Capitolo 4
Regressione Lineare con un Regressore

■ Soluzioni

1. (a) La predizione della regressione relativa al punteggio medio del test per tale classe è

$$\widehat{TestScore} = 520.4 - 5.82 \times 22 = 392.36$$

(b) La predizione della regressione relativa alla varaizione del punteggio medio del test nella classe è

$$\widehat{\Delta TestScore} = (-5.82 \times 19) - (-5.82 \times 23) = 23.28$$

(c) L'intervallo di confidenza al 95% è 5.82±(1.96*2.21)=(1.4884, 10.15)

(d) Il valore-p di un test bilaterale per l'ipotesi nulla H_0: $\beta_1=0$ è

$$valore - p = \Pr(|Z| > |t^{act}|) = 2\Phi(-|t^{act}|) = 2\Phi(-2.633) = 0.0086$$

L' ipotesi nulla si rigetta sia al 5% che all'1%.

(c) Utilizzando la formula per $\hat{\beta}_0$, sappiamo che la media campionaria del punteggio del test nelle 100 classi è

$$\overline{TestScore} = \hat{\beta}_0 + \hat{\beta}_1 \times \overline{CS} = 520.4 - 5.82 \times 21.4 = 395.85.$$

(d) Utilizziamo la formula per l'errore standard della regressione (SER) per ottenere la somma dei quadrati residui:

$$SSR = (n-2)SER^2 = (100-2) \times 11.5^2 = 12961.$$

Utilizziamo la formula di R^2 per calcolare la somma dei quadrati totale:

$$TSS = \frac{SSR}{1-R^2} = \frac{12961}{1-0.08^2} = 13044.$$

La varianza del campione è $s_Y^2 = \frac{TSS}{n-1} = \frac{13044}{99} = 131.8.$ Quindi la deviazione standardizzata campionaria è $s_Y = \sqrt{s_Y^2} = 11.5.$

2. (a) La differenza stimata dovuta al genere è \$2.79/hour.

(b) La verifica d'ipotesi circa la differenza dovuta al genere è $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0.$ La statistica-t è

$$t^{act} = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{2.79}{0.84} = 3.32,$$

il valore-p per il test è

$$valore - p = 2\Phi(-|t^{act}|) = 2\Phi(-3.32) = 2 \times 0.000 = 0.000 \text{ (fino a 4 decimali)}$$

Il valore-p è minore di 0.01, quindi rigettiamo l'ipotesi nulla che non vi è differenza di genere all'1%.

(c) L'intervallo di confidenza al 95% per la differenza di genere β_1 è $\{2.79 \pm 1.96 \times 0.84\} = (1.1436, 4.4364)$

(d) Il salario medio delle donne nel campione è $\hat{\beta}_0 = \$12.68/ora$ Il salario medio degli uomini nel campione è $\hat{\beta}_0 + \hat{\beta}_1 = \$12.68 + \$2.79 = \$15.47/ora$

(e) Il modello di regressione con variabile binaria (dummy) può essere scritto sia

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

sia

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

Nella prima regressione, $Male$ è uguale ad 1 per gli uomini e a 0 per le donne; β_0 è la media della popolazione per le donne e $\beta_0 + \beta_1$ è la media della popolazione per gli uomini. Nella seconda regressione, $Female$ è uguale ad 1 per le donne e 0 per gli uomini; γ_0 è la media della popolazione per gli uomini e $\gamma_0 + \gamma_1$ è la media della popolazione per le donne. Quindi troviamo la seguente relazione tra i coefficienti delle due regressioni:

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

Avendo le stime dei coefficienti $\hat{\beta}_0$ e $\hat{\beta}_1$, otteniamo

$$\hat{\gamma}_0 = \hat{\beta}_0 + \hat{\beta}_1 = 15.47$$
$$\hat{\gamma}_1 = \hat{\beta}_0 - \hat{\gamma}_0 = -2.79$$

Sulla base della relazione tra i coefficienti stimati, per ogni osservazione individuale, i residui OLS sono gli stessi nelle due regressioni: $\hat{u}_i = \hat{v}_i$. Quindi la somma dei quadrati dei residui, $SSR = \sum_{i=1}^{n} \hat{u}_i^2$, è la stessa nelle due regressioni. Questo implica che sia $SER = \left(\frac{SSR}{n-1}\right)^{\frac{1}{2}}$ che $R^2 = 1 - \frac{SSR}{TSS}$ non mutano.

In conclusione, facendo la regressione di $Wages$ su $Female$, otteniamo

$$W\hat{a}ge = 15.47 - 2.79 Female, \; R^2 = 0.06, \; SER = 3.10$$