9. Dummy Variable Regression Models

= Regression model with qualitative variable(s).

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.

9.1 Describing Qualitative Information

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

Example:

[1] A person is female or male ~~[illegible]~~

FEMALE = 1 if ♀
= 0 if ♂

[2] A firm offers a certain kind of employee pension plan or it does not.

PENSION = 0 if pension is NOT provided
= 1 if pension is provided

[3] A farm is located nearby the dam or not.

DAM < 1, 0

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.

Usually, we mostly use w/ cross-sectional data

i	GPA	EXPENDITURE	INCOME	READING HR.	LAPTOP OR COMPUTER	FEMALE
1					0	1
2					1	0
3					1	0
1					1	0
⋮					0	1
					1	0
					0	1
					1	0
545					⋮	1

Page 2

Tuesday, April 10, 2018 7:55 AM

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.

Page 3

Tuesday, April 10, 2018 7:55 AM

9.1.1 A Single Dummy Independent Variable

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

$$wage_i = \beta_0 + \delta_0 \text{female} + \beta_1 \text{edu} + u_i$$

FEMALE = 1 IF FEMALE

= 0 IF MALE

δ_0 will reflect the wage differential between male and female, given the same level of education.

$E[\text{WAGE} \mid \text{FEMALE}=1, \text{EDU}] = \beta_0 + \delta_0 \cdot 1 + \beta_1 \text{edu} = (\beta_0+\delta_0) + \beta_1 \text{edu}.$ — ①

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

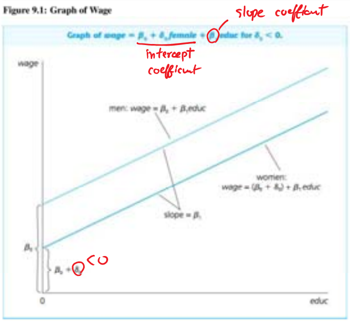
① - ② gives $E[\text{WAGE} \mid \text{FEMALE}=1, \text{EDU}] - E[\text{WAGE} \mid \overset{(=\text{male})}{\text{FEMALE}=0}, \text{EDU}]$

$= \delta_0$

If $\delta_0 < 0$, then, on average, female earns les than male. Wage discrimination exists!

154 Chapter 9. Dummy Variable Regression Models

Figure 9.1: Graph of Wage



WE SET FEMALE = 1 IF FEMALE
FEMALE = 0 IF MALE

The catagory that we assign value of "0" is called "Base group" or "Base catagory" or "reference group"

When we report our result, we compare other group/catagory w/ the base group.

(m = 2)

Ex: we say "Female earn lower wage on averge than "the male group"", given level of education.

OUR BASE GROUP

Example 2 5 geographic regions

- N & NE → D_1
- CENTRAL → D_2
- EAST → D_3
- WEST → D_4
- SOUTH → D_5

(m = 5)
↑
# of catagories

D_1 = 1 IF N & NE
0 OTHERWISE

D_2 = 1 IF CENTRAL
= 0 OTHER WISE

D_3 = 1 IF EAST
0 OTHERWISE

D_4 = 1 IF WEST
= 0 OTHERWISE

D_5 = 1 IF SOUTH
= 0 OTHERWISE

$$\text{TEACHER'S WAGE}_i = \beta_0 + \beta_1 D_1 + \beta_2 D_3 + \beta_3 D_4 + \beta_4 D_5 + u_i$$

DATA SET LOOKS LIKE THIS:

PROVINCES (i)	N&NE D_1	EAST D_3	WEST D_4	SOUTH D_5
CHIANG MAI	1	0	0	0
MAE HONG SON	1	0	0	0
NONG KAI	1	0	0	0
UDON THANI	1	0	0	0
CHANTABURI	0	1	0	0
TRAD	0	1	0	0
RAYONG	0	1	0	0
TAK	0	0	1	0
PRACHUAB	0	0	0	1
PECTHBURI	0	0	0	1
PHUKET	0	0	0	1
RANONG	0	0	0	1
BANGKOK	0	0	0	0
NONTABURI	0	0	0	0
SAMUTPRAKARN	0	0	0	0

CENTRAL

NOTE: WHEN $D_1 = 0, D_3 = 0, D_4 = 0, D_5 = 0$ → CENTRAL PROVINCES

* When # of catagories = m, the maximum numbers of dummy variable = m − 1 variables.

