Printout

Tuesday, April 05, 2016 6:39 AM

Lecture 21
CHAPTER 9: Dummy Variable Regression Models

In the previous chapter, the dependent and independent variables in our multiple regression models have had **quantitative** meaning. For example, the salary of CEO.annual firm sales,return on equity in percent, and return on firm's stock. In each case the magnitude of the variable conveys useful information.

However, in the empirical work, we must also incorporate **qualitative factors** into regression models. The gender or race of an individual, the industry of a firm (manufacturing, retail, and so on), and the region in Thailand where a city is located (north, south, west, and so on) are all considered as the qualitative factors.

8.1 Describing Qualitative Information

Normally, qualitative factors often come in the form of binary information:

Example:

[1] A person is female or male or female. GENDER

[2] A firm offers a certain kind of employee pension plan or it does not. YES 1 / NO 0

[3] A farm is located nearby the dam or not. YES / NO

All of these examples, the relevant information can be captured by defining a **binary variable** or a zero-one variable.

In econometrics, binary variables are most commonly called **dummy variables**, although this name is not especially descriptive.

In defining a dummy variable, we must decide which event is assigned the value one and which is assigned the value zero.

Question: Why do we use the the values zero and one to describe qualitative information?

Answer: These values are arbitrary: any two different values would do. The real benefit of capturing qualitative information using zero-one variable is that it leads to regression models where the parameters have very natural interpretations.

QUANTITATIVE

(GENDER) 0 IF ♀ / 1 IF ♂

(RACE) 0 WHITE / 1 BLACK

Firm i | X_2 SALARY | X_3 YEAR | DUM1 | DUM2

i: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ..., 200

DUM1: 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, ...

DUM2: 0, 0, 0, 0, 1, 0, 0, 1, 0

8.2 A Single Dummy Independent Variable

Suppose we would like to estimate the following simple model of hourly wage determination:

$$wage_i = \beta_0 + \delta_0 female + \beta_1 edu + u_i$$

$n = 526$

$WAGE = \beta_0 + \delta_0 D_1 + \beta_1 edu + u_i$

WHERE $D_1 = 1$ IF FEMALE
$D_1 = 0$ IF MALE

FEMALE IS A BINARY VARIABLE TAKING VALUE OF ONE FOR FEMALE AND VALUE OF ZERO FOR MALE

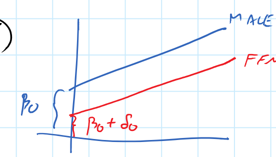
FEMALE = 1 IF THE PERSON IS FEMALE
FEMALE = 0 IF THE PERSON IS MALE

δ_0 REFLECTS WAGE DIFFERENTIAL BET. MALES AND FEMALE, GIVEN THE SAME LEVEL OF EDUCATION

IF $\delta_0 < 0$, IT IMPLIES THAT, ON AVERAGE, FEMALES EARN LESS WAGES THAN MALES, GIVEN THE SAME LEVEL OF EDUCATION.

$E(WAGE \mid FEMALE = 1, EDU) = \beta_0 + \delta_0 \cdot 1 + \beta_1 edu$
$= (\beta_0 + \delta_0) + \beta_1 edu$ — ①

$E(WAGE \mid FEMALE = 0 \text{ (MALE)}, EDU) = \beta_0 + \delta_0 \cdot 0 + \beta_1 edu$
$= \beta_0 + \beta_1 edu$ — ②



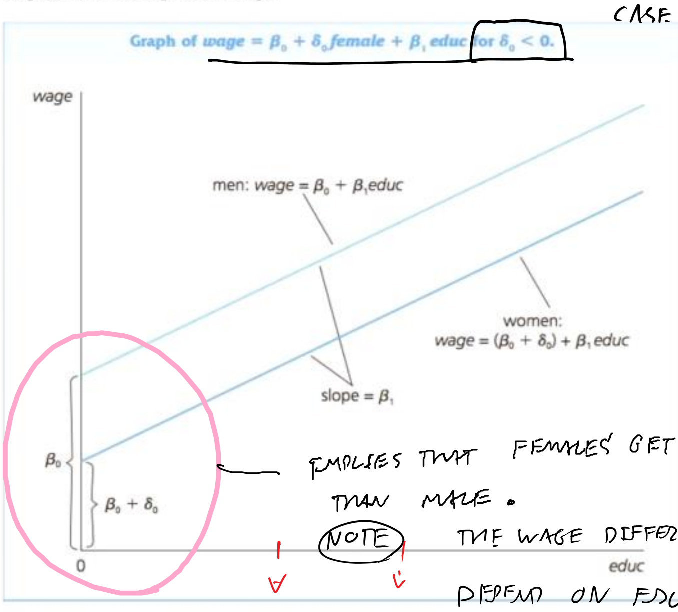
① − ②;

