17.871, Political Science Lab
Spring 2015
Problem set # 4

Handed out: April 1
Due back: April 13

Part I: Interpreting coefficients recoded to a 0-1 scale (6 points, 2 points apiece)

Return to Part V on Problem Set 3. That problem required you to calculate regression coefficients using the variance/covariance formula. In this problem, you will do all the steps using Stata, including producing the regression coefficient.

1. Transform the independent variable (per capita GDP) to lie within the 0-1 interval.
2. Perform a regression, using the transformed independent variable, in which you regress (untransformed) CO_2 emissions on (transformed) per capita GDP.
3. Interpret the coefficients in substantive terms.

Part II. Calculating multiple regression coefficients (10 points — 2 points for questions 1, 3, and 4, and 4 points for question 2)

One of the issues surrounding elections in the United States is the question of whether voters should be required to show photo ID in order to vote. Supporters of photo ID say requiring them is necessary to battle voter fraud. Opponents of photo ID say they are being required because Republican state legislatures (which are the legislatures passing photo ID laws) are trying to make it harder for Democrats to vote. (Implied in this charge is that Democrats are less likely to possess a photo ID than Republicans.)

How much of the support expressed for photo ID is due to partisanship, and how much is due to beliefs about the prevalence of voter fraud? (We need multivariate analysis here because Republicans are more likely to believe voter fraud is widespread than are Democrats.)

The point of this question is to have you calculate how much of the statistical relationship between attitudes about voter fraud and attitudes about voter ID are directly due to attitudes about fraud, and the degree to which the indirect link through partisanship is involved.

The data for this part are in the file /mit/17.871/Examples/cces13_mit.dta. (This is the "MIT module" from the 2013 Cooperative Congressional Election Study, which was conducted immediately after the November 2013 elections.)

The variables I want you to focus on are the following:

MIT403: Degree of support for requiring voters to show photo ID

pid3: Three-point party identification variable

MIT402C: Fraud frequency - Pretending to be someone else

Before beginning the data analysis exercise, you should use Stata commands (such as "describe" and "tabulate") to see how these variables are coded. *All of them will need to be recoded or transformed somehow to respond to the questions posed below.*

In general, the following questions proceed with the following as the variables of interest, which you will construct from the three variables I have just identified:

- **ID support**, coded so that high values are associated with high levels of approval.
- **Republican Party identification,** coded in a single scale, so that being a Republican is associated with greater values than being a Democrat. (Think about this: what do you do with Independents?)
- **Fraud frequency,** coded so that people who believe that voters pretending to be someone else is common are given higher values of this variable.

The weight variable you need to use is creatively named *weight.*

1. Run three regressions and compare the results in a single table. All three regressions treat the "ID support" variable as the dependent variable and the other two as the independent variables. The first two regressions are bivariate regressions with "party identification" and "fraud frequency" as the independent variables. The third regression is the multivariate regression with both variables as independent variables. *Make sure your do-file contains the proper commands for coding the variables properly. Put comments in your do-file documenting why you are making the transformations you are using.*
2. Explain why the bivariate and multivariate coefficients are different (assuming that they are).
3. Construct a table in which you decompose the total regression effects into direct and indirect effects.
4. The following shows the regression output in which all the independent variables have been recoded to lie within the 0-1 interval. Interpret the coefficients.

```
. reg id_support rep01 fraud_freq01 [aw=weight]
(sum of wgt is   7.2329e+02)

      Source |       SS       df       MS              Number of obs =     719
-------------+------------------------------           F(  2,   716) =  110.52
       Model |  186.816118     2  93.4080588           Prob > F      =  0.0000
    Residual |  605.153991   716  .845187139           R-squared     =  0.2359
-------------+------------------------------           Adj R-squared =  0.2338
       Total |  791.970109   718  1.10302244           Root MSE      =  .91934

------------------------------------------------------------------------------
  id_support |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rep01 |    .596045   .0878665     6.78   0.000     .4235383    .7685517
fraud_freq01 |   1.068792   .0979743    10.91   0.000     .8764405    1.261143
       _cons |   2.427437   .0617943    39.28   0.000     2.306117    2.548756
------------------------------------------------------------------------------
```

5. If we were concerned about omitted variables bias in the multivariate analysis you conducted in Question 1 (and we always should be), what should concern us in this regression? Write a short paragraph in which you discuss (in substantive terms) what might contribute to omitted variables bias in this regression.

Part III. Dummy variables and interaction terms (10 points — 2 points for questions 1–3 and 4 points for question 4)

Return to Part II. Test the proposition that Democrats are more likely to form their opinions about photo ID based on their beliefs about fraud than are Republicans.

1. Perform this test using two separate regressions (one for Democrats and one for Republicans; you may omit Independents for this part). Draw up a table that reports the results side-by-side.
2. Draw a nice-looking graph to illustrate your answer to the first question.
3. Using the regression coefficients and the graphic visualization, how do these results confirm, or disconfirm, the idea that attitudes toward fraud influence opinions about photo ID, depending on one's partisanship?
4. Perform the same regression as the first question, this time performing one regression involving a dummy variable and interaction term. Report the coefficients. Show why these coefficients are, or are not, consistent with what you got in the two bivariate regressions.