

EXERCISE 5
KEY

Purpose: To learn more about multiple regression, specification bias, and tests of linear hypotheses. **This exercise is due on Tuesday, October 4.**

Work computer exercise C10 on page 100 and exercise C11 on page 148 in your textbook. I will take up both exercises.

C10 page 100

- (i) What is the range of the educ variable in the sample? Min: 6, Max: 20, Range = 14
 What percentage of men completed twelfth grade but no higher grade?
 512 for education level 12
 $(512)/1230 = 41.62\%$
 Do men or their parents have, on average, higher levels of education?
 Men do with a mean of 13.03 years of education compared to 12.17 for mothers and 12.44 for fathers.

(ii) $educ = B_0 + B_1motheduc + B_2fatheduc + u$
 $educ = 6.96 + .304motheduc + .190fatheduc + u$
 (0.319) (0.031) (0.022)

Source	SS	df	MS	Number of obs	=	1,230
Model	1697.9676	2	848.9838	F(2, 1227)	=	203.68
Residual	5114.31207	1,227	4.1681435	Prob > F	=	0.0000
				R-squared	=	0.2493
				Adj R-squared	=	0.2480
Total	6812.27967	1,229	5.54294522	Root MSE	=	2.0416

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
motheduc	.3041971	.0319266	9.53	0.000	.2415603 .366834
fatheduc	.1902858	.0222839	8.54	0.000	.1465669 .2340046
_cons	6.964355	.3198205	21.78	0.000	6.336899 7.59181

The sample variation explained by the mother's and father's education is computed by looking at the R² of the fitted model above. The answer is 24.9%.
 The coefficient on *motheduc* means that for every additional grade/year of education completed by the mother, the adult male's education will be 0.304 grades higher.

(iii) Add *abil* to the regression, report the results in equation form. Does ability help explain variations in education, even after controlling for parents' education? Explain.

Source	SS	df	MS	Number of obs	=	1,230
Model	2912.30705	3	970.769018	F(3, 1226)	=	305.17
Residual	3899.97262	1,226	3.18105434	Prob > F	=	0.0000
				R-squared	=	0.4275
				Adj R-squared	=	0.4261
Total	6812.27967	1,229	5.54294522	Root MSE	=	1.7836

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
motheduc	.1891314	.0285062	6.63	0.000	.1332051 .2450578
fatheduc	.1110854	.0198849	5.59	0.000	.0720733 .1500976
abil	.5024829	.025718	19.54	0.000	.4520268 .552939
_cons	8.44869	.2895407	29.18	0.000	7.88064 9.01674

$$\text{Equation Form: } Educ = 8.44 + .189\text{motheduc} + .111\text{fatheduc} + .502\text{abil} + u$$

$$(0.289) \quad (0.028) \quad (0.019) \quad (0.025)$$

Ability helps explains variations in education, as the absolute value of 19.52 is far greater than 2, thus indicating it is helpful for determining variation in the dependent variable, *educ*.

(iv)

Source	SS	df	MS	Number of obs	=	1,230
Model	3027.03706	4	756.759264	F(4, 1225)	=	244.91
Residual	3785.24262	1,225	3.08999397	Prob > F	=	0.0000
				R-squared	=	0.4444
				Adj R-squared	=	0.4425
Total	6812.27967	1,229	5.54294522	Root MSE	=	1.7578

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
motheduc	.1901261	.0280957	6.77	0.000	.1350051 .2452472
fatheduc	.1089387	.0196014	5.56	0.000	.0704827 .1473946
abil	.4014624	.0302875	13.26	0.000	.3420413 .4608835
abil2	.050599	.0083039	6.09	0.000	.0343076 .0668905
_cons	8.240226	.2874099	28.67	0.000	7.676356 8.804097

Calculus:

$$Educ = 8.24 + .190\text{motheduc} + .109\text{fatheduc} + .401\text{abil} + .050\text{abil2} + u$$

$$d(Educ)/d(abil) = .401 + .10\text{abil} = 0$$

$$\text{abil}^* = -0.401/0.10 = -4.01$$

$$d^2(Educ)/d^2(abil) = 0.10 > 0 \text{ and therefore we have a minimum.}$$

(v) Only 14 men out of the 1,230 have an ability level below -4.01.