$$\underbrace{E(WAGE \mid FEMALE = 1, EDU) - E(WAGE \mid FEMALE = 0, EDU) = \delta_0}$$

GIVEN THE SAME LEVEL OF EDUCATION, WAGE DIFFERENTIAL IS DUE TO GENDER ONLY!

Figure 9.1: Graph of Wage

CASE WHERE $\delta_0 < 0$



IMPLIES THAT FEMALES GET LOWER HOURLY WAGE THAN MALE.

NOTE THE WAGE DIFFERENTIAL DOES NOT DEPEND ON EDUCATION.

THIS EXPLAINS WHY THE WAGE-EDUCATION PROFILES OF THE TWO ARE PARALLEL.

$$WAGE = \beta_0 + \delta_0 \, FEMALE + \beta_1 EDU + u_i$$

FEMALE = 1 IF FEMALE
FEMALE = 0 IF MALE

MALE IS BENCHMARK GROUP

VS. FEMALE = 0 IF FEMALE
FEMALE = 1 IF MALE

THE GROUP THAT WE ASSIGN "0" IS CALLED "BENCHMARK GROUP" OR "BASE GROUP"

THE GROUP THAT YOU MAKE A COMPARISON AGAINST IT

VS MALE = 1 IF MALE
MALE = 0 IF FEMALE

FEMALE IS USED AS "BENCHMARK GROUP"

EX

$$WAGE = \alpha_0 + \gamma_0 \, MALE + \beta_1 EDU + u_i$$

MALE = 1 IF MALE
MALE = 0 IF FEMALE

→ FEMALE IS USED AS BENCHMARK GROUP

$$E(WAGE \mid MALE = 1, EDU) = (\alpha_0 + \gamma_0) + \beta_1 EDU \quad \text{--- (1)}$$

$$E(WAGE \mid MALE = 0, EDU) = \alpha_0 + \beta_1 EDU \quad \text{--- (2)}$$

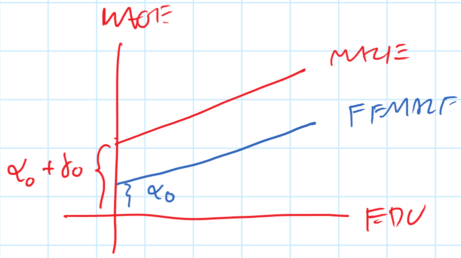
$$(1) - (2) \quad E(WAGE \mid MALE = 1, EDU) - E(WAGE \mid MALE = 0, EDU) = \gamma_0$$

$E(WAGE \mid MALE=0, EDU) = \alpha_0 + \beta_1 EDU$ —— (2)

(1) − (2) $E(WAGE \mid MALE=1, EDU) - E(WAGE \mid MALE=0, EDU) = \gamma_0$!

INTERCEPT FOR MALE'S REGRESSION LINE

" ———— " FEMALE'S REGRESSION LINE



Page 4

Tuesday, April 05, 2016 6:39 AM

Now, we added more variables into the wage model. Taking males as the base group, a model that controls for experience and tenure in addition to education is

$$wage_i = \beta_0 + \delta_0 \text{female} + \beta_1 \text{edu} + \beta_2 \text{exper} + \beta_3 \text{tenure} + u_i$$

If edu, expert, and tenure are all relevant productivity characteristics, the null hypothesis of no difference between men and women (No wage discrimination) is:

$H_0: \delta = 0$ (NO WAGE DISCRIMINATION)

$H_1: \delta < 0$ (WAGE DISCRIMINATION)

In table 9.1, it represents the partial listing of the sample data of wage model. We see that Person 1 is female, Person 2 is female , Person 3 is male, and so on.

