Part I: Interpreting coefficients recoded to a 0-1 scale (8 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. 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 — 2 points for questions 1, 3, and 4, and 4 points for question 2)

One of the issues swirling around the upcoming midterm congressional elections is the degree to which the Affordable Care Act (“Obamacare”) has shifted attitudes of the electorate toward or away from support from President Obama. Of course one of the tricks in studying the relationship between attitudes toward Obamacare and attitudes toward President Obama is that they are likely to be contaminated by one’s partisanship (which is generally considered fixed in the short term).

The point of this question is to have you calculate how much of the statistical relationship between attitudes toward Obamacare and attitudes toward President Obama are directly due to attitudes toward Obama, 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:
CC312a: Job approval of President Obama.

pid3: Three-point party identification variable

MIT418C: Policy issue – Government should help people get low cost health care.

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:

- **Obama approval**, coded so that high values are associated with high levels of approval.
- **Party identification**, coded in a single scale, so that being a Democrat is associated with greater values than being a Republican. (Think about this: what do you do with Independents?)
- **Government involvement in health care**, coded so that people who have the strongest beliefs that government should help people get low cost health care are coded with the highest values.

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 “Obama approval” variable as the dependent variable and the other two as the independent variables. The first two regressions are bivariate regressions with “party identification” and “government involvement” 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 variables have been recoded to lie within the 0-1 interval. Interpret the coefficients.
. reg obama_approve01 govhealth01 dem01 [aw=weight]
(sum of wgt is \(8.589e+02\))

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 871</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>55.5790354</td>
<td>2</td>
<td>27.7895177</td>
<td>(F(2, 868) = 408.38)</td>
</tr>
<tr>
<td>Residual</td>
<td>59.0651891</td>
<td>868</td>
<td>.068047453</td>
<td>R-squared = 0.4848</td>
</tr>
<tr>
<td>Total</td>
<td>114.644224</td>
<td>870</td>
<td>.131774971</td>
<td>Adj R-squared = 0.4836</td>
</tr>
</tbody>
</table>

| obama_app01 | Coef.   | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|-------------|---------|-----------|-----|-----|----------------------|
| govhealth01 | .352597 | .0271322  | 13.00 | 0.000 | .2993445 -.4058494 |
| dem01 | .3882384 | .0254519 | 15.25 | 0.000 | .338284 -.4381928 |
| _cons | -.0602261 | .0182119 | -3.31 | 0.001 | -.0959706 -.0244817 |

5. If we were concerned about specification error (and we always should be), how much should we be concerned about specification bias in these regressions? Be as specific as possible. (In other words, answers like “a lot,” “a little,” or “depends” aren’t enough.)

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 Independents are more likely to form their evaluations of President Obama based on their attitudes toward government involvement in health care than are Democrats and Republicans.

1. Perform this test using three separate regressions (one each for Democrats, Republicans, and Independents). 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 government involvement in health care influence attitudes toward Obama differently, depending on one’s partisanship?
4. Perform the same regression as the first question, this time performing one regression involving dummy variables and interaction terms. Report the coefficients. Show why these coefficients are, or are not, consistent with what you got in the three bivariate regressions.