PSC 508

Jim Battista

Univ. at Buffalo, SUNY

Interactions

A model with two dummies

- Say we have a model predicting personal ideology with sex and college education

```
. reg lib women college

      Source |       SS       df       MS              Number of obs =     897
-------------+------------------------------           F(  2,   894) =    4.36
       Model |  18.6866953     2  9.34334767           Prob > F      =  0.0131
    Residual |  1916.94095   894  2.14422925           R-squared     =  0.0097
-------------+------------------------------           Adj R-squared =  0.0074
       Total |  1935.62765   896  2.16029871           Root MSE      =  1.4643

------------------------------------------------------------------------------
    lib_cons |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       women |  -.2342654   .0978602    -2.39   0.017    -.4263278    -.042203
     college |  -.1829796   .1020319    -1.79   0.073    -.3832295    .0172703
       _cons |   4.457896   .0797195    55.92   0.000     4.301437    4.614355
------------------------------------------------------------------------------
```

So?

- So this regression suggests that women and college graduates are more liberal than men and non-college-graduates
- It says that if you're a woman, and you graduated from college, we move you .23 points to the left for being a woman and than another .18 points for your graduation
- But what if sex and education don't work that way?
- What if the effect is concentrated in just college-educated women being particularly liberal?
- We can look for this with an interaction
- Interactions allow the effect of education to be different for men and women, and for the effect of sex to be different for grads and nongrads

Warning! DANGER WILL ROBINSON!

- Interactions can be hard to interpret
- Our discussion might make your brain melt a little; that's normal
- Interactions are fairly serious juju that should not be employed willy-nilly; you should have a good reason to mess with them
- We're going to talk about them primarily because it's relatively common for models you'll read to use them, so you should have a pretty good idea what they're saying

The Not Rocket Surgery Part

- To interact two variables, you just multiply them together
- In stata, just `gen interaxn=variable1*variable2`
- Even easier in R; don't need to actually generate anything
  - Just specify it in your lm() command (more later)

Interacting sex and education

- So we interact sex and education
- The new variable looks like this:

	Men	Women
Non-Grads	0	0
Grads	0	1

Running the model

- Then just run the model with the original two variables, and the interaction
- Stata
  - `gen interaxn=iv1*iv2`
  - `reg dv iv1 iv2 interaxn`
- R
  - `model<-lm(dv~iv1*iv2)`
  - See? Just specify the interaction directly

Interpreting interactions – BRING THE PAIN!

- If interacting IV1 and IV2
- Coef. on IV1: effect of IV1 when IV2 is zero
- Coef. on IV2: effect of IV2 when IV1 is zero
- Coef. on interaction: change in coef on IV1 when IV2 is 1
    - –OR– change in coef on IV2 when IV1 is 1!
    - Kinda different ways to say same thing

A real example in Stata

```
. gen womenXcoll=women*college

. reg lib women college womenXcol

      Source |       SS       df       MS              Number of obs =     897
-------------+------------------------------           F(  3,   893) =    4.59
       Model |  29.3798004     3  9.79326679           Prob > F      =  0.0034
    Residual |  1906.24785   893  2.13465604           R-squared     =  0.0152
-------------+------------------------------           Adj R-squared =  0.0119
       Total |  1935.62765   896  2.16029871           Root MSE      =   1.461

------------------------------------------------------------------------------
    lib_cons |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       women |  -.0709383    .121898    -0.58   0.561    -.3101781    .1683016
     college |   .0474903   .1448019     0.33   0.743    -.2367015    .3316821
  womenXcoll |  -.4557331    .203621    -2.24   0.025    -.8553645   -.0561016
       _cons |   4.372263    .088265    49.54   0.000     4.199032    4.545494
------------------------------------------------------------------------------
```

What does all this mean?

Same example in R

```
> model2<-lm(lib_cons~women*college)
> summary(model2)

Call:
lm(formula = lib_cons ~ women * college)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4198 -0.8931 -0.3013  1.5802  3.1069

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.37226    0.08827  49.536   <2e-16 ***
women         -0.07094    0.12190  -0.582   0.5607
college        0.04749    0.14480   0.328   0.7430
women:college -0.45573    0.20362  -2.238   0.0255 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.461 on 893 degrees of freedom
  (279 observations deleted due to missingness)
Multiple R-squared:  0.01518,Adjusted R-squared:  0.01187
F-statistic: 4.588 on 3 and 893 DF,  p-value: 0.003397
```

Another way to think about this example

We can also just look directly at the average ideology for each category

```
. tab women coll, summ(lib_cons)

          Means, Standard Deviations and Frequencies of lib_cons

           |       college
     women |         0          1 |     Total
-----------+----------------------+----------
         0 | 4.3722628  4.4197531 | 4.3899083
           | 1.4064597  1.5024965 |  1.441356
           |       274        162 |       436
-----------+----------------------+----------
         1 | 4.3013245  3.8930818 | 4.1605206
           | 1.3995484  1.6171204 | 1.4891794
           |       302        159 |       461
-----------+----------------------+----------
     Total | 4.3350694  4.1588785 | 4.2720178
           | 1.4020678  1.5800487 | 1.4697955
           |       576        321 |       897
```

But wait! There's more!

- You can also interact a dummy with a continuous variable
- Same way; just multiply them (or tell R you want them multiplied)
- Coefficient on continuous: effect of continuous when dummy=0
- Coefficient on interaction: change in coefficient on continuous when dummy=1

EVEN MOAR!

- BUT AT THE SAME TIME
- Coefficient on dummy: effect of dummy when continuous=0
- Coefficient on interaction: how much the effect of the dummy changes when you increase the continuous variable by 1

Thinking about this graphically

- Dummy variable – change in intercept
- Interaction – change in slope
- Try this with rejection rates as a function of public, tuition, interaction

A final note

- Can you interact a continuous variable with another continuous variable?
  - Yes
  - Coef on variable1: effect of variable1 when variable2=0
  - Coef on variable2: effect of variable2 when variable1=0
  - Interaction: how much the effect of variable1 (2) changes when you increase variable2 (1) by 1
- Can you interact more than two variables?
  - Yes – "three way interaction"
  - Say race, sex, education, or race, sex, income
  - Real pain to interpret