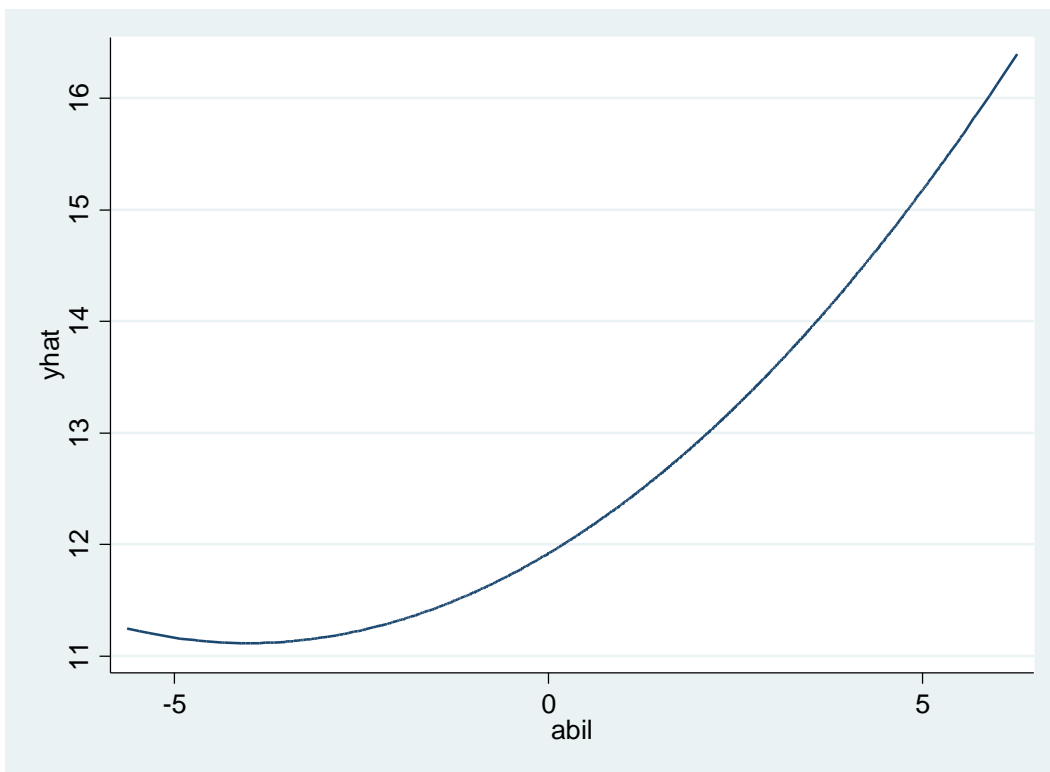
```
. summarize educ if abil < -4.01
```

Variable	Obs	Mean	Std. Dev.	Min	Max
educ	14	10.57143	1.650841	8	12

This is important because only 14 individuals will have a higher expected education level than individuals slightly to the right of -4.01. That is, only 14 fitted values of the above equation are counter-intuitive.

(vi)

```
generate yhat = 11.92 + 0.401*abil + 0.05*abil2  
twoway (line yhat abil, sort), ytitle(yhat) xtitle(abil)
```



Here is the STATA code that gave rise to the above answers:

```
summarize educ  
summarize educ if educ == 12  
summarize educ motheduc fatheduc  
regress educ motheduc fatheduc  
regress educ motheduc fatheduc abil  
generate abil2 = abil*abil  
regress educ motheduc fatheduc abil abil2  
summarize educ if abil < -4.01
```

```
generate yhat = 11.92 + 0.401*abil + 0.05*abil2
twoway (line yhat abil, sort), ytitle(yhat) xtitle(abil)
```

C11 page 148

(i) The estimated equation, with standard errors in parentheses below coefficient estimates, is

$$educ = 8.24 + .190 motheduc + .109 fatheduc + .401 abil + .0506 abil^2$$

$$(0.29) \quad (.028) \quad (.020) \quad (.030) \quad (.0083)$$

$$n = 1,230, R^2 = .444.$$

The null hypothesis of a linear relationship between *educ* and *abil* is $H_0 : \beta_4 = 0$ and the alternative is that H_0 does not hold. The *t* statistic is about $.0506 / .0083 \approx 6.1$, which is a very large value for a *t* statistic. The *p*-value against the two-sided alternative is zero to more than four decimal places.

