EEP/IAS 118 - Introductory Applied Econometrics, Lecture 9

Gregory Lane

Summer 2017

This Lecture

Topics

- Dummy Variables
- Interactions
- Chow Test

Assignments

- Problem Set 4 due Monday, July 24th
- Quiz 4 next Tuesday, July 25th

Dummy Variables

Dummy variables: are binary variables (or zero-one variables).
For example: urban or rural

How do we interpret dummies? Let's look at a classic question, the wage gap between men and women:

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + u$$

Then, β_1 equals:

$$E[wage|female = 0, educ] = \beta_0 + \beta_2 educ$$
$$E[wage|female = 1, educ] = \beta_0 + \beta_1 + \beta_2 educ$$
$$E[wage|female = 1, educ] - E[wage|female = 0, educ] = \beta_1$$

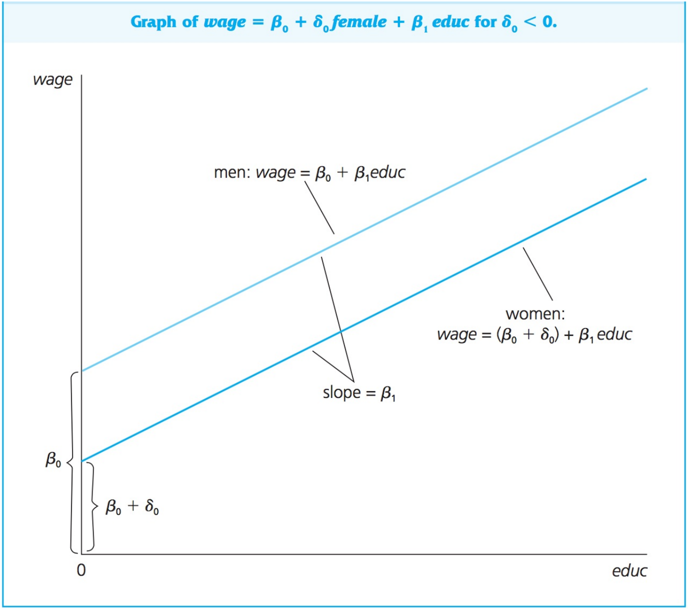
Q: Why don't we include *both* male and female in this regression?

Dummy Variables

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + u$$

- Difference between females and males wages at a given education level is β_1.
- We can also think of this dummy as introducing an intercept shift between males and females:
  - The intercept for males is β_0
  - The intercept for female is $\beta_0 + \beta 1_0$

Dummy Variables (note $\delta = \beta_1$)



Policy and Dummy Variables

This type of analysis will simply tell us if there is a gap between men and women's earnings

In the policy world, we need to go further:

- Is this result "robust"
- Why is there a gap?
    - Different jobs?
    - Discrimination?
    - Differential returns to eduction?
- The answers to these questions will determine the appropriate policy response

Wage Gap

```
. sum female wage if female==1
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      female |      1033           1           0          1          1
        wage |      1033    16.12258    9.715608      2.125     72.125
. sum female wage if female==0;
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      female |       967           0           0          0          0
        wage |       967    20.72326    12.71402         .7   82.42857
```

