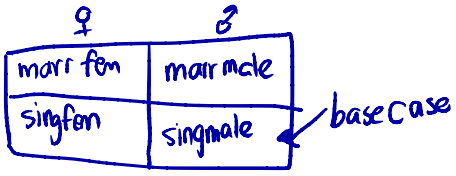


Consider a model which includes dummy variables for each gender/marital status combination– $marrmale$, $marrfem$ and $singfem$.

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

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

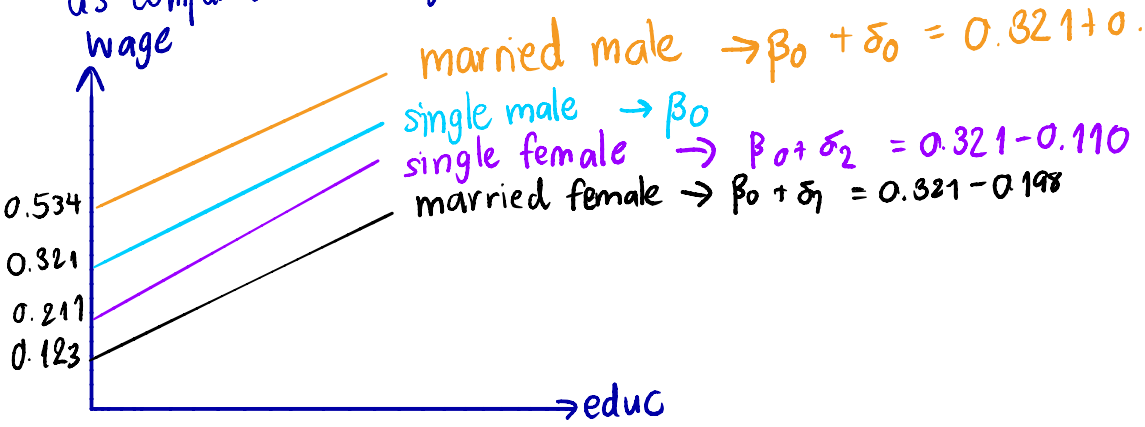
Source	SS	df	MS		
Model	68.3617623	8	8.54522029	Number of obs	= 526
Residual	79.9679891	517	.154676961	F(8, 517)	= 55.25
				Prob > F	= 0.0000
				R-squared	= 0.4609
				Adj R-squared	= 0.4525
Total	148.329751	525	.28253286	Root MSE	= .39329

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
δ_0 marrmale	.2126757	.0553572	3.84	0.000	.103923	.3214284
δ_1 marrfem	-.1982676	.0578355	-3.43	0.001	-.311889	-.0846462
δ_2 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 as it uses "single male" as base group (The previous one uses male & single as 2 base groups)

- δ_0 measures the expected diff in wage of married male as compared with single males, holding other factors constant

- δ_1 measures the expected diff in wage of married female as compared with single male



Case 2 We can use dummy variables to represent multiple categories of a variable Consider the relationship between law school rankings and starting salaries

$$\log(salary) = \beta_0 + \delta_0 top10 + \delta_1 r11\_25 + \delta_3 r26\_40 + \delta_4 r41\_60 + \beta_1 LSAT + \beta_2 GPA + \beta_3 \log(libvol) + \beta_4 \log(cost) + u.$$

* In many cases, "range of values" serve as a better explainatory variable than value itself

where $top10$, $r11\_25$, $r26\_40$, $r41\_60$ would be equal to 1 when the variable $rank$ falls into the appropriate range.

↳ e.g. age may explain the model better if split into generations e.g. gen x, y, z

** Rank below 60 would be the base case.

```
. regress lsalary top10 r11_25 r26_40 r41_60 LSAT GPA llibvol lcost
```

Source	SS	df	MS		
Model	9.16538532	8	1.14567316	Number of obs =	136
Residual	1.2109665	127	.009535169	F(8, 127) =	120.15
				Prob > F =	0.0000
Total	10.3763518	135	.076861865	R-squared =	0.8833
				Adj R-squared =	0.8759
				Root MSE =	.09765

lsalary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
top10	.5393428	.053542	10.07	0.000	.4333927	.6452928
r11_25	.4716199	.0390921	12.06	0.000	.3942637	.548976
r26_40	.2790977	.0346972	8.04	0.000	.2104383	.3477571
r41_60	.182382	.0283098	6.44	0.000	.126362	.238402
LSAT	.0060482	.0034919	1.73	0.086	-.0008616	.012958
GPA	.1305893	.0818678	1.60	0.113	-.0314122	.2925908
llibvol	.0725522	.0289213	2.51	0.013	.0153221	.1297824
lcost	.0249169	.0283224	0.88	0.381	-.031128	.0809619
_cons	8.363103	.4457314	18.76	0.000	7.481081	9.245125

baseline is ranking 61 and worse →

Comments:

rank	top 10	11-25	25-40
1	1	0	0
2	1	0	
3	1	0	
⋮	⋮	⋮	⋮
10	0	1	
11	0	1	
⋮	0	⋮	
25	0	0	
⋮	⋮	⋮	1
40			1

1) δ_0 measures the difference in expected log(salary) of a law school grad from a top-10 university compared to expected log(salary) of those who graduate from the school ranked 61th and worse

2) δ_1 ⇝ use the same rationale