

**17.871, Political Science Lab**  
**Spring 2013**  
**Problem set # 4**

Handed out: March 13  
Due back: April 3

**Part I: Interpreting coefficients recoded to a 0-1 scale (8 points, 2 points apiece)**

Return to Part IV on Problem Set 3. That problem required you to hand-calculate regression coefficients. In this problem, you will do all the steps you need to do using Stata.

1. For the year 2007, generate variables that measure per capita military spending and per capita iron and steel production.
2. Transform these variables to lie within the 0-1 interval.
3. Perform a regression, using the transformed variables, in which you regress (transformed) per capita military spending on (transformed) per capita iron/steel production.
4. Interpret the coefficient in substantive terms.

**Part II. Calculating multiple regression coefficients (10 points — 4 points for question 1 and 2 points each of the rest)**

We are interested in the role that the state of the economy had in evaluating the job that Barack Obama was doing in 2012. The problem, of course, is that Democrats probably will say the economy is doing better than Republicans, as a way of expressing support for the incumbent president, and vice versa. So, the question is what is the direct effect of the economy on evaluations of Obama and what is the indirect effect, due to partisanship?

The data are taken from the MIT module of the 2012 Congressional Cooperative Election Study. The following are the three relevant variables:

**obama\_approve:** 4 = strongly approve the job that Obama is doing as president, 3 = somewhat approve, 2 = somewhat disapprove, 1 = strongly disapprove

**economy\_good:** 1 = the economy has gotten much worse over the past year, 2 = the economy has gotten worse over the past year, 3 = the economy has stayed about the same, 4 = the economy has gotten better, 5 = the economy has gotten much better

**dem7:** 1 = strong Republican, 2 = not very strong Republican, 3 = lean Republican, 4 = independent, 5 = lean Democrat, 6 = not very strong Democrat, 7 = strong Democrats.

The following is the variance-covariance matrix between the relevant variables:

```
. corr obamaapprove economygood dem7,cov
(obs=922)
```

	obamaa~e	econom~d	dem7
obamaapprove	1.58385		
economygood	1.06552	1.26169	
dem7	2.27577	1.63027	5.00448

1. Calculate by hand both the *bivariate* and *multivariate* regression coefficients with **obama\_approve** as the dependent variable and **economygood** and **dem7** as the independent variable(s). (In other words, calculate four different coefficients, two in separate bivariate regressions, and then two more in a single multivariate regression. You will need to consult the relevant slide from the multivariate regression lecture to do this.)
2. Explain why the bivariate and multivariate coefficients are different (assuming that they are, in fact, different).
3. Construct a table in which you decompose the total regression effects into direct and indirect effects.
4. The following shows the regression output when all the variables have been recoded to lie within the 0-1 interval. Interpret the coefficients.

```
. reg obama01 economy01 dem701
```

Source	SS	df	MS	Number of obs = 922		
Model	120.622902	2	60.3114508	F( 2, 919) = 1336.92		
Residual	41.458204	919	.0451123	Prob > F = 0.0000		
Total	162.081106	921	.175983828	R-squared = 0.7442		
				Adj R-squared = 0.7437		
				Root MSE = .2124		

  

obama01	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
economy01	.5915803	.0327517	18.06	0.000	.5273034	.6558572
dem701	.6204223	.0246674	25.15	0.000	.5720114	.6688332
_cons	-.1497416	.0137507	-10.89	0.000	-.1767279	-.1227553

### Part III. Dummy variables and interaction terms

Go to the Examples folder of the 17.871 course locker. There you will find a file called section5.dta. The purpose of this exercise is to estimate the relationship between the percentage of votes cast for Barack Obama in and the percentage of the population that is nonwhite. The data are at the county level.

This relationship has a direct bearing on the workings of Section 5 of the Voting Rights Act, which is currently being challenged in the U.S. Supreme Court. You may read about the arguments made in the Supreme Court two weeks ago by following [this link](#).

Section 5 of the Voting Rights Act mandates that certain parts of the U.S. obtain “preclearance,” from the U.S. Justice Department or a federal court, of any change in their voting laws before they can be implemented. [This link](#) tells you more about Section 5.

One important empirical question that surrounds Section 5 is whether whites and black vote in racially polarized patterns in the areas of preclearance, compared to the rest of the country.

To see whether this pattern holds, we will run a regression. The `section5.dta` dataset has the following variables in it:

**state:** state

**county:** county

**nonwhitepct:** percent of the county that was nonwhite in the 2010 census

**obamapct:** percent of the two-party vote received by Obama in 2012

**covered:** dummy variable equal to 1 if the county is covered by Section 5, 0 otherwise.

Do the following:

1. Perform the regression of the Obama vote percent on the nonwhite population percent. Interpret the coefficients.
2. *Using interaction terms* test whether the slope coefficient and intercept of the previous regression are the same in counties covered by Section 5 and those not covered by Section 5.
3. Interpret the coefficients from the previous step in substantive terms.
4. Plot the predicted values of the last regression against the independent variable, to compare what the regression lines look like in covered and non-covered counties. Make sure to make the graph pretty (i.e., don't take the defaults for things like axis labels and the legend.)