Yes, there is a wage gap

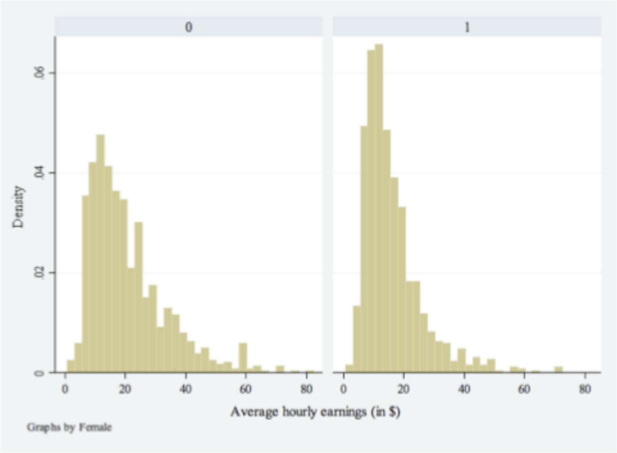
Wage Gap

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     967    20.72326    .4088552    12.71402    19.92091    21.52561
       1 |    1033    16.12258    .3022872    9.715608    15.52942    16.71575
---------+--------------------------------------------------------------------
combined |    2000    18.34701    .2570348    11.49495    17.84293    18.85109
---------+--------------------------------------------------------------------
    diff |            4.600677    .5040778                3.612104     5.58925
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   9.1269
Ho: diff = 0                                     degrees of freedom =     1998

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Yes, it is statistically significant at the 1% level

Wage Gap



0
1
Density
0
.02
.04
.06
0
20
40
60
80
0
20
40
60
80
Average hourly earnings (in $)
Graphs by Female

Policy and Dummy Variables

Now we have to investigate why:

1. We want to investigate whether there are other differences between men and women that could drive this effect
   - I.e. is there an omitted variable?
2. We want to test whether the *return* to certain characteristics is different for men and women
   - In our example of wages across men and women, this would mean asking whether the coefficient on $educ$ or $exper$ is different across these groups

Policy and Dummy Variables

```
. tabstat wage educ exper union service profocc , by(female)

Summary statistics: mean
  by categories of: female (Female)
```

female	wage	educ	exper	union	services	profocc
0	20.72326	13.5274	20.3061	.1323681	.1323681	.1664943
1	16.12258	13.73185	20.84608	.1452081	.1703775	.2545983
Total	18.34701	13.633	20.585	.139	.152	.212

There are some differences, particularly in job type

Policy and Dummy Variables

How much of the gap do these difference explain?

```
. reg wage female

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  1,  1998) =   83.30
       Model |   10571.589     1   10571.589           Prob > F      =  0.0000
    Residual |  253563.889  1998  126.908853           R-squared     =  0.0400
-------------+------------------------------           Adj R-squared =  0.0395
       Total |  264135.478  1999  132.133806           Root MSE      =  11.265

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -4.600677   .5040778    -9.13   0.000     -5.58925   -3.612104
       _cons |   20.72326   .3622703    57.20   0.000     20.01279    21.43373
------------------------------------------------------------------------------
```

Policy and Dummy Variables

```
. reg wage female educ exper union service profocc

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  6,  1993) =  118.29
       Model |  69360.4741     6  11560.079           Prob > F      =  0.0000
    Residual |  194775.003  1993  97.7295551           R-squared     =  0.2626
-------------+------------------------------           Adj R-squared =  0.2604
       Total |  264135.478  1999  132.133806           Root MSE      =  9.8858

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -5.161436   .4466179   -11.56   0.000    -6.037323   -4.285549
        educ |   2.000723   .1182173    16.92   0.000      1.76888    2.232565
       exper |   .1711896   .0175183     9.77   0.000     .1368336    .2055457
       union |   2.229349   .6460632     3.45   0.001      .962319    3.496379
    services |  -2.729558   .6422974    -4.25   0.000    -3.989203   -1.469914
     profocc |   1.531844   .6118991     2.50   0.012     .3318149    2.731873
       _cons |  -10.00724   1.645929    -6.08   0.000    -13.23516   -6.779315
------------------------------------------------------------------------------
```

The coefficient on $female$ remains negative and significant. This is a “robust” result so far

Interactions: Dummy Variables

Now we move to considering whether women have differential returns to their characteristics. Consider this model:

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 female \times educ + u$$

- What is the marginal effect of increasing education?
- Intuitively, think about regrouping all the terms that have education in them:

$$E[wage|educ, female] = \beta_0 + \beta_1 female + \underbrace{(\beta_2 + \beta_3 female)}_{regrouped} educ$$

So the "education effect", i.e the marginal effect of education is $\beta_2 + \beta_3 female$.