If you assign # of dummy varibles = # of catagories = m, we will encounter "dummy variable trap" due to we will encounter "dummy variable trap" due to perfect multicollinearity. As a result, model cannot be estimated!

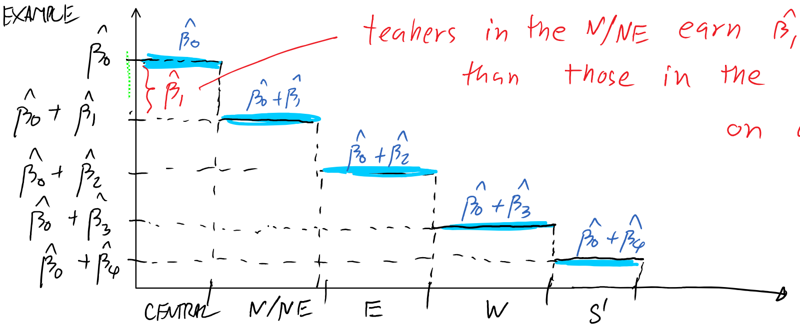
N/NE, EAST, WEST, SOUTH

$$\text{Teacher's wage} = \hat{\beta}_0 + \hat{\beta}_1 D_1 + \hat{\beta}_2 D_3 + \hat{\beta}_3 D_4 + \hat{\beta}_4 D_5$$

Base group

- $E(\text{wage} \mid D_1 = D_3 = D_4 = D_5 = 0) = \hat{\beta}_0$ → Teacher's wage in central region.
- $E(\text{wage} \mid \underbrace{D_1 = 1}_{N/NE}, D_3 = D_4 = D_5 = 0) = \hat{\beta}_0 + \hat{\beta}_1$ " ——— " N/NE.
- $E(\text{wage} \mid \underbrace{D_3 = 1}_{EAST}, D_1 = D_4 = D_5 = 0) = \hat{\beta}_0 + \hat{\beta}_2$ " ——— " EAST
- $E(\text{wage} \mid \underbrace{D_4 = 1}_{WEST}, D_1 = D_3 = D_5 = 0) = \hat{\beta}_0 + \hat{\beta}_3$ " ——— " WEST
- $E(\text{wage} \mid D_5 = 1, D_1 = D_3 = D_4 = 0) = \hat{\beta}_0 + \hat{\beta}_4$ " ——— " SOUTH

FOR EXAMPLE



teachers in the N/NE earn $\hat{\beta}_1$ baht lower than those in the central region, on average.

Page 5

Tuesday, April 10, 2018 7:55 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: $\delta_0 \geq 0$ (no wage discrimination)

