

Esercizi empirici

X

E7.1 Si usi l'insieme di dati **CPS08** descritto nell'Esercizio empirico 4.1 per rispondere alle seguenti domande.

- Si effettui una regressione della retribuzione oraria media (*AHE*) sull'età (*Age*). Qual è l'intercetta stimata? Qual è la pendenza stimata?
- Si effettui una regressione di *AHE* su *Age*, genere (*Female*) e istruzione (*Bachelor*). Qual è l'effetto stimato di *Age* sulla retribuzione? Si costruisca un intervallo di confidenza al 95% per il coefficiente di *Age* nella regressione.
- I risultati derivati dalla regressione nella (b) sono sostanzialmente diversi dai risultati nella (a) riguardanti l'effetto di *Age* e *AHE*? La regressione nella (a) sembra soffrire di distorsione da variabili omesse?
- Bob è un lavoratore di 26 anni con un diploma di scuola media superiore. Si predica la retribuzione di Bob usando la regressione stimata nella (b). Alexis è una lavoratrice laureata di 30 anni. Si predica la retribuzione di Alexis usando la regressione.
- Si confronti l'adattamento delle regressioni nella (a) e nella (b) usando il SER, l' R^2 e l' $\bar{R}^2$. Perché l' R^2 e l' $\bar{R}^2$ sono così simili nella regressione (b)?
- Genere e istruzione sono determinanti delle retribuzioni? Si verifichi l'ipotesi nulla che *Female* può essere omissa dalla regressione. Si verifichi l'ipotesi nulla che *Bachelor* può essere omissa dalla regressione. Si verifichi l'ipotesi nulla che sia *Female* che *Bachelor* possono essere omissi entrambi dalla regressione.
- Una regressione soffre di distorsione da variabili omesse quando valgono due condizioni. Quali sono queste due condizioni? Sembrano valere qui?

E7.2 Usando l'insieme di dati **TeachingRatings** descritto nell'Esercizio empirico 4.2, si svolgano i seguenti esercizi.

- Si effettui una regressione di *Course_Eval* su *Beauty*. Si costruisca un intervallo di confidenza al 95% per l'effetto di *Beauty* su *Course_Eval*.
- Si considerino le diverse variabili di controllo disponibili nei dati. Quali si ritiene di dover includere nella regressione? Usando una tabella simile alla Tabella 7.1, si esamini la robustezza dell'intervallo di confidenza costruito nella (a). Qual è un ragionevole intervallo di confidenza al 95% per l'effetto di *Beauty* su *Course_Eval*.

E7.3 Si usi l'insieme di dati **CollegeDistance** descritto nell'Esercizio empirico 4.3 per rispondere alle seguenti domande.

- Un gruppo di pressione afferma che, in media, il risultato scolastico di una persona aumenterebbe di approssimativamente 0,15 anni se la distanza dall'università più vicina diminuisse di 20 miglia. Si effettui una regressione del numero di anni di istruzione completati (*ED*) sulla distanza dall'università più vicina (*Dist*). L'affermazione del gruppo di pressione è coerente con la regressione stimata? Si argomenti la risposta.
- Anche altri fattori influenzano gli anni di università che una persona completa. Controllando per questi altri fattori cambia l'effetto stimato della distanza sugli anni di università completati? Per rispondere a questa domanda, si costruisca una tabella simile alla Tabella 7.1. Si includa una specificazione semplice (da riportare nella colonna (a)), una specificazione di base (che includa una serie di importanti variabili di controllo) e varie modifiche della specificazione di base. Si discuta come l'effetto stimato di *Dist* su *ED* cambia tra le diverse specificazioni.

- Si è affermato che, con di università dei bianchi, la parte (b)?

X

E7.4 Usando l'insieme di dati su Malta, si si

- Si effettui una regressione di *Assassinations* e *RGDI* su *TradeShare*.
- Si verifichi se, presi *RGDP60* possano essere correlati? Qual è l'intercetta



Soluzioni e dati

Appendice 7.1: Il test di Bonferroni

Il metodo del Paragrafo 7.2 è un test di ipotesi multipla. Tuttavia, se l'ipotesi nulla non sottopone a verifica la regressione, non si potrà calcolare un modo per verificare ipotesi individuali. Una tabella con i risultati di un test molto generale alla verifica di ipotesi individuali.

Il test di Bonferroni consente di verificare ipotesi individuali; in altre parole, si verifica "coefficiente per coefficiente".

Il test di Bonferroni dell'ipotesi nulla $H_0: \beta_j = 0$, segue la regola:

accetta l'ipotesi nulla se $|t_j| \leq t_{\alpha/c}$ (test di Bonferroni)

dove t_1 e t_2 sono le statistiche di test.

Il trucco sta nello scegliere il valore critico c in modo da tenere conto delle correzioni, sia della possibile correlazione tra le variabili che della possibile correlazione tra le variabili di controllo.

Disuguaglianza di Bonferroni

La disuguaglianza di Bonferroni riguarda due eventi. Sia A e B due eventi. Sia $A \cup B$ l'evento " A o B ". Poiché $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$. Poiché $\Pr(A \cap B) \geq 0$, questa disuguaglianza a sua volta implica che $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$. Poiché A^c e B^c i complementi di A e B , il complemento di $A \cup B$ è $A^c \cap B^c$, e la disuguaglianza di Bonferroni: $\Pr(A^c \cap B^c) \geq 1 - \Pr(A) - \Pr(B)$.