Interactions: Dummy Variables

We can express the marginal effect of educ is:

$$\frac{\partial E[wage]}{\partial educ} = \beta_2 + \beta_3 female$$

The marginal effect of education *depends* on the value female take. So we need to plug in:

- If female==0 (i.e. male) then the marginal effect of $educ$ is:

$$\frac{\partial E[wage]}{\partial educ} = \beta_2$$

- If female==1 (i.e. female) then the marginal effect is:

$$\frac{\partial E[wage]}{\partial educ} = \beta_2 + \beta_3$$

Interactions: Dummy Variables

Similarly, we can ask what the marginal effect of $female$ is.

$$\frac{\Delta E[wage]}{\Delta fem} = \beta_2 + \beta_3 educ$$

- The "effect" of being female depends on the value of $educ$
- To evaluate this, you will be asked to look at a particular value of education
- Typically we use the median value of the continuous variable
- Substituting educ=10 for example gives:
$$\frac{\Delta E[wage]}{\Delta fem} = \beta_1 + \beta_3 * 10$$

Interactions: Dummy Variables

How do we interpret each coefficient in the regression?

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 female \times educ + u$$

- **Interpreting** β_0: parameter is the intercept for of males.
- **Interpreting** β_1: parameter is the difference in the intercepts between women and men.
- **Interpreting** β_2: parameter reflects the effect of education for males.
- **Interpreting** β_3: parameter reflects the difference in the marginal returns to education between males and females

Interactions: Dummy Variables

How do we interpret each coefficient in the regression?

$$wage = 65 - 10 female + 3 educ - 2 female \times educ + \hat{u}$$

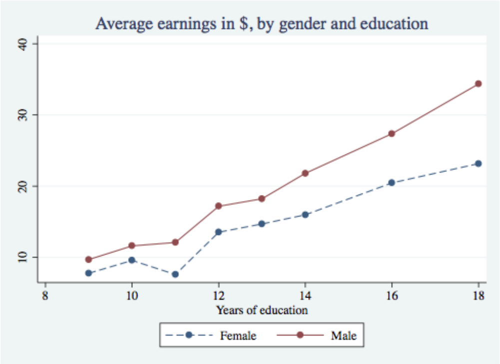
1. Interpret each coefficient above in words
2. Draw a graph that depicts the $\widehat{wage}$ for males and females
3. Redraw the same graph for the new coefficients below

$$wage = 65 - 15 female + 2 educ + 3 female \times educ + u$$

Interactions: Dummy Variables

Graphs of equation (7.16): (a) $\delta_0 < 0, \delta_1 < 0$; (b) $\delta_0 < 0, \delta_1 > 0$.

Wage Gap and Interactions



Looks like the first panel. The wage gap is *larger* for individuals with more education

Based on this picture, what are your predictions for each of the four $\hat{\beta}_j$?

Wage Gap and Interactions

What does this look like in regression:

```
. g femeduc=female*educ
. reg wage female educ femeduc
      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  3,  1996) =  182.86
       Model |  56945.4372     3  18981.8124           Prob > F      =  0.0000
    Residual |   207190.04  1996  103.802625           R-squared     =  0.2156
-------------+------------------------------           Adj R-squared =  0.2144
       Total |  264135.478  1999  132.133806           Root MSE      =  10.188

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   8.838706   3.013838     2.93   0.003     2.928108     14.7493
        educ |   2.772576   .1560434    17.77   0.000     2.466551    3.078601
     femeduc |  -1.019981   .2186047    -4.67   0.000    -1.448698   -.5912633
       _cons |   -16.7825   2.136137    -7.86   0.000    -20.97179   -12.59321
------------------------------------------------------------------------------
Female effect on wage = (8.8 – 1.02 educ)
Education effect on wage = (2.77 - 1.02 female)
```

Does each coefficient conform to our predictions? Why or why not?