(ii) We could rewrite the model by defining, say, $\theta_1 = \beta_1 - \beta_2$ and then substituting in $\beta_1 = \theta_1 + \beta_2$, just as we did with the example in Section 4.4. These days, it is easier to use a special command in statistical software. The estimated difference in the coefficients is about .081. The instructor could use the *lincom* command in Stata to get a *t* statistic of about 1.94 and an associated two-sided *p*-value of about .053. So there is some evidence against the null hypothesis.

(iii) The instructor could use the *test* command in Stata to test the joint significance of the tuition variables. With 2 and 1,223 degrees of freedom the instructor get an *F* statistic of about .84 with association *p*-value of about .43. Thus, the tuition variables are jointly insignificant at any reasonable significance level.

(iv) Not surprisingly, the correlation between *tuit17* and *tuit18* is very high, about .981: there is very little change in tuition over a year that cannot be explained by a common inflation factor. The instructor could generate the variable $avgtuit = (tuit17 + tuit18)/2$, and then added it to the regression from part (i). The coefficient on *avgtuit* is about .016 with $t = 1.29$. This certainly helps with statistical significance but the two-sided *p*-value is still only about .20.

(v) The positive coefficient on *avgtuit* does not make a lot of sense if we think that, all other things fixed, higher tuition makes it less likely that people go to college. But we are only controlling for parents' levels of education and a measure of ability. It could be that higher tuition indicates higher quality of the state colleges. Or, it could be that tuition is higher in states with higher average incomes, and higher family incomes lead to higher education. In any case, the statistical link is not very strong.

Here is the STATA code that gave rise to the above answers:

generate abil2 = abil*abil

regress educ motheduc fatheduc abil abil2

* Here we use the "lincom" command to test a linear combination of coefficients

* In particular, we are interested in whether an additional year of mother's

* education has the same effect as an additional year of father's education.

* There is some suggestion that the effects are not the same. The p-value of

* the t-ratio is $p = 0.053$ which is barely larger than 0.05. There is a

* statistical difference at the 10% level. Maybe the mother's education is

* of slightly greater influence than the father's education.

lincom motheduc - fatheduc

* Here is alternative way to test the same hypothesis using the F-statistic.

* We get the same p-value.

test motheduc - fatheduc == 0

* Here we add two additional explanatory variables to our equation: tuit17 and

* tuit18. Let's examine their joint significance.

regress educ motheduc fatheduc abil abil2 tuit17 tuit18

* Both of the tuit variables are singularly insignificant. However, let us

* test the joint significance of the two variables.

test tuit17 tuit18

* Noting the collinearity between the two tuition variables.

corr tuit17 tuit18

generate avgtuit = (tuit17 + tuit18)/2

* Even taking the average of the tuition variables does not provide a

* very satisfactory result. Tuition continues to have an unexpected positive

* coefficient.

regress educ motheduc fatheduc abil abil2 avgtuit

```

. use "E:\E5350\Wooldridge_Intro_Econometrics_ed6\Data Sets\HTV.DTA", clear

. save "E:\E5350\E5350 f16\Exercises\HTV.dta"
file E:\E5350\E5350 f16\Exercises\HTV.dta saved

. do "C:\Users\00008904\AppData\Local\Temp\STD02000000.tmp"

. generate abil2 = abil*abil

. regress educ motheduc fatheduc abil abil2

```

Source	SS	df	MS	Number of obs	=	1,230
Model	3027.03706	4	756.759264	F(4, 1225)	=	244.91
Residual	3785.24262	1,225	3.08999397	Prob > F	=	0.0000
Total	6812.27967	1,229	5.54294522	R-squared	=	0.4444
				Adj R-squared	=	0.4425
				Root MSE	=	1.7578

