17.871, Political Science Lab
Spring 2014
Problem set # 3

Handed out: March 12
Due: March 19

Submit

1. A paper copy of your answers to each question, your graphs, and calculations.
2. Your log file. Please use one do file for the entire problem set.

You may work on these together, but you must write them up separately.

Part I: Placement of political parties (12 points, 2 points for each part)

In /mit/17.871/Examples you will find cces12_common_subset, which is a subset of the 2012 Cooperative Congressional Election Study. You will use that dataset for this question.

1. Using the regression command, estimate the effect of one's own ideological self-placement on how positively a respondent views the Tea Party movement. (Note, the proper weighting variable to use in this study is variable V103.)
2. Interpret the slope coefficient.
3. Interpret the confidence interval of the slope coefficient.
4. Interpret the Standard Error of Regression (SER, Stata calls it Root MSE).
5. Draw a graph that plots the estimated regression line against the independent variable. There is no need to plot the original data. (Why?)
6. Write a paragraph or two in which you explain, in substantive terms, what the slope coefficient means, so far as how voters view the Tea Party movement.

Part II: Scatter plots and ecological inference (14 points total, two points each)

Use the cces12_common_subset dataset that was used previously.

1. Create a publishable scatter plot of party identification (pid7) by family income *at the individual-level*. By publishable, I mean
    a) Recode income to meaningful values (and code irrelevant values to missing).
    b) Code irrelevant values to missing on party identification.
    c) Label both variables.
    d) You may use the "jitter" subcommand, if you think it helps.
2. Add a best-fit linear regression line to this graph.
3. Conduct the regression implied by # 2 and interpret the coefficients.

Note: To see how variables are coded, use tabulate, e.g., `tab faminc`. The variables in this data set have value labels, so tabulate does not reveal the actual values of the

variable. To see the actual values, use tabulate with the no label option, e.g., `tab faminc, nol`.
4. Create a publishable scatter plot of average partisan identification by average family income *at the state level*. You should have one data point for each state in your scatter plot, with a state's average partisanship on the y-axis and a state's average income on the x-axis. On the scatter plot, label the data points with state abbreviations. (Hint: Use the collapse command to average partisanship and income by state.)
5. Add a best-fit linear regression line to this graph.
6. Conduct the regression implied by # 5 and interpret the coefficients.
7. Does the relationship between partisan identification and income differ when analyzed at the state level vs the individual level? Briefly suggest an explanation for any difference you might see.

Part III: Interpreting regression coefficients (eight points total, two points each)

Using the data set **quartet.dta** in the Examples folder of the course locker:

1. Regress each y on its corresponding x (e.g., y1 on x1, y2 on x2). Present the results in a table with four columns, one for each regression. The rows of the table should be the slopes, the constant, the confidence intervals, and the Standard Error of Regression (Root MSE). You can create this table in Word, Excel, Open office, etc. No need to recode variables before running these regressions.
2. Interpret the coefficients and the Standard Error of Regression.
3. Do you believe these estimates? Explain.
4. What should you conclude about the use of regression (and other fancy statistical procedures/predictions) from this example?

Part IV: Functional forms (9 points, 3 points each)

Revisit Part II of this problem set. Because income is strongly right-skewed, this is a case for taking logarithms of the independent variable.

1. For the individual analysis in Part II, perform the regression in which the income variable has been transformed using logarithms.
2. Interpret the slope coefficient from this regression.
3. Draw a graph that compares the individual-level regression line generated in Part II of the problem set with the regression line from this part. Write a short paragraph (or two) in which you describe how the interpretation of the two regressions is different.

Part V: Calculation of regression coefficient (9 points, 3 points each)

This problem requires you to use Correlates of War dataset you downloaded and analyzed in Problem Set 2 (Part III).

We are interested in whether military spending is related to industrial capacity, measured in terms of iron and steel production. Using the Correlates of War dataset, do the following:

1. For the year 2007, generate variables that measure per capita military spending and per capita iron and steel production.
2. Calculate the slope coefficient for the regression of per capita military spending on per capita iron/steel production, using the formula for a slope coefficient that uses variances and covariances. (Hint: you can add a suboption to the correlate command in Stata to produce a variance-covariance matrix.)
3. Interpret the coefficient in substantive terms.