Wage Gap and Interactions

We can run the same analysis with union membership: Do unions have a differential impact for men and women?

```
. gen femunion=female*union
. reg wage female union femunion educ exper

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  5,  1994) =  134.42
       Model |  66586.3651     5   13317.273           Prob > F      =  0.0000
    Residual |  197549.112  1994  99.0717715           R-squared     =  0.2521
-------------+------------------------------           Adj R-squared =  0.2502
       Total |  264135.478  1999  132.133806           Root MSE      =  9.9535

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -5.094864   .4801934   -10.61   0.000    -6.036597    -4.15313
       union |   2.577524   .9480136     2.72   0.007     .7183235    4.436725
    femunion |   -.612306    1.29469    -0.47   0.636    -3.151394    1.926782
        educ |   2.228979   .1073513    20.76   0.000     2.018447    2.439512
       exper |   .1756898   .0175311    10.02   0.000     .1413085    .2100711
       _cons |  -13.33871   1.533331    -8.70   0.000     -16.3458   -10.33161
------------------------------------------------------------------------------
```

Wage Gap Research

This is as far as we'll take this investigation on our own. However there is much on-going work on this topic

The most recent paper on the wage gap is a working paper by Barth, Kerr, and Olivetti (2017):

- Wage gap starts small as men and women enter the workforce
- The wage gap increases sharply in the late 20s and early 30s, concentrated in high education jobs
- This effect is driven by *married* women - implies that having children is a big driver of the wage gap
    - High education jobs seem to place premiums on long hours in the office and face time, something women are more likely to sacrifice when they have children
    - Women are more likely to change jobs without a wage increase (likely following their spouse)

Hypotheses with Interactions

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 female \times educ + u$$

1. Write the null and alternative hypothesis to test that the return to education is the same for women and men
2. Write the null and alternative hypothesis to test that average wages are identical for men and women who have the same levels of education:

Hypotheses with Interactions

How can we test hypotheses when the same variable appears multiple time:

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 female \times educ + u$$

1. Write the null and alternative hypothesis to test that the return to education is the same for women and men:

$$H_0 : \beta_3 = 0 \quad vs. \quad H_1 : \beta_3 \neq 0$$

2. Write the null and alternative hypothesis to test that average wages are identical for men and women who have the same levels of education:

$$H_0 : \beta_1 = 0 \;\; \& \;\; \beta_3 = 0 \quad vs. \quad H_1 : \beta_1 \neq 0 \;\; \&/or \;\; \beta_3 \neq 0$$

Hypotheses with Interactions

What if we want to test whether the an entire model is different for two groups

- E.g. Is the wage equation different for men and women

There are two ways we might do this:

1. Interact *all* variables with the dummy for our group and then run on F-test for the interaction terms.
   - This can be tedious
2. Conduct a "Chow test"
   - A Chow test is the same as an F-test, but can save you time in certain situations

Chow Test

To conduct a Chow test you

1. Run three regressions without any interaction terms:
   - The regression for each group separately (e.g. for men then women)
   - A "pooled" regression with both groups included
2. Plugging the results of these three regressions into this formula to get an F-stat

$$F = \frac{\left(SSR_{pooled} - (SSR_w + SSR_m)\right)/q}{(SSR_w + SSR_m)/(n-(2k+1))}$$

   Where $q = k + 1$

3. Proceed as you would on an F-test

Chow Test

$$F = \frac{\Big(SSR_{pooled} - (SSR_w + SSR_m)\Big)/q}{(SSR_w + SSR_m)/(n - (2k+1))}$$

Note that this equation is identical to the normal F-test because

$$SSS_{UR} = SSR_w + SSR_m$$

And the SSR_{pooled} is the same as the SSR_R. If we plug this in we get our old F-stat:

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n - k_{UR} - 1)}$$

So, this is the exact same test, just a different way to calculate it

Interactions: Two Continuous Variables

Now consider this model:

$$wage = \beta_0 + \beta_1 age + \beta_2 educ + \beta_3 age \times educ + u$$

What is the marginal effect of age? Again, re-group all the terms with age in them:

$$E[wage|educ, age] = \beta_0 + \underbrace{(\beta_1 + \beta_3 educ)}_{regrouped} age + \beta_2 educ$$

As before, the "age effect" i.e the marginal effect of age is $\beta_1 + \beta_3 educ$.