H_1: $\delta_0 < 0$ (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

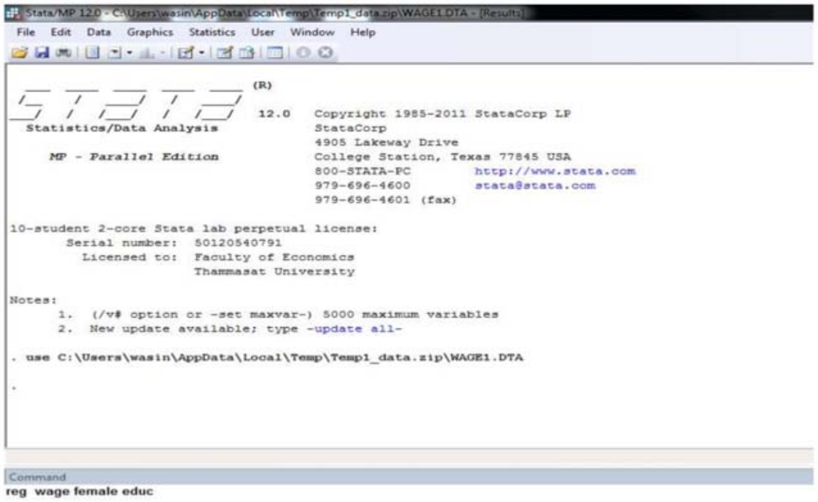
wage[1] 3.0999999

	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 6

Tuesday, April 10, 2018 7:55 AM

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



Page 7

Tuesday, April 10, 2018 7:55 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.[illegible]2317           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
------------------------------------------------------------------------------
```

ESS

RSS

TSS

$\hat{t} = \frac{\text{Coeff}}{SE}$

Example: the Hourly Wage Equation:

they are controlled variables

$$\widehat{wage} = -1.5679 - 1.8109\ \text{female} + 0.5715\ \text{edu} + 0.025\ \text{exper} + 0.141\ \text{tenure}$$
$$\text{s.e.} \quad = (0.7246) \quad (0.2648) \quad (0.0493) \quad (0.0116) \quad (0.0212) \qquad (9.1)$$

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

Interpret the model:

The intercept: $-1.5679 \rightarrow$ not so meaningful since nobody has zero values for all of edu, exper, tenure in the sample.

$\hat{\delta}_0 < 0$

$-1.8109 \rightarrow$ statistically significant at 99%.

$\downarrow$ shows the average difference in hourly wage between male and female, given the same level of education, experience, and tenure!

Female earn, on average, 1.8109 \$ less per hour

Female earn, on average, 1.8109 $ less per hour than male.

Note: The data set was in 1976

Page 8

Tuesday, April 10, 2018 7:55 AM

158 Chapter 9. Dummy Variable Regression Models

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
------------------------------------------------------------------------------
```

$\hat{\delta}_0$

Recall that female = 1 if ♀

= 0 if ♂

7.0995 = Hourly wages for male, on average

$E(wage \mid female = 0) = 7.0995$ \$

The coefficient on female

−2.51183 = average wage differentials between male and female

So, average wage for female in the sample = 7.10 − 2.51 = 4.59 \$ per hour.

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:

$$\widehat{wage} = 7.0995 - 2.5118\ \text{female}$$
$$se = (0.2100) \quad (0.3034) \qquad (9.2)$$

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

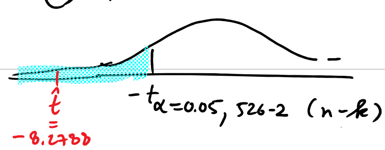
Hypothesis Testing

① $H_0: \delta_0 \geq 0$

$H_1: \delta_0 < 0$

② $\hat{t} = \dfrac{-2.5118}{0.3034} = -8.2788$

③ Find critical t



④ Since $\hat{t} = -8.2788 < -\hat{t}_{CRITICAL}$, then we can reject H_0.

Wage differential between the two groups is statistically significant. In particular, Female earn less than male on average by 2.5 \$ / hour.

\# Let's compare the two models:

(9.1) vs. (9.2)

$wage = f(\text{female}, \text{edu}, \text{exper}, \text{tenure})$ | $wage = f(\text{female})$

$\hat{\delta}_0 = -1.8$ < $\hat{\delta}_0 = -2.5$ (WHY?)

B/c 9.2 does not "control" for the differences . . . .

in education, experience, and tenure!

that's why larger wage differential may come from those differences mentioned above.

9.1 gives a more reliable estimate for what we called "the ceteris paribus gender wage gap".

Once we hold edu, exper, tenure FIXED, it still reveals a large wage differential of -1.81 \$/hr.

Page 9

Tuesday, April 10, 2018 7:55 AM

Page 10

Tuesday, April 10, 2018 7:55 AM

9.2 Interpreting Coefficients on Dummy Explanatory Variables When the Dependent Variable is log(y)

In this section, we will study a model that has the dependent variable appearing in logarithmic form, with one or more dummy variables approving as independent variables.
Question: How do we interpret the dummy variable coefficients in this case?

Answer: Not surprisingly, the coefficients have a percentage interpretation.

Let us reestimate the wage equation , using log (wage) as the dependent variable and adding quadratics in expert and tenure:

$$log(wage_i) = \beta_0 + \delta_0 female + \beta_1 edu + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + u_i$$

The Stata result is shown in table 9.5.

