Chapter 7

Hypothesis Tests and Confidence Intervals in Multiple Regression

■ Solutions to Exercises

1.

Regressor	**(1)**	**(2)**	**(3)**
College (X_1)	5.46** (0.21)	5.48** (0.21)	5.44** (0.21)
Female (X_2)	–2.64** (0.20)	–2.62** (0.20)	–2.62** (0.20)
Age (X_3)		0.29** (0.04)	0.29** (0.04)
Ntheast (X_4)			0.69* (0.30)
Midwest (X_5)			0.60* (0.28)
South (X_6)			–0.27 (0.26)
Intercept	12.69** (0.14)	4.40** (1.05)	3.75** (1.06)

(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 $5.46 \pm 1.96 \times 0.21$.

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

3. (a) Yes, age is an important determinant of earnings. Using a t-test, the t-statistic is $\frac{0.29}{0.04} = 7.25,$ with a p-value of 4.2×10^{-13}, implying that the coefficient on age is statistically significant at the 1% level. The 95% confidence interval is $0.29 \pm 1.96 \times 0.04$.

(b) $\Delta Age \times [0.29 \pm 1.96 \times 0.04] = 5 \times [0.29 \pm 1.96 \times 0.04] = 1.45 \pm 1.96 \times 0.20 = \$1.06$ 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 $F_{3,\infty}$ distribution) is 3.78. Because $6.10 > 3.78$, the regional effects are significant at the 1% level.

(b) The expected difference between Juanita and Molly is $(X_{6,\text{Juanita}} - X_{6,\text{Molly}}) \times \beta_6 = \beta_6$. Thus a 95% confidence interval is $-0.27 \pm 1.96 \times 0.26$.

(c) The expected difference between Juanita and Jennifer is $(X_{5,\text{Juanita}} - X_{5,\text{Jennifer}}) \times \beta_5 + (X_{6,\text{Juanita}} - X_{6,\text{Jennifer}}) \times \beta_6 = -\beta_5 + \beta_6$. A 95% confidence interval could be contructed using the general methods discussed in Section 7.3. In this case, an easy way to do this is to omit *Midwest* from the regression and replace it with $X_5 = West$. In this new regression the coefficient on *South* measures the difference in wages between the *South* and the *Midwest*, and a 95% confidence interval can be computed directly.

5. The *t*-statistic for the difference in the college coefficients is $t = (\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992})/SE(\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992})$. Because $\hat{\beta}_{college,1998}$ and $\hat{\beta}_{college,1992}$ are computed from independent samples, they are independent, which means that $\text{cov}(\hat{\beta}_{college,1998}, \hat{\beta}_{college,1992}) = 0$ Thus, $\text{var}(\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992}) = \text{var}(\hat{\beta}_{college,1998}) + \text{var}(\hat{\beta}_{college,1998})$. This implies that $SE(\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992}) = (0.21^2 + 0.20^2)^{\frac{1}{2}}$. Thus, $t^{act} = \frac{5.48-5.29}{(0.21^2+0.20^2)^{\frac{1}{2}}} = 0.6552$. There is no significant change since the calculated *t*-statistic is less than 1.96, the 5% critical value.

6. In isolation, these results do imply gender discrimination. Gender discrimination means that two workers, identical in every way but gender, are paid different wages. Thus, it is also important to control for characteristics of the workers that may affect their productivity (education, years of experience, etc.) If these characteristics are systematically different between men and women, then they may be responsible for the difference in mean wages. (If this were true, it would raise an interesting and important question of why women tend to have less education or less experience than men, but that is a question about something other than gender discrimination.) These are potentially important omitted variables in the regression that will lead to bias in the OLS coefficient estimator for *Female*. Since these characteristics were not controlled for in the statistical analysis, it is premature to reach a conclusion about gender discrimination.

7. (a) The *t*-statistic is $\frac{0.485}{2.61} = 0.186 < 1.96$. Therefore, the coefficient on BDR is not statistically significantly different from zero.

(b) The coefficient on *BDR* measures the *partial effect* of the number of bedrooms holding house size (*Hsize*) constant. Yet, the typical 5-bedroom house is much larger than the typical 2-bedroom house. Thus, the results in (a) says little about the conventional wisdom.

(c) The 99% confidence interval for effect of lot size on price is $2000 \times [.002 \pm 2.58 \times .00048]$ or 1.52 to 6.48 (in thousands of dollars).

(d) Choosing the scale of the variables should be done to make the regression results easy to read and to interpret. If the lot size were measured in thousands of square feet, the estimate coefficient would be 2 instead of 0.002.

(e) The 10% critical value from the $F_{2,\infty}$ distribution is 2.30. Because 0.08 < 2.30, the coefficients are not jointly significant at the 10% level.

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

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

$$\text{Column 1: } R^2 = 1 - \frac{420-1-1}{420-1}(1-0.049) = 0.051$$

$$\text{Column 2: } R^2 = 1 - \frac{420-2-1}{420-1}(1-0.424) = 0.427$$

$$\text{Column 3: } R^2 = 1 - \frac{420-3-1}{420-1}(1-0.773) = 0.775$$

$$\text{Column 4: } R^2 = 1 - \frac{420-3-1}{420-1}(1-0.626) = 0.629$$

$$\text{Column 5: } R^2 = 1 - \frac{420-4-1}{420-1}(1-0.773) = 0.775$$

(b)

$$H_0: \beta_3 = \beta_4 = 0$$
$$H_1: \beta_3 \neq, \beta_4 \neq 0$$

Unrestricted regression (Column 5): $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$, $R^2_{\text{unrestricted}} = 0.775$

Restricted regression (Column 2): $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, $R^2_{\text{restricted}} = 0.427$

$$F_{HomoskedasticityOnly} = \frac{(R^2_{\text{unrestricted}} - R^2_{\text{restricted}})/q}{(1-R^2_{\text{unrestricted}})/(n-k_{\text{unrestricted}}-1)}, \ n = 420, \ k_{\text{unrestricted}} = 4, \ q = 2$$
$$= \frac{(0.775-0.427)/2}{(1-0.775)/(420-4-1)} = \frac{0.348/2}{(0.225)/415} = \frac{0.174}{0.00054} = 322.22$$

5% Critical value form $F_{2,00} = 4.61$; $F_{HomoskedasticityOnly} > F_{2,00}$ so H_o is rejected at the 5% level.

(c) $t_3 = -13.921$ and $t_4 = 0.814$, $q = 2$; $|t_3| > c$ (Where $c = 2.807$, the 1% Benferroni critical value from Table 7.3). Thus the null hypothesis is rejected at the 1% level.

(d) $-1.01 \pm 2.58 \times 0.27$

9. (a) Estimate

$$Y_i = \beta_0 + \gamma X_{1i} + \beta_2 (X_{1i} + X_{2i}) + u_i$$

and test whether $\gamma = 0$.

(b) Estimate

$$Y_i = \beta_0 + \gamma X_{1i} + \beta_2 (X_{2i} - aX_{1i}) + u_i$$

and test whether $\gamma = 0$.

(c) Estimate

$$Y_i - X_{1i} = \beta_0 + \gamma X_{1i} + \beta_2 (X_{2i} - X_{1i}) + u_i$$

and test whether $\gamma = 0$.

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

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