Table 9.1: A Partial Listing of the Wage Data.

Data Editor (Browse) - [WAGE1]

File Edit View Data Tools

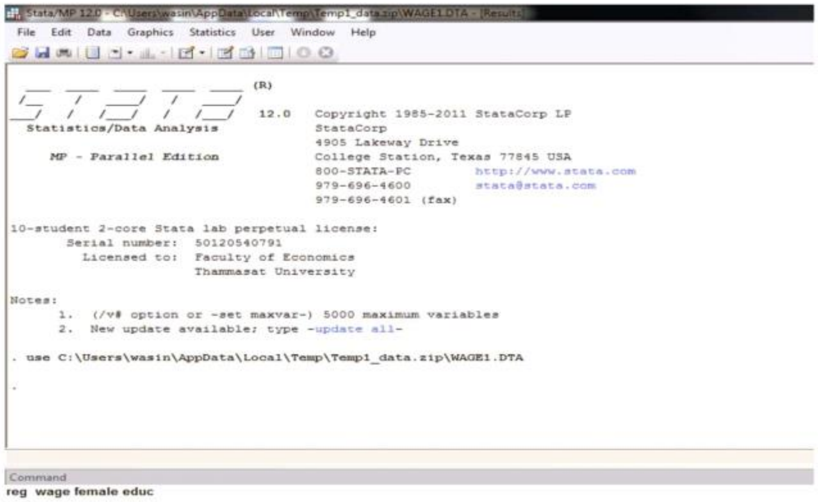
wage[1] 3.0999999

	wage	educ	exper	tenure	female
1	3.1	11	2	0	1
2	3.2	12	22	2	1
3	3	11	2	0	0
4	6	8	44	28	0
5	5.3	12	7	2	0
6	8.8	16	9	8	0
7	11	18	15	7	0
8	5	12	5	3	1
9	3.6	12	26	4	1
10	18	17	22	21	0
11	6.3	16	8	2	1
12	8.1	13	3	0	1
13	8.8	12	15	0	0
14	5.5	12	18	3	0
15	22	12	31	15	0
16	17	16	14	0	0
17	7.5	12	10	0	1
18	11	13	16	10	1
19	3.6	12	13	0	1
20	4.5	12	36	6	1
21	6.9	12	11	4	1
22	8.5	12	29	13	0
23	6.3	16	9	9	1
24	.53	12	3	1	1
25	6	11	37	8	1
26	9.6	16	3	3	0
27	7.8	16	11	10	0
28	13	16	31	0	0
29	13	15	30	0	0
30	3.3	8	9	1	1
31	13	14	23	5	0
32	4.5	14	2	5	1
33	9.7	13	16	16	1

Page 5

Tuesday, April 05, 2016 6:39 AM

Table 9.2: The command function to estimate the wage model in STATA program



Page 6

Tuesday, April 05, 2016 6:39 AM

Table 9.3: $wage_i = \beta_0 + \delta_0 \textbf{female} + \beta_1 \textbf{edu} + \beta_2 \textbf{exper} + \beta_3 \textbf{tenure} + u_i$

```
. reg wage female educ exper tenure

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  4,   521) =   74.40
       Model |  2603.10658     4  650.776644           Prob > F      =  0.0000
    Residual |  4557.30771   521   8.7472317           R-squared     =  0.3635
-------------+------------------------------           Adj R-squared =  0.3587
       Total |  7160.41429   525  13.6388844           Root MSE      =  2.9576

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -1.810852   .2648252    -6.84   0.000    -2.331109   -1.290596
        educ |   .5715048   .0493373    11.58   0.000     .4745802    .6684293
       exper |   .0253959   .0115694     2.20   0.029     .0026674    .0481243
      tenure |   .1410051   .0211617     6.66   0.000     .0994323    .1825778
       _cons |  -1.567939   .7245511    -2.16   0.031    -2.991339    -.144538
------------------------------------------------------------------------------
```

