ECO 441 Fall 2009

Solutions-4

1)

	(1)	(2)	(3)
College	5.46**	5.48**	5.44**
	(0.21)	(0.21)	(0.21)
Female	-2.64**	-2.62**	-2.62**
	(0.20)	(0.20)	(0.20)
Age		0.29**	0.29**
		(0.04)	(0.04)
Northeast			0.69*
			(0.30)
Midwest			0.60*
			(0.28)
South			-0.27
			(0.26)
Intercept	12.69**	4.40**	3.75**
	(0.14)	(1.05)	(1.06)

2) a) The $t-statistic$ is $5.46/0.21 = 26.0 > 1.96$, so the coefficient is statistically significant at the 5% level. The 95% confidence interval is $5.46 \pm 1.96 * 0.21$.

b) The $t-statistic$ is $-2.64/0.20 = -13.2 > 1.96$, so the coefficient is statistically significant at the 5% level. The 95% confidence interval is $-2.64 \pm 1.96 * 0.20$.

3) a) Yes, age is an important determinant of earnings. Using a $t-test$,$the t-statistics$ is $\frac{0.29}{0.04} = 7.25$ implying that the coefficient on age is statistically significant at the 1% level. The 95% confidence interval is $0.29 \pm 1.96 * 0.04$.

b) $\Delta Age * [0.29 \pm 1.96 * 0.04] = 5 * [0.29 \pm 1.96 * 0.04] = 1.45 \pm 1.96 * 0.20 = \$1.06$ to $\$1.84$.

4) a) The $F-statistic$ testing the coefficients on the regional regressors are zero is 6.10. The 1% critical value from the F distribution is 3.78. Because $6.10 > 3.78$, the regional effects are significant at the 1% level.

5) a) Using the expressions for R^2 and $\bar{R}^2$, algebra shows that

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2),\ so\ R^2 = 1 - \frac{n-k-1}{n-1}(1-\bar{R}^2)$$

$$Col1: R^2 = 1 - \frac{420-1-1}{420-1}(1-0.049) = 0.051$$
$$Col2: R^2 = 1 - \frac{420-2-1}{420-1}(1-0.424) = 0.427$$
$$Col3: R^2 = 1 - \frac{420-3-1}{420-1}(1-0.773) = 0.775$$

$$Col4: R^2 = 1 - \frac{420-3-1}{420-1}(1-0.626) = 0.629$$
$$Col5: R^2 = 1 - \frac{420-4-1}{420-1}(1-0.773) = 0.775$$

b)

$$H_0 \quad : \quad \beta_3 = \beta_4 = 0$$
$$H_0 \quad : \quad \beta_3 \neq, \beta_4 \neq 0$$

$$Unrestricted\ Regression\ (Col5) = Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4,\ R^2_{unrestricted} = 0.775$$

$$Restricted\ Regression (Col2) = Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2, R^2_{restricted} = 0.427$$

$$\begin{aligned} F_{\hom oskedasticity} &= \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1-R^2_{unrestricted})/n - k_{unrestricted} - 1},\ n = 420, k_{unrestricted} = 4, q = 2 \\ &= \frac{(0.775 - 0427)/2}{(1-0.775)/(420-4-1)} = \frac{0.174}{0.00054} = 322.22 \end{aligned}$$

5% critical value for F is $F_{2,\infty} = 4.61$, therefore, we reject the null hypothesis. (I used Appendix Table 4 for F since the homoskedastic/heteroskedastic critical values converges to each other in large samples. The conclusion is similar when you use Appendix Table 5).

6) Since $R^2 = 1 - \frac{SSR}{TSS}$, $R^2_{unrestricted} - R^2_{restricted} = \frac{SSR_{restricted} - SSR_{unrestricted}}{TSS}$ and $1 - R^2_{unrestricted} = \frac{SSR_{unrestricted}}{TSS}$. Thus,

$$\begin{aligned} F &= \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1-R^2_{unrestricted})/n - k_{unrestricted} - 1} \\ &= \frac{(\frac{SSR_{restricted} - SSR_{unrestricted}}{TSS})/q}{(\frac{SSR_{unrestricted}}{TSS})/(n - k_{unrestricted} - 1)} \\ &= \frac{(SSR_{restricted} - SSR_{unrestricted})/q}{SSR_{unrestricted}/(n - k_{unrestricted} - 1)} \end{aligned}$$

7) a) The estimated intercept is -1.133, the estimated slope is 0.591.

b) The estimated effect of age on earnings is 0.618 dollars, while keeping other regressors constant. The 95% confidence interval is from 0.435 to 0.800.

c) The results are quite similar. Evidently the regression in (a) does not suffer from important omitted variable bias.

d) Bob's predicted earnings is 12.40, Alexis's predicted earnings is 18.76.

e) The regression in part (b) fits the data much better. Gender and education are important predictors of earnings. The R^2 and $\bar{R}^2$ are similar because the sample size is large.

f) Gender and education are statistically significant determinants of earnings. Therefore, we can not delete them from the regression. The joint hypothesis yields a $p-value$ of 0.00, which indicates the rejection of the null hypothesis.

g) The omitted variables must have non-zero coefficients and must be correlated with the included regressor. From (f) *Female* and *Bachelor* have non-zero coefficients; yet there does not seem to be important omitted variable bias, suggesting that the correlation of *Age* and *Female* and *Age* and *Bachelor* is small.

8) a) The group's claim is that the coefficient on *Dist* is$-0.075 (= -0.15/2)$. The confidence interval for *Dist* is $-0.109\ to - 0.002$. the group's claim is included in the 95% confidence interval so that it is consistent with the estimated regression.

b) I chose the following specifications (you may choose more specifications):

Spec 1: Regress Edu on Dist

Spec 2: Regress Edu on Dist Bytest Female Black Hispanic Incomehi Ownhome Dadcoll Momcoll Cue80 Stwmfg80

Spec 3: Regress Edu on Dist Bytest Female Black Hispanic Incomehi Ownhome Dadcoll Momcoll Cue80 Stwmfg80 Urban Tuition

The estimated coefficient on *Dist* decreases from -0.056 to 0.004 from Spec 1 to Spec 2. More importantly, the coefficient is no longer statistically significant as we augment the additional regressors to the model. Including urban and tuition, on the other hand, does not seem to affect the results qualitatively given that the coefficient on *Dist* remains insignificant.

c) Yes, the estimated coefficients for *Black* and *Hispanic* are positive, large and statistically significant.