Interactions: Two Continuous Variables

$$\frac{\partial E[wage]}{\partial age} = \beta_1 + \beta_3 educ$$

- Marginal effect varys with $educ$. To get one value, we can plug in for $educ$ (usually with the media)
- The expected wage for people with 10 years of education is $\beta_1 + \beta_3 * 10$

We can follow the same intuition for education:

$$\frac{\partial E[wage]}{\partial educ} = \beta_2 + \beta_3 age$$

- The marginal effect of education on expected wage for people with 20 years of age is $\beta_2 + \beta_3 * 20$

Interacting Two Dummy Variables

Changing the model so we have two dummy variables:

$$wage = \beta_0 + \beta_1 married + \beta_2 female + \beta_3 female \times married + u$$

Regrouping terms on married:

$$E[wage|married, female] = \beta_0 + \underbrace{(\beta_1 + \beta_3 female)}_{regrouped} married + \beta_2 female$$

The marginal effect of marriage on wage:

$$\frac{\Delta E[wage]}{\Delta mar} = \beta_1 + \beta_3 female$$

Interacting Two Dummy Variables

Now, when we plug in for $female$ we only have two options (0,1)

- Substituting female=0 (i.e you are male) gives:

$$\frac{\Delta E[wage]}{\Delta mar} = \beta_1$$

  The effect of being married on expected wage for **males** is β_1

- Substituting female = 1 (i.e you are female) gives:

$$\frac{\Delta E[wage]}{\Delta mar} = \beta_1 + \beta_3$$

  The effect of being married on expected wage for **females** is $\beta_1 + \beta_3$

Interaction Two Dummy Variables

$$wage = \beta_0 + \beta_1 married + \beta_2 female + \beta_3 female \times married + u$$

Interprete each coefficient:

1. β_0:
2. β_1:
3. β_2:
4. β_3:

Q5: Calculate the average wage of married females?

Interaction Two Dummy Variables

$$wage = \beta_0 + \beta_1 married + \beta_2 female + \beta_3 female \times married + u$$

Interprete each coefficient:

1. β_0: The average wage for single males.
2. β_1:The average effect of being married for a male.
3. β_2: The average effect of being female for a single individual.
4. β_3: The differential effect of being married for a woman relative to what it is for a man.

Q5: Calculate the average wage of married females?

$$= \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$$

Another Example

$$colGPA = \beta_0 + \beta_1 female + \beta_2 athlete + \beta_3 female \times athlete$$

colgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
female	.1248531	.0208526	5.99	0.000	.0839707	.1657355
athlete	-.3015775	.0553063	-5.45	0.000	-.4100075	-.1931474
fem_ath	.1963892	.1129543	1.74	0.082	-.0250621	.4178405
_cons	2.608557	.0141477	184.38	0.000	2.58082	2.636294

1. What is the effect of $female$ on predicted wages?
2. What's the effect of $athlete$ on predicted wages?
3. Interpret each coefficient

Another Example

1. What's the effect of $female$ on predicted wages?

$$0.1248531 + 0.1963892 athlete$$

2. What's the effect of $athlete$ on predicted wages?

$$-0.3015775 + .1963892 female$$

3. Interpret each coefficient:
   - β_0: Average GPA for male non-athletes
   - β_1: Predicted average effect of being a female for non-athletes
   - β_2: Predicted average effect of being an athlete for males
   - β_3: Differential effect of being an athlete for a woman relative to what it is for a man