ECON 6306:501 – Applied Econometrics
Fall 2017 – Beron
Problem Set 1
Due Date is September 12

Instructions: This problem set must be **typed** and printed with writing only on **one side** of each sheet of paper. You may work together on these questions, *except* where the problem states differently, but all questions must be written up separately and evidence that this has not occurred will be subject to UTD's plagiarism policy. Throughout, *all* work leading to your answers must be shown for full credit.

You will turn in a **hard copy** that includes your answers and any Stata output that supports your answers (regression tables, correlation tables, etc.) making sure it is easily read, with Stata output not ragged, as discussed in class and Stata graphs formatted clearly. Your *error-free* Stata log that shows all the requested work (descriptions, generated variables, etc.) should be uploaded to the eLearning web site, making sure the code is separated by question. Report to the third significant digit. For each part, you should explain in words what has been done and found. Data files are on the web site, either in the problem set folder or in the Stata textbook data files folder. ***If*** a level of significance is not specifically asked for, then use a 0.05 level for your work. Where a statistical test is called for make sure you state both the null and alternative hypotheses in proper form and, at the end, state your conclusion in words with reference to your stated hypotheses.

1. Hill, Griffith, and Lim, problem 5.12, pp. 204 – 205, using significance levels as given below.
   Further instructions: NOTE that the data is called cocaine.dta and is in the course directory that has the Stata files for the textbook. You are to write the null and alternative hypotheses by hand for each part, being careful of what is a one-tail and what is a two-tail test. You are to use Stata to determine the critical *t*-values as necessary (make sure to show your code and output). You are to write out the formula for the calculated *t*-value in each case based on the specific problem, but you do not have to calculate it out by hand. An example on p. 118 of the textbook is the first equality in the *t* formula rather than the second or third equalities, whereas an example of writing out the null and alternative hypotheses is given at the top of p. 118. After writing out the null and alternative hypotheses, showing the critical value, and writing the formula for the calculated *t*, you will then use Stata to actually do all the calculations. Based on the above you will answer each question in words. **For part (d) use $\alpha = 0.01$**. **For part (e) use $\alpha = 0.10$**.

2. In this problem you will be working with some test score data from California. The data are from all 420 Kindergarten through 6th (K-6) and (K-8) grade districts with data available for 1998 and 1999 and are in the file calif_test_scores_data.csv, a comma-delimited file. The variables in the data set, which we will use at various times during the semester, are

Variable		Variable label
dist_code		district code
read_scr		average reading score
math_scr		average math score
county		county
enrl_tot		total enrollment
teachers		number of teachers
computer		number of computers
expn_stu		expenditures per student ($'s)
el_pct		percent of English learners (non-native English speakers)
meal_pct		percent qualifying for reduced-price lunch
pub_assist		percent qualifying for public-assistance
avginc		district average income (in $1000's)

Throughout this problem, you will use Stata except where it says something like "interpret in words." You will copy graphs and regression output from Stata into your answer, but you may write equations by hand (though you should try not to!). Remember that all commands and results must show up in your log file.

a. Import the Excel file into Stata.
b. Now, label the variables in Stata with the labels listed above.
c. Create two new variables, as defined here, and label them as well:

testscr		avg test score = (read_scr + math_scr)/2
str		student teacher ratio (enrl_tot/teachers)

d. Show what your current data set looks like with Stata's describe command and save it for yourself for future use (you will use this data set in a future problem set).
e. Compute the descriptive statistics for testscr and str and construct a scatterplot of testscr and str. Describe in a sentence or two what you visually see in the graph.
f. Many policymakers are interested in what would be the effect on test scores of changing the student-teacher ratio, since teachers are costly inputs into the production of the output, education. Run a regression of testscr on str. Interpret the numeric coefficient on str in a sentence. Explain whether this makes sense to you by thinking about what you have estimated. For the rest of the question I will refer to this model as **SLR**.
g. Predict out the fitted value of testscr and then plot the estimated line in the same graph as the scatterplot of the two variables.
h. Now consider whether there might be any influence from including an additional variable in the regression, the percent of English learners in a school district. Estimate your new model with both independent variables now included. Carefully interpret the numeric values of the estimated coefficients on both str and on el_pct, each in a sentence. For the rest of the question I will refer to this model as **MLR**.
i. Compare your MLR results with your SLR results, and explain briefly **both** how they compare **and** why you would or would not expect the results to be as they are.
j. Write out the estimated equation from your MLR results using the form I showed in class, using variable names, "hats," as appropriate, standard errors, and R^2.

k. Interpret the MLR R^2 and compare it to the one from the SLR model. Has there been an improvement and is any observed change expected or not? Show how to compute the MLR R^2 using the numbers for SST, SSR, and SSE.

l. For the MLR, calculate the elasticity of test scores with respect to the percent of English learners <u>at the mean</u> both by hand (showing the values for each part, though you may use Stata to do this) **and** confirming using Stata's margin command. Interpret numerically, and in a sentence, what would happen with a 10% increase in English learners in a school district.

m. For the MLR, calculate the <u>average</u> (marginal) elasticity of test scores with respect to the percent of English learners using Stata's margin command. Interpret numerically, and in a sentence, what would happen with a 10% increase in English learners in a school district.

n. Predict out the expected test score for a school district with a student-teacher ratio of 19 and 15.5% English learners. Make sure it is clear how you are doing your prediction.

o. Using your MLR specification, construct a 90 percent interval estimate around the estimated coefficient on str, showing where all your numbers come from.

p. Using your MLR specification, **manually** test the null hypothesis at the 0.05 level of significance that the coefficient on English learners is not equal to -0.75. Make sure all steps to your solution are clear and state your conclusion in words. Then, confirm with Stata using the lincom command.

q. A policy is being considered to increase teacher funding if there is evidence that reducing the student-teacher ratio by 1 increases test scores by **more** than 1 point. Set up this statistical test at the 0.05 level of significance, and using the MLR specification, being very **careful** about what is being asked. You may test this manually or with Stata but make sure all steps are clearly laid out and state your conclusion in words.