Chapter 5
Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals

5.1 (a) The 95% confidence interval for β_1 is $\{-5.82 \pm 1.96 \times 2.21\}$, that is $-10.152 \leq \beta_1 \leq -1.4884$.

(b) Calculate the *t*-statistic:

$$t^{act} = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)} = \frac{-5.82}{2.21} = -2.6335.$$

The *p*-value for the test $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ is

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-2.6335) = 2 \times 0.0042 = 0.0084.$$

The *p*-value is less than 0.01, so we can reject the null hypothesis at the 5% significance level, and also at the 1% significance level.

(c) The *t*-statistic is

$$t^{act} = \frac{\hat{\beta}_1 - (-5.6)}{\mathrm{SE}(\hat{\beta}_1)} = \frac{0.22}{2.21} = 0.10$$

The *p*-value for the test $H_0: \beta_1 = -5.6$ vs. $H_1: \beta_1 \neq -5.6$ is

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-0.10) = 0.92$$

The *p*-value is larger than 0.10, so we cannot reject the null hypothesis at the 10%, 5% or 1% significance level. Because $\beta_1 = -5.6$ is not rejected at the 5% level, this value is contained in the 95% confidence interval.

(d) The 99% confidence interval for β_0 is $\{520.4 \pm 2.58 \times 20.4\}$, that is, $467.7 \leq \beta_0 \leq 573.0$.

5.2. (a) The estimated gender gap equals \$2.12/hour.

(b) The null and alternative hypotheses are $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$. The *t*-statistic is

$$t^{act} = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{2.12}{0.36} = 5.89,$$

and the p-value for the test is

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-5.89) = 2 \times 0.0000 = 0.000 \text{ (to four decimal places)}$$

The p-value is less than 0.01, so we can reject the null hypothesis that there is no gender gap at a 1% significance level.

(c) The 95% confidence interval for the gender gap β_1 is $\{2.12 \pm 1.96 \times 0.36\}$, that is, $1.41 \le \beta_1 \le 2.83.$

(d) The sample average wage of women is $\hat{\beta}_0 = \$12.52$/hour. The sample average wage of men is $\hat{\beta}_0 + \hat{\beta}_1 = \$12.52 + \$2.12 = \$14.64$/hour.

(e) The binary variable regression model relating wages to gender can be written as either

$$Wage = \beta_0 + \beta_1 Male + u_i,$$

or

$$Wage = \gamma_0 + \gamma_1 Female + v_i.$$

In the first regression equation, $Male$ equals 1 for men and 0 for women; β_0 is the population mean of wages for women and $\beta_0 + \beta_1$ is the population mean of wages for men. In the second regression equation, $Female$ equals 1 for women and 0 for men; γ_0 is the population mean of wages for men and $\gamma_0 + \gamma_1$ is the population mean of wages for women. We have the following relationship for the coefficients in the two regression equations:

$$\gamma_0 = \beta_0 + \beta_1,$$
$$\gamma_0 + \gamma_1 = \beta_0.$$

Given the coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, we have

$$\hat{\gamma}_0 = \hat{\beta}_0 + \hat{\beta}_1 = 14.64,$$
$$\hat{\gamma}_1 = \hat{\beta}_0 - \hat{\gamma}_0 = -\hat{\beta}_1 = -2.12.$$

Due to the relationship among coefficient estimates, for each individual observation, the OLS residual is the same under the two regression equations: $\hat{u}_i = \hat{v}_i$. Thus the sum of squared residuals, $SSR = \sum_{i=1}^{n} \hat{u}_i^2$, is the same under the two regressions. This implies that both $SER = \left(\frac{SSR}{n-1}\right)^{1/2}$ and $R^2 = 1 - \frac{SSR}{TSS}$ are unchanged.

In summary, in regressing $Wages$ on $Female$, we will get

$$Wages = 14.64 - 2.12 Female, \quad R^2 = 0.06, \quad SER = 4.2.$$

5.3. The 99% confidence interval is 1.5 × {3.94 ± 2.58 × 0.31) or 4.71 lbs ≤ WeightGain ≤ 7.11 lbs.

5.4. (a) − 5.38 + 1.76 × 16 − $22.78 per hour

(b) The wage is expected to increase by 1.76 × 2 = $3.52 per hour.

(c) The increase in wages for college education is $\beta_1 \times 4$. Thus, the counselor's assertion is that $\beta_1 = 10/4 = 2.50$. The *t*-statistic for this null hypothesis is $t^{act} = \dfrac{1.76 - 2.50}{0.08} - 9.25$, which has a *p*-value of 0.00. Thus, the counselor's assertion can be rejected at the 1% significance level. A 95% confidence for $\beta_1 \times 4$ is $4 \times (1.76 \pm 1.96 \times 0.08)$ or $\$6.41 \leq \text{Gain} \leq \$7.67$.

5.5 (a) The estimated gain from being in a small class is 13.9 points. This is equal to approximately 1/5 of the standard deviation in test scores, a moderate increase.

(b) The *t*-statistic is $t^{act} = \dfrac{13.9}{2.5} = 5.56$, which has a *p*-value of 0.00. Thus the null hypothesis is rejected at the 5% (and 1%) level.

(c) $13.9 \pm 2.58 \times 2.5 = 13.9 \pm 6.45$.

5.7. (a) The *t*-statistic is $\dfrac{3.2}{1.5} = 2.13$ with a *p*-value of 0.03; since the *p*-value is less than 0.05, the null hypothesis is rejected at the 5% level.

(b) $3.2 \pm 1.96 \times 1.5 = 3.2 \pm 2.94$

(c) Yes. If *Y* and *X* are independent, then $\beta_1 = 0$; but this null hypothesis was rejected at the 5% level in part (a).

(d) β_1 would be rejected at the 5% level in 5% of the samples; 95% of the confidence intervals would contain the value $\beta_1 = 0$.

5.8. (a) $43.2 \pm 2.05 \times 10.2$ or 43.2 ± 20.91, where 2.05 is the 5% two-sided critical value from the t_{28} distribution.

(b) The *t*-statistic is $t^{act} = \frac{61.5-55}{7.4} = 0.88$, which is less (in absolute value) than the critical value of 20.5. Thus, the null hypothesis is not rejected at the 5% level.

(c) The one sided 5% critical value is 1.70; t^{act} is less than this critical value, so that the null hypothesis is not rejected at the 5% level.