Table 9.5:

```
reg lwage female educ exper expersq tenure tenursq
```

Source	SS	df	MS
Model	65.3791009	6	10.8965168
Residual	82.9506505	519	.159827843
Total	148.329751	525	.28253286

Number of obs	= 526
F(6, 519)	= 68.18
Prob > F	= 0.0000
R-squared	= 0.4408
Adj R-squared	= 0.4343
Root MSE	= .39978

$\hat{\delta}_0$

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
female	-.296511	.0358055	-8.28	0.000	-.3668524	-.2261696
educ	.0801967	.0067573	11.87	0.000	.0669217	.0934716
exper	.0294324	.0049752	5.92	0.000	.0196585	.0392063
expersq	-.0005827	.0001073	-5.43	0.000	-.0007935	-.0003719
tenure	.0317139	.0068452	4.63	0.000	.0182663	.0451616
tenursq	-.0005852	.0002347	-2.49	0.013	-.0010463	-.0001241
_cons	.416691	.0989279	4.21	0.000	.2223425	.6110394

$\log(wage) = f(\text{female})$

$\hat{\delta}_0 > |0.2965|$ as we expected.

Example: Log Hourly Wage Equation:

$$\widehat{\log(\text{wage})} = \underset{(0.0989)}{0.4167} - \underset{(0.0358)}{0.2965}\ \text{female} + \underset{(0.0068)}{0.0802}\ \text{edu} + \underset{(0.0050)}{0.0294}\ exper - \underset{(0.0001)}{0.0006}\ exper^2 + \underset{(0.0068)}{0.0317}\ tenure - \underset{(0.0002)}{0.0006}\ tenure^2 \quad (9.3)$$

$R^2 = 0.4408 \quad n = 526$

Recall this

$\ln y = \beta_0 + \beta_1 x$

$\frac{d \ln y}{dx} = \beta_1$

$\frac{1}{y}\frac{dy}{dx} = \beta_1$

$\frac{\frac{dy}{y} \times 100}{dx} = \beta_1 \times 100$

$\boxed{\frac{\% \Delta y}{\Delta x} = \beta_1 \times 100}$

Interpret the model:

The coefficient on female -0.2965

For the same level of education, years of experience, and tenure, female earns $100 \cdot (0.2965) = 29.65\%$ less than male. (approximately)

The above is just an approximation of the wage differential. We can have a precise calculation by computing "the exact percentage difference" in predicted wages:

Holding other factors constant, we know that

$$\widehat{\log(wage_F)} - \widehat{\log(wage_M)} = -0.2965$$

$$\log\left(\frac{\widehat{wage_F}}{wage_M}\right) = -0.2965$$

$$\frac{\widehat{wage_F} - \widehat{wage_M}}{\widehat{wage_M}} = EXP(-0.2965) - 1 \approx -0.257$$

More accurate estimate says that female's wage is, on average, $0.257 \cdot 100 = 25.7\%$ below a comparable male's wage!!!

♀ Vs. ♂

FEMALE MALE

same edu, exper, tenure

Page 12

Tuesday, April 10, 2018 7:55 AM

9.3 Using Dummy Variables for Multiple Categories

We can use several dummy independent variables in the same equation. For example, we could add the dummy variable **married** to the wage model.

The previous model:

$$log(wage_i) = \beta_0 + \delta_0 female + \beta_1 edu + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + u_i$$

Now, Let us estimate a model that allows for wage differences among four groups:

[1.] Married Men

[2] Married Women

Page 13

Tuesday, April 10, 2018 7:55 AM

[3] Single Men

[4] Single Women

Page 14

Tuesday, April 10, 2018 7:55 AM

To do this, we must select a base group:

Let us use " Single men " as our base group

		MARRIED 0	MARRIED 1
FEMALE	0	OUR BASE GROUP (0,0) single men	(0,1) married males
	1	(1,0) single females	(1,1) married females

NOTE: (GENDER, MARITIAL STATUS)

Now, we need to define dummy variables for each of the remaining groups.

(0,1) → Married Males → MARRMALE

(1,0) → Single Females → SINGFEM

(1,1) → Married Females → MARR FEM

m = number of catagories.

