■ Soluzioni Esercizi quarta settimana

1.

Regressori	**(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)
Intercetta	12.69** (0.14)	4.40** (1.05)	3.75** (1.06)

3. (a) La statistica t è 5.46/0.21 = 26.0 > 1.96, quindi il coefficiente è statisticamente significativo al 5%. L'intervallo di confidenza al 95% è 5.46 ± 1.96 × 0.21.

   (b) La statistica t è –2.64/0.20 = –13.2, e 13.2 > 1.96, quindi il coefficiente è statisticamente significativo al 5%. L'intervallo di confidenza al 95% è –2.64 ± 1.96 × 0.20.

4. (a) Si, l'età è un importante determinante dei guadagni. La statistica t è $\frac{0.29}{0.04} = 7.25$, con un valore-p di 4.2×10^{-13}, implicando che il coefficiente del'età è statisticamente significativo all'1%. L'intervallo di confidenza al 95% è 0.29 ± 1.96 × 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, \$1.84)

5. (a) La statistica F usata per testare l'ipotesi nulla che i coefficienti dei regressori regionali siano zero è 6.10. Il valore critico all'1% è (dalla distribuzione $F_{3,\infty}$) 3.78. Dato che 6.10 > 3.78, gli effetti regionali sono significativi all'1%.

   (b) Collinearità perfetta.

   (ci) La differenza attesa tra Juanita e Molly è $(X_{6,\text{Juanita}} - X_{6,\text{Molly}}) \times \beta_6 = \beta_6$. L'intervallo di confidenza al 95% è –0.27 ± 1.96 × 0.26.

   (cii) La differenza attesa tra Juanita e Jennifer è $(X_{5,\text{Juanita}} - X_{5,\text{Jennifer}}) \times \beta_5 + (X_{6,\text{Juanita}} - X_{6,\text{Jennifer}}) \times \beta_6 = -\beta_5 + \beta_6$= -0.60-0.27=-0.87.

(ciii) Per costruire un intervallo di confidenza al 95% si può escludere *Midwest* dalla regressione e inserire X_5 = *West*. Nella nuova regressione il coefficiente di *South* misura la differenza nei salari tra *South* e *Midwest*, e così si può realizzare l'intervallo di confidenza desiderato.

5. La statistica t per la differenza tra i coefficienti relativi alla variabile College è $t = (\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992})/SE(\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992})$. Dato che $\hat{\beta}_{college,1998}$ e $\hat{\beta}_{college,1992}$ sono ottenuti da campione indipendenti, sono anch'essi indipendeti e quindi $\text{cov}(\hat{\beta}_{college,1998}, \hat{\beta}_{college,1992}) = 0$. Quindi, $\text{var}(\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992}) = \text{var}(\hat{\beta}_{college,1998}) + \text{var}(\hat{\beta}_{college,1998})$. Ciò implica che e che $t^{act} = \frac{5.48-5.29}{(0.21^2+0.20^2)^{\frac{1}{2}}} = 0.6552$. Concludendo, non vi sono cambiamenti significativi dato che la statistica t calcolata è minore del valore critico 1.96.

6. Astraendo dal contesto, tale risultato implica una discriminazione dovuta al genere. Questo fenomeno significa che due lavoratori, identici sotto ogni profilo eccetto il genere, percepiscono salari diversi. È perciò importante controllare per le caratteristiche dei lavoratori che possono influenzare la loro produttività (istruzione, esperienza, etc.). Se tali caratteristiche sono sistematicamente differenti tra uomini e donne, allora esse potrebbero essere le responsabili delle differenze nei salari. Queste sono variabili potenzialmente importanti omesse dalla regressione che comporteranno delle distorsioni sui sugli stimatori OLS dei coefficienti della variabile *Female*. Dato che tali caratteristiche non sono state prese in considerazione nell'analisi statistica, sembra avventato giungere a conclusioni definitive circa la discriminazione di genere