

Multiple regression

Data source: Current Population Survey 2006.

wage average hourly earnings (in \$)
educ years of education
exper years potential experience
female 1=female, 0=male
nonwhite =1 if nonwhite
services =1 if in services industry
profocc =1 if in professional occupation
union =1 if respondent is union member

Summary statistics of the variables

. sum wage educ exper female nonwhite services profocc union

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	2000	18.34701	11.49495	.7	82.42857
educ	2000	13.633	2.0877	9	18
exper	2000	20.5885	12.75769	0	67
female	2000	.5165	.4998527	0	1
nonwhite	2000	.149	.3561775	0	1
services	2000	.152	.3591107	0	1
profocc	2000	.212	.4088271	0	1
union	2000	.139	.346033	0	1

Regression of log(wage) on education (with increasing number of other controls)

$$\widehat{\log(\text{wage})} = 1.19 + .114 \text{ educ} \quad R^2 = .17$$

(.08) (.006) n = 2000

$$\widehat{\log(\text{wage})} = 0.97 + .114 \text{ educ} + .011 \text{ exp} \quad R^2 = .22$$

(.08) (.005) (.0009) n = 2000

$$\widehat{\log(\text{wage})} = 1.06 + .117 \text{ educ} + .011 \text{ exp} - .25 \text{ female} \quad R^2 = .27$$

(.08) (.005) (.0009) (.02) n = 2000

$$\widehat{\log(\text{wage})} = 1.21 + .109 \text{ educ} + .010 \text{ exp} - .24 \text{ female} - .23 \text{ services} \quad R^2 = .29$$

(.08) (.005) (.0009) (.02) (.03) n = 2000

. g lwage=log(wage)
. reg lwage educ exper female services;

Source	SS	df	MS	Number of obs =	2000
Model	194.949697	4	48.7374243	F(4, 1995) =	203.57
Residual	477.62717	1995	.239412115	Prob > F =	0.0000
				R-squared =	0.2899
				Adj R-squared =	0.2884
Total	672.576867	1999	.336456662	Root MSE =	.4893

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.1086052	.005367	20.24	0.000	.0980798 .1191306
exper	.0103387	.0008616	12.00	0.000	.0086489 .0120284
female	-.2437118	.0219738	-11.09	0.000	-.2868059 -.2006178
services	-.2272342	.0313323	-7.25	0.000	-.2886816 -.1657868
_cons	1.20755	.0776504	15.55	0.000	1.055266 1.359834

$$\widehat{\log(\text{wage})} = 1.06 + .117 \text{educ} + .011 \text{exp} - .25 \text{female} \quad R^2 = .27$$

(.08) (.005) (.0009) (.02) n = 2000

Adding/omitting an irrelevant variable:

$$\widehat{\log(\text{wage})} = 1.06 + .117 \text{educ} + .011 \text{exp} - .25 \text{female} - .037 \text{nonwhite} \quad R^2 = .27$$

(.08) (.005) (.0009) (.02) (.031) n = 2000

Adding/omitting an important variable not correlated with the other independent variables:

$$\widehat{\log(\text{wage})} = 1.28 + .117 \text{educ} - .25 \text{female} \quad R^2 = .21$$

(.08) (.006) (.02) n = 2000

Adding/omitting an important variable correlated with the other independent variables:
Omitted variable bias

$$\widehat{\log(\text{wage})} = 1.17 + .106 \text{educ} + .011 \text{exp} - .26 \text{female} + .012 \text{profocc} \quad R^2 = .28$$

(.08) (.005) (.0009) (.02) (.03) n = 2000

$$\widehat{\log(\text{wage})} = 2.57 + .011 \text{exp} - .26 \text{female} + .358 \text{profocc} \quad R^2 = .16$$

(.03) (.0009) (.02) (.03) n = 2000

```
. corr lwage educ exp female profocc nonwhite
(obs=2000)
```

	lwage	educ	exper	female	profocc	nonwhite
lwage	1.0000					
educ	0.4097	1.0000				
exper	0.2358	0.0010	1.0000			
female	-0.1935	0.0489	0.0210	1.0000		
profocc	0.2181	0.4276	-0.0383	0.1077	1.0000	
nonwhite	-0.0379	-0.0051	-0.0200	0.0368	-0.0143	1.0000