Here, m = 4

Then maximum number of dummy variables we can put into the model = m − 1 = 4 − 1 = 3 variables

WARNING! : If we add " single male" into this current model, we will encounter " perfect collinearity " and the model cannot be estimated!

This event is called " Dummy Variable Trap "

Therefore, our model is:

$$log(wage_i) = \beta_0 + \delta_0 marrmale + \delta_1 marrfem + \delta_2 singfem +$$
$$\beta_1 edu + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + u_i$$

i	marrmale	marrfem	sing fem	
1	0	0	0	→ single male
2	1	0	0	→ married male
3	0	0	1	→ single female
4	0	1	0	→ married female.
5				
⋮				

We of course drop the dummy variable (female). (Why?)

Page 15

Tuesday, April 10, 2018 7:55 AM

Table 9.5:

```
reg lwage marrmale marrfem singfem educ exper expersq tenure tenursq
```

Source	SS	df	MS
Model	68.3617623	8	8.54522029
Residual	79.9679891	517	.154676961
Total	148.329751	525	.28253286

Number of obs =	526
F(8, 517) =	55.25
Prob > F =	0.0000
R-squared =	0.4609
Adj R-squared =	0.4525
Root MSE =	.39329

BASE GROUP is SINGLE MALE (0,0)

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
marrmale	.2126757	.0553572	3.84	0.000	.103923	.3214284
marrfem	-.1982676	.0578355	-3.43	0.001	-.311889	-.0846462
singfem	-.1103502	.0557421	-1.98	0.048	-.219859	-.0008414
educ	.0789103	.0066945	11.79	0.000	.0657585	.092062
exper	.0268006	.0052428	5.11	0.000	.0165007	.0371005
expersq	-.0005352	.0001104	-4.85	0.000	-.0007522	-.0003183
tenure	.0290875	.006762	4.30	0.000	.0158031	.0423719
tenursq	-.0005331	.0002312	-2.31	0.022	-.0009874	-.0000789
_cons	.3213781	.100009	3.21	0.001	.1249041	.5178521

$$\begin{aligned}
\widehat{\log(wage)} &= 0.3214 + 0.2127\ marrmale - 0.1983\ marrfem - 0.1104\ singfem \\
\text{s.e} \leftarrow\ & (0.1000) \quad (0.0554) \quad (0.0578) \quad (0.0557) \\
& + 0.0789\ edu + 0.0268\ exper - 0.0005\ exper^2 + 0.0291\ tenure - 0.0005\ tenure^2 \\
\text{s.e} \leftarrow\ & (0.0067) \quad (0.0268) \quad (0.0001) \quad (0.0068) + \quad (0.0002)
\end{aligned} \tag{9.4}$$

$R^2 = 0.4609 \quad n = 526$

- +0.2127 FOR MARRMALE (1,0,0)
  w/ the same qualifications, married males are expected to earn 21.27% higher than single males (OUR base group)
- −0.1983 FOR MARRFEM (0,1,0)
  w/ the same qualifications, married females are expected to earn 19.83%. lower than single males.
- −0.1104 FOR single females (0,0,1)
  w/ the same qualifications, single females are expected to earn 11.04%. lower than single males

Q; How about wage differential between

"SINGLE FEMALE" and "MARRIED FEMALE"?

" SINGLE FEMALES" and " MARRIED FEMALE"?

marrfem → -0.1983

singfem → -0.1104

$\Delta \hat{W} = 0.1983 - 0.1104 = 0.0879$

♀ + married   ♀ + single

Single females earn about $0.0879 \times 100 = 8.8\%$ higher than married females, on average

Page 16

Tuesday, April 10, 2018 7:55 AM

Interpret the model:

Page 17

Tuesday, April 10, 2018 7:55 AM

9.3.1 Interactions Involving Dummy Variables

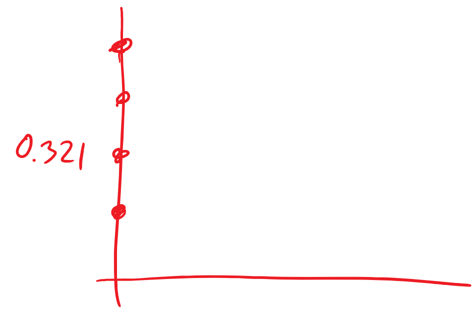
8.5.1 The Interactions Among Dummy Variables:

