].
5) The expectations operator E is a linear operator, that is if Z = X + Y then E[Z] = E[X] + E[Y].
6) Likewise E[aX] = aE[X] if a is a constant and X is a random variable.
7) Finally, E[XY] where both X and Y are random variables is equal to E[X]×E[Y] only if X and Y are independent (that is Cov[X,Y] = 0).

1. Here are some paired observations of two random variables:

Observation	X	Y
1	5	3
2	6	5
3	9	12
4	1	-4
5	2	0

Find E[X], E[Y], V[X], V[Y], E[XY], and Cov[XY] and describe in words what each of these are.

2. One can show that when E[X] = 0 then the variance of X is equal to $E[X^2]$. Prove this (and don't forget it—we will see it over and over again)!

3. Use the facts that X is a random variable with mean 4 and variance 8, Y is a random variable independent of X with mean 3 and variance 5 to answer the following questions:
a. What is E[X] and E[Y]?
b. Think of a random variable $Z = 5+ 2X + Y$. What is E[Z] and V[Z]?

c. How would your answer to b change if there was a positive correlation between X and Y rather than X and Y being independent? Specifically, what happens to V[Z] as the correlation between X and Y becomes closer to 1? Why?

4. The data set "freshmen 2002 data" consists of observations of each freshmen (new students and running start students) who began college in 2002. Use this data set to answer the following questions.
a. Estimate the regression $gpa_i = B_0 + B_1 hsgpa_i + \varepsilon_i$. Does high school GPA impact first-quarter GPA? How do you interpret B_1?

b. I remember high school counselors telling me that the average college student earned a college GPA one point lower than their high school GPA. Using the regression in part a, test if this is true.

c. Estimate the regression $gpa_i = B_0 + B_1 hsgpa_i + B_2 satverb_i + B_3 satmath_i + B_4 male_i + B_5 firstgen_i + \varepsilon_i$. How do you interpret B_1? How does this differ from your answer in part a?

d. Does the SAT test help predict first-quarter GPA?

e. I notice the coefficient on male is negative. Offer some explanations for this. Do any of your explanations violate the classical assumptions of OLS?