Table 9.4: $wage_i = \beta_0 + \delta_0 \textbf{female} + u_i$

```
reg wage female

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  1,   524) =   68.54
       Model |  828.220467     1  828.220467           Prob > F      =  0.0000
    Residual |  6332.19382   524  12.0843394           R-squared     =  0.1157
-------------+------------------------------           Adj R-squared =  0.1140
       Total |  7160.41429   525  13.6388844           Root MSE      =  3.4763

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -2.51183   .3034092    -8.28   0.000    -3.107878   -1.915782
       _cons |   7.099489   .2100082    33.81   0.000     6.686928     7.51205
------------------------------------------------------------------------------
```

Example: the Hourly Wage Equation:

$\hat{\beta}_0$ $\hat{\delta}_0$

$$\widehat{\text{wage}} = -1.5679 - 1.8109 \text{ female} + 0.5715 \text{ edu} + 0.025 \text{ exper} + 0.141 \text{ tenure}$$
$$= (0.7246) \quad (0.2648) \quad (0.0493) \quad (0.0116) \quad (0.0212) \qquad \text{(Eq.32)}$$

MODEL 1

$R^2 = 0.3635 \quad n = 526$

Interpret the model:

The intercept: -1.5679 IS NOT SO MEANINGFUL AS NO ONE HAS ZERO OR NEGATIVE WAGES IN THE SAMPLE.

The coefficient on female -1.8109 => SHOWS THE AVERAGE DIFFERENCE IN HOURLY WAGES BETWEEN MEN AND WOMEN, WITH THE SAME LEVEL OF EDUCATION, EXPERIENCE, AND TENURE!

[EDU, EXPER, TENURE ARE CONTROLLED VARIABLES]

COMPARED W/ MEN (BASE GROUP), WOMEN RECEIVE 1.81 $ LESS PER HOUR THAN MEN, ON AVERAGE.

It is informative to compare the coefficient on female in the above equation to the estimate we get when all other explanatory variables are dropped from the equation:

$$\widehat{\text{wage}} = 7.0995 - 2.5118 \text{ female}$$
$$se = (0.2100) \quad (0.3034) \qquad \text{(Eq.33)}$$

MODEL 2

$R^2 = 0.1157 \quad n = 526$

-2.51183 = AVERAGE WAGE DIFFERENTIALS BETWEEN MALES AND FEMALES

$E(\text{WAGE} \mid \text{FEMALE}=1) = 7.0995 - 2.51183 \cdot (1)$

$= 4.59$ \$ PER HOUR

$E(\text{WAGE} \mid \text{FEMALE}=0) = 7.0995 - 2.51183 \cdot (0)$
(MALE)

$= 7.0995$ \$ PER HOUR

SO $E(\text{WAGE} \mid \text{FEMALE}=1) - E(\text{WAGE} \mid \text{FEMALE}=0) = 4.59 - 7.0995$

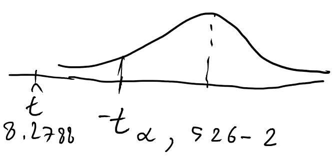
$= -2.51183$ \$ PER HOUR

LET'S DO HYPOTHESIS TESTING

$H_0: \delta_0 = 0$ (NO WAGE DISCRIMINATION)

$H_1: \delta < 0$ (WAGE DISCRIMINATION)

- COMPUTE $\hat{t}$: $\hat{t} = \dfrac{\hat{\delta}_0 - \delta_0}{SD} = -\dfrac{2.51183 - 0}{0.3034} = -8.2788$
- FIND $t_{CRITICAL}$: $t_{CRITICAL} = t_{0.05, 524} = 1.645$



- COMPARE $\hat{t}$ w/ $t_{CRITICAL}$: SINCE $-\hat{t} < -t_{CRITICAL}$,

THEN WE REJECT THE NULL HYPOTHESIS.

WAGE DIFFERENTIAL BET. MALES AND FEMALES IS STATISTICALLY SIGNIFICANT AT 95%

IN PARTICULAR, COMPARED W/ MALES, FEMALES EARN 2.51 \$ LESS, ON AVERAGE.

MODEL 1

WAGE = f(FEMALE, EDU, EXPER, TENURE)

$\hat{\delta}_0 = -1.81$

MODEL (I) IS THEN MORE RELIABLE

MODEL2

WAGE = f(FEMALE)

$\hat{\delta} = -2.51$

HIGHER B/C WE DO NOT CONTROL FOR DIFFERENCES IN EDU, EXPER, TENURE