We can recast the model by adding an **interaction term** between female and married to the model where female and married appear separately. This allows the marriage premium to depend on gender. The estimated model with the female-married interaction term is :

$$\begin{aligned}\widehat{\log(wage)} &= \underset{(0.100)}{0.321} - \underset{(0.056)}{0.110}\ female + \underset{(0.055)}{0.213}\ married \\ &\quad + \underset{(0.072)}{0.301}\ female \cdot married + \ldots,\end{aligned} \tag{9.5}$$

let's consider the output :

- Intercept for single males = 0.321
- Intercept for single females = 0.321 − 0.110(1) + 0.213(0) + 0.301(1·0)
  = 0.321 − 0.110
  = +0.211
- Intercept for married males = 0.321 + 0.213(1)
  = +0.534
- Intercept for married females = 0.321 − 0.110(1) + 0.213(1) + 0.301(1·1)
  = +0.725



8.5.2 The interaction between Dummy Variable/s and Explanatory Variable/s: the Allowing for the Different Slopes

There are also occasions for interacting dummy variables with explanatory variables that are not dummy variables to allow for a **difference in slope.**

To see the interaction between female and edu, we can rewrite the model as follow:

$$wage_i = \beta_0 + \delta_0 \text{female} + \beta_1 \text{edu} + \underline{\delta_1 \textit{female} \cdot \textit{edu}} + u_i$$

Men Group we plug female =0

Therefore: $wage = \beta_0 + \beta_1 edu + u_i$ → FOR MEN

reflects " Return of Education "
for male (β_1)
and for female ($\beta_1 + \delta_1$)

Women Group we plug female =1

Therefore: $wage = \beta_0 + \delta_0 (\cdot 1) + \beta_1 edu + \delta_1 (1 \cdot edu) + u_i$

$= (\beta_0 + \delta_0) + (\beta_1 + \delta_1) edu + u_i$

δ_0 captures the difference in "intercept" between men and women groups.

δ_1 capture the difference in "slope" between men and women groups

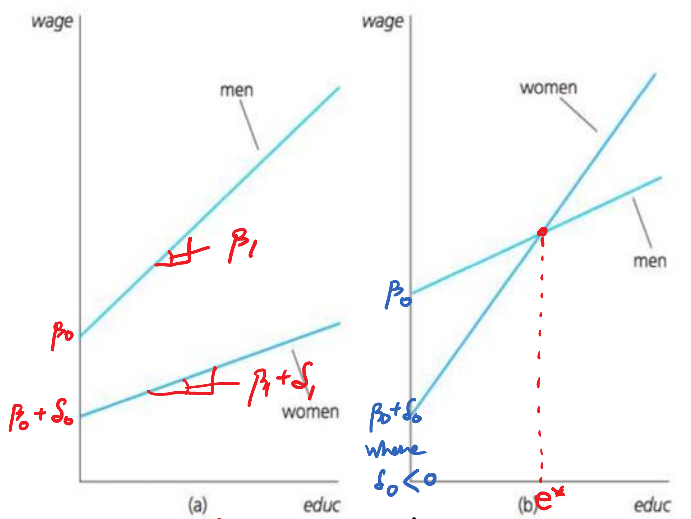
$\boxed{H_0: \delta_0 = 0 \quad H_1: \delta_1 \neq 0}$ and $\boxed{H_0: \delta_1 = 0 \quad H_1: \delta_1 \neq 0}$

Page 20

Tuesday, April 10, 2018 7:55 AM

Figure 9.2: Graph of the Wage Model with an Interaction between female and education

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



- For any level of edu, women earn less than men, on average.
- The higher the education, the larger the wage gap between the two groups.

for $e < e^*$, men wage > women wage

+

The larger the edu level, the narrower the wage gap.

for $e > e^*$, women wage > men wage

+

The larger the edu level, the wider the wage gap.

