Problem set 2

July 14th, 2011
due July 19th, 2011

Question 1

Consider changing the units of measurment for observations in a dataset. This is equivalent to multiplying each observation on a given variable by a constant scaling parameter. (Refer to chapter 2.4 for help if needed.)

- Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the intercept and slope coefficient from the regression of y_i on x_i. Let $\tilde{\beta}_0$ and $\tilde{\beta}_1$ be the intercept and slope coefficient from the regression of c_1y_i on c_2x_i, where c_1 and c_2 are non-zero constants. Express $\tilde{\beta}_0$ and $\tilde{\beta}_1$ as functions of $\hat{\beta}_0$, $\hat{\beta}_1$, c_1 and c_2.
- Let $\hat{\alpha}_0$ and $\hat{\alpha}_1$ be the intercept and slope coefficient from the regression of $ln(y_i)$ on $ln(x_i)$. Let $\tilde{\alpha}_0$ and $\tilde{\alpha}_1$ be the intercept and slope coefficient from the regression of $ln(c_1y_i)$ on $ln(c_2x_i)$, where c_1 and c_2 are non-zero constants. Express $\tilde{\alpha}_0$ and $\tilde{\alpha}_1$ as functions of $\hat{\alpha}_0$, $\hat{\alpha}_1$, c_1 and c_2.

Question 2

Consider the following econometric model relating the birth weight (in ounces) of babies in the US to the cigarettes consumption of the mother:

$$bweight_i = \beta_0 + \beta_1 cigs_i + \epsilon_i$$

To estimate the model, infant birth weight (*bweight*) is regressed on the average number of cigaretts the mother smoked per day during pregnancy (*cigs*). The following estimates are obtained:

$$\hat{\beta}_0 = 119.77$$
$$\hat{\beta}_1 = -0.514$$

- what is the interpretation of $\hat{\beta}_1$?
- what is the interpretation of $\hat{\beta}_0$? Does it makes sense in this example?
- if one mom smoked on average 10 cigarettes more per day than the other mom, what would you predict the babies weight difference to be?
- you found out about the babies from the previous point that one baby is 3 ounces heavier that the other. What does this imply for previous point?
- is the cetris paribus condition satisfied here? what does you answer imply for unbiasdness of applied estimator?

Question 3

Consider the following simple regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

The Zero Conditional Mean assumption holds, and it is known that the error term has a standard normal distribution. The sample size is 100 and the $x's$ have variance equal to 16 in the sample. What is the variance of the estimator $\hat{\beta}_1$?

Question 4

Answer this question using STATA and the data set WAGE1.dta (N=526). The variables of interest are *hourly wage* (w_i) and *education* (s_i). Attach the Do-File and the output log from your work using Stata.
Consider the following equation:

$$w_i = \beta_0 + \beta_1 s_i + \epsilon_i$$

- find the OLS estimates of β_0 and β_1
- state and verify the first two algebraic properties of the OLS estimator (consult the class notes)
- report the R^2 and comment on the value obtained
- now do the same for the following model:

$$ln(w_i) = \alpha_0 + \alpha_1 s_i + \epsilon_i$$

  compare the estimates of $\alpha's$ with the estimates of $\beta's$ and commnet on the difference.

Question 5

Sample variances of x are usually calculated by dividing the sum of x minus $\bar{x}$ squared by (n-1) which "corects for degrees of freedom". What does this correctness accomplish?