Homework #3
***** Due Tuesday, February 19, 2013 at the beginning of lecture *****

Please work on **L1** during lab so you can ask questions of the lab instructor.

Lab Problems

L1. The dataset `TeachingRatings.dta` contains data on course evaluations, course characteristics and professor characteristics for 463 at the University of Texas at Austin. This question examines whether the "beauty" of an instructor impacts course evaluations. Create a do file `teaching.do` and a log file `teaching.log` for this problem.

- **a.** Create a graph that includes both a scatter plot of average course evaluation (`course_eval`) vs. instructor's beauty (`beauty`) and the fitted line. Save this scatter plot as `scattereval_beauty.emf` so it can be printed and turned in with your assignment.
- **b.** Using the scatter plot and the fitted line from a), does it appear that course evaluations and beauty are related? Explain why.
- **c.** Use the `summarize` command to look at the description of the variables in this dataset.
- **d.** Use the covariance option for the correlate command to get the covariance between `course_eval` and `beauty`.
- **e.** Use the `regression` command to estimate the following linear regression model $Course\ Evaluations_i = \beta_0 + \beta_1 Beauty + u_i$.
- **f.** What is the interpretation of the intercept term β_0 in *this* regression model.
- **g.** Using the regression output, what is the value of the estimated intercept $\hat{\beta}_0$?
- **h.** What is the interpretation of the slope term β_1 in *this* regression model.
- **i.** Using the regression output, what is the value of the estimated slope $\hat{\beta}_1$? Is it significantly different from zero? How do you know?
- **j.** Suppose Professor Smith is of average beauty, while Professor Miller is one standard deviation above the average beauty. Using the regression output and the summary statistics, predict Professor Smith's and Professor Miller's course evaluations.
- **k.** Does this model explain a large fraction of the variance in course evaluations? Explain why or why not.

Questions

Q1. Suppose you are interested in estimating the effect of hours spent in an SAT preparation course (*hours*) on the total SAT score (*sat*). The population is all college-bound high school seniors graduating in a particular year.

a) Suppose you are given a grant to run a controlled experiment. Explain how you would structure the experiment in order to estimate the causal effect of *hours* on *sat*.

b) Consider the more realistic case where students choose the number of hours they spend on SAT preparation and whether or not to take an SAT-prep course. In this case, you can only randomly select *sat* and *hours* from the population of students. The population model is:

$$sat = \beta_0 + \beta_1 hours = u$$

where, because of the intercept, E(u) =0. List one factor that might explain *sat* that is likely included in *u*. Do you think this factor is correlated with *hours*? If this factor were correlated with *hours*, why might that be a problem?

c) In the equation from part (ii), what should the sign of β_1 be if the preparation course is effective?

Q2. Use the results from L1 to calculate the following. Even though the regression output tells you the answers, show how they are calculated assuming you had the other output but you did not have the regression results. Please show what formula you used and the steps you used to get to the final answer. You can always use the output to see if you did it right!

a) The coefficient for beauty, using the formula. (Hint: What is the numerator? What is the denominator?)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

b) Write down the null and alternative hypotheses to test if the coefficient on beauty is equal to zero.

c) Calculate the t-statistic to test the null hypothesis from part b). Note: In this part it is OK to use the standard error calculated by the regression output rather than having to calculate it.

d) To test the hypothesis in part b) with a confidence level of 95% level, then the critical t-value is 1.96. Would you reject or accept (fail to reject) the null hypothesis? Explain.

e) Calculate the R^2 for this model. What does it represent?

f) Compare your results to a), c) and e) to the output from your regression in L1. Are they the same?

- *Submit your written responses to the questions, your do file, log file, and your exported scatterplot. The responses to L1 should be in complete sentences. The answers to anything that asks you to explain should also be in complete sentences.*
- *Be sure to follow the "Problem Set Guidelines" distributed at the beginning of the semester.*