	δ_0	δ_1
(c)	>0	>0
(d)	>0	<0

DIY: visualize it

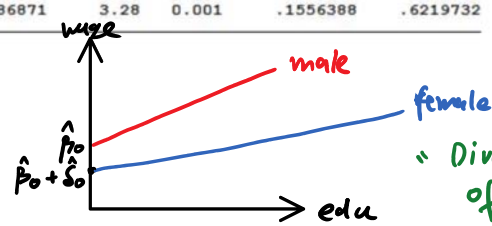
Example

Table 9.5:

```
gen femed = female*educ

 reg lwage female educ femed exper expersq tenure tenursq

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  7,   518) =   58.37
       Model |  65.4081534     7  9.34402192           Prob > F      =  0.0000
    Residual |  82.921598    518  .160080305           R-squared     =  0.4410
-------------+------------------------------           Adj R-squared =  0.4334
       Total |  148.329751   525  .28253286            Root MSE      =   .4001

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.2267886   .1675394    -1.35   0.176    -.5559289    .1023517
        educ |   .0823692   .0084699     9.72   0.000     .0657296    .0990088
       femed |  -.0055645   .0130618    -0.43   0.670    -.0312252    .0200962
       exper |   .0293366   .0049842     5.89   0.000      .019545    .0391283
     expersq |  -.0005804   .0001075    -5.40   0.000    -.0007916   -.0003691
      tenure |   .0318967    .006864     4.65   0.000      .018412    .0453814
     tenursq |   -.00059    .0002352    -2.51   0.012    -.001052    -.000128
       _cons |   .388806    .1186871     3.28   0.001     .1556388    .6219732
------------------------------------------------------------------------------
```

$\hat{\delta}_0$ $\hat{\beta}_1$ $\hat{\delta}_1$ $\hat{\beta}_0$

i	female	· edu	femed
1	0	17	
2	1	21	
3	0	16	
4	0	12	
	1		
	0		
	0		
	1		
	0		



" Divergence of wage gap

$$\widehat{\log(\text{wage})} = \underset{(0.1187)}{0.3889} - \underset{(0.1675)}{0.2268}\ \text{female} + \underset{(0.0085)}{\overset{\hat{\beta}_1}{0.082}}\ \text{edu} \underset{(0.0131)}{\overset{\hat{\delta}_1}{- 0.0056}}\ female \cdot edu$$
$$+\underset{(0.0050)}{0.0293}\ exper - \underset{(0.0001)}{0.0006}\ exper^2 + \underset{(0.0069)}{0.0319}\ tenure + - \underset{(0.0002)}{0.00059}\ tenure^2 \qquad (9.6)$$

$R^2 = 0.4410 \quad n = 526$

$0.082 \times 100 = 8.2\,\%$

Given same qualifications (experience and tenure), 1 year increase in education would increase male wage by 8.2% on average.

$\hat{\beta}_1 + \hat{\delta}_1 = 0.082 - 0.0056 = 0.0764$

$0.0764 \times 100 = 7.64\%$ → Estimated Return of Female group.

Q: Is there any significance in the difference in ROE bet. the two group?

A: do hypothesis testing.

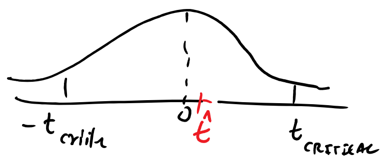
$H_0: \delta_1 = 0$ (No difference in ROE)

$H_1: \delta_1 \neq 0$ (Difference in ROE exists)

Models $\log(\text{wage}) = \beta_0 + \beta_1 \text{edu} + \cdots + u_i$ → For men

$\log(\text{wage}) = \beta_0 + \delta_0 + (\beta_1 + \delta_1)\text{edu} + \cdots + u_i$ → For Women

$$\hat{t} = \frac{-0.0056}{0.0131} = -0.43$$



Since $-t_{critical} < \hat{t} < t_{critical}$,

then we fail to reject $H_0: \delta_0 = 0$.

Therefore, there is No significant difference in ROE wage bet. the two groups.

$H_0: \delta_0 = 0$

$H_1: \delta_0 \neq 0$

Fail to reject $H_0: \delta = 0$