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
motheduc	.1901261	.0280957	6.77	0.000	.1350051 .2452472
fatheduc	.1089387	.0196014	5.56	0.000	.0704827 .1473946
abil	.4014624	.0302875	13.26	0.000	.3420413 .4608835
abil2	.050599	.0083039	6.09	0.000	.0343076 .0668905
_cons	8.240226	.2874099	28.67	0.000	7.676356 8.804097

```

. * Here we use the "lincom" command to test a linear combination of coefficients
. * In particular, we are interested in whether an additional year of mother's
. * education has the same effect as an additional year of father's education.
. * There is some suggestion that the effects are not the same. The p-value of
. * the t-ratio is p = 0.053 which is barely larger than 0.05. There is a
. * statistical difference at the 10% level. Maybe the mother's education is
. * of slightly greater influence than the father's education.
. lincom motheduc - fatheduc

```

(1) motheduc - fatheduc = 0

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	.0811875	.0419431	1.94	0.053	-.0011009	.1634758

. * Here is alternative way to test the same hypothesis using the F-statistic.
 . * We get the same p-value.
 . test motheduc - fatheduc == 0

(1) motheduc - fatheduc = 0

F(1, 1225) = 3.75
 Prob > F = 0.0531

. * Here we add two additional explanatory variables to our equation: tuit17 and
 . * tuit18. Let's examine their joint significance.
 . regress educ motheduc fatheduc abil abil2 tuit17 tuit18

Source	SS	df	MS	Number of obs	=	1,230
Model	3032.22546	6	505.37091	F(6, 1223)	=	163.51
Residual	3780.05421	1,223	3.09080475	Prob > F	=	0.0000
Total	6812.27967	1,229	5.54294522	R-squared	=	0.4451
				Adj R-squared	=	0.4424
				Root MSE	=	1.7581

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
motheduc	.1928925	.0281804	6.84	0.000	.1376052	.2481797
fatheduc	.1084427	.019618	5.53	0.000	.0699541	.1469314
abil	.3990426	.0303493	13.15	0.000	.3395002	.458585
abil2	.0505545	.0083126	6.08	0.000	.0342459	.066863
tuit17	.0157588	.0625029	0.25	0.801	-.106866	.1383836
tuit18	.0000603	.0636477	0.00	0.999	-.1248105	.1249311
_cons	8.081865	.3127685	25.84	0.000	7.468242	8.695487

. * Both of the tuit variables are singularly insignificant. However, let us
 . * test the joint significance of the two variables.
 . test tuit17 tuit18

(1) tuit17 = 0

(2) tuit18 = 0

F(2, 1223) = 0.84
 Prob > F = 0.4322

```

. * Noting the collinearity between the two tuition variables.
. corr tuit17 tuit18
(obs=1,230)

```

	tuit17	tuit18
tuit17	1.0000	
tuit18	0.9808	1.0000

```

. generate avgtuit = (tuit17 + tuit18)/2

```

```

. * Even taking the average of the tuition variables does not provide a
. * very satisfactory result. Tuition continues to have an unexpected positive
. * coefficient.
. regress educ motheduc fatheduc abil abil2 avgtuit

```

Source	SS	df	MS	Number of obs	=	1,230
Model	3032.17713	5	606.435426	F(5, 1224)	=	196.36
Residual	3780.10254	1,224	3.08831907	Prob > F	=	0.0000
Total	6812.27967	1,229	5.54294522	R-squared	=	0.4451
				Adj R-squared	=	0.4428
				Root MSE	=	1.7574

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
motheduc	.1928633	.0281681	6.85	0.000	.1376001 .2481264
fatheduc	.1083681	.019601	5.53	0.000	.0699127 .1468234
abil	.3990808	.0303355	13.16	0.000	.3395654 .4585962
abil2	.0505989	.0083017	6.10	0.000	.0343119 .066886
avgtuit	.015963	.0123734	1.29	0.197	-.0083125 .0402384
_cons	8.081339	.3126144	25.85	0.000	7.468019 8.694658

```

.
end of do-file

```