

5 Using dummy variables for multiple categories

Case 1 We can use many dummy variables in the same model

Consider a model which includes 2 dummy variables— *female* and *married*.

$$\log(\text{wage}) = \beta_0 + \delta_0 \text{female} + \delta_1 \text{married} + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + \beta_4 \text{tenure} + \beta_5 \text{tenure}^2 + u$$

1 if female 1 if married
 0 if otherwise 0 if otherwise

```
regress lwage female married educ exper expersq tenure tenursq
```

Source	SS	df	MS	Number of obs = 526	
Model	65.6482326	7	9.37831895	F(7, 518) =	58.76
Residual	82.6815188	518	.159616832	Prob > F =	0.0000
				R-squared =	0.4426
				Adj R-squared =	0.4351
Total	148.329751	525	.28253286	Root MSE =	.39952

y variable

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
female	-.2901838	.0361121	-8.04	0.000	-.3611279	-.2192396
married	.0529219	.0407561	1.30	0.195	-.0271456	.1329894
educ	.0791547	.0068003	11.64	0.000	.0657952	.0925143
exper	.0269535	.0053258	5.06	0.000	.0164907	.0374163
expersq	-.0005399	.0001122	-4.81	0.000	-.0007603	-.0003196
tenure	.0312962	.0068482	4.57	0.000	.0178426	.0447499
tenursq	-.0005744	.0002347	-2.45	0.015	-.0010355	-.0001134
_cons	.4177837	.0988662	4.23	0.000	.2235557	.6120116

Aj. ⇒ But since, $|t| < 1.96$ or $P > 0.05$, we do not reject H_0 of no impact

$1.96 > 1.30$ $0.05 < 0.195$
 ∴ we do not reject null hypothesis of $\beta = 0$
 ; marriage doesn't have impact on wage

Comments:

1.) δ_0 measure the expected difference between female & male workers given the same marital status and other factors.

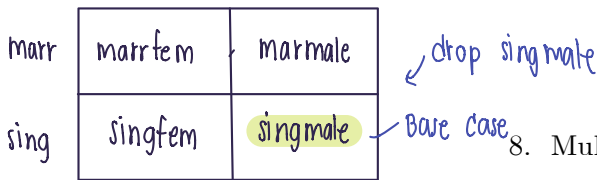
$$\frac{\partial \log(\text{wage})}{\partial \text{female}} = \frac{1}{\text{wage}} \frac{\partial \text{wage}}{\partial \text{female}} = -0.29$$

$$100 \cdot \frac{1}{\text{wage}} \frac{\partial \text{wage}}{\partial \text{female}} = 100 \cdot -0.29$$

$$\frac{\% \Delta \text{wage}}{\partial \text{female}} = 29.02\%$$

• female workers are expected to earn less than male worker by 29.02%, holding other factor the same

2.) δ_1 measures the impact of be married (marriage premium)



Consider a model which includes dummy variables for each gender/marital status combination— *marmale*, *marrfem* and *singfem*. or *singmale* ← used as the base case

$$\log(\text{wage}) = \beta_0 + \delta_0 \text{marmale} + \delta_1 \text{marrfem} + \delta_3 \text{singfem} + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + \beta_4 \text{tenure} + \beta_5 \text{tenure}^2 + u. \quad (8.1)$$

```
regress lwage marmale marrfem singfem educ exper expersq tenure tenursq
```

Source	SS	df	MS	Number of obs =	526
Model	68.3617623	8	8.54522029	F(8, 517) =	55.25
Residual	79.9679891	517	.154676961	Prob > F	= 0.0000
Total	148.329751	525	.28253286	R-squared	= 0.4609
				Adj R-squared	= 0.4525
				Root MSE	= .39329

dummy V.
same as ↑ case except for dummy V.

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
marmale	.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

Comments:

This regression is not the same as the previous one. It uses "single male" as the base group. (The previous one use male & single as 2 base groups.)

- δ_0 measure the expected different in wage of married male as compared with single males, holding other factors constant.
- δ_1 measure married female as compared with single males, holding other factors constant.
- δ_2 same rationale.

intercept

