

Example 1 – Soluzione completa

Testo del problema (riassunto)

Un ricercatore ha stimato il seguente modello di regressione lineare multipla:

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \beta_5 x_{5,t} + \varepsilon_t. \quad (1)$$

L'output di stima fornisce:

Parametro	Coefficiente	Errore Std.	Z-statistic	p-value
α	1.436	0.264	5.439	0.000
β_1	-0.645	0.536	-1.203	0.229
β_2	2.664	1.132	(manc.)	0.019
β_3	3.623	1.187	(manc.)	(manc.)
β_4	0.994	0.226	4.398	0.000
β_5	-2.644	6.624	-0.399	(manc.)

Sono richiesti i seguenti punti:

1. Test di significatività del coefficiente β_2 .
2. Numero di parametri (inclusa l'intercetta) significativi all'1%.
3. Rappresentazione del vincolo $\beta_3 - \beta_2 = \beta_4$ nella forma $R\beta = r$ e forma dell'F-test (tipo Wald).
4. Scrivere il modello sotto la restrizione e la forma dell'F-test basato sul confronto tra modello ristretto e non ristretto.
5. Dati $RSS_U = 543,2$, $RSS_R = 552,12$, $T = 240$, calcolare la statistica F e decidere il rifiuto di H_0 al 5%.

Nel seguito risolviamo passo-passo tutti i punti.

Punto 1: test di significatività per β_2

1. Formulazione delle ipotesi

Il test di significatività di un coefficiente di regressione è, salvo diversa indicazione, un test a due code. Per β_2 :

$$H_0 : \beta_2 = 0, \quad H_1 : \beta_2 \neq 0.$$

L'ipotesi nulla afferma che il regressore $x_{2,t}$ non ha effetto lineare (statisticamente rilevante) su y_t .

2. Statistica test teorica

La statistica del test (tipo Z o t) per un singolo coefficiente è:

$$Z = \frac{\hat{\beta}_2 - \beta_2^{(0)}}{SE(\hat{\beta}_2)}, \quad (2)$$

dove $\beta_2^{(0)}$ è il valore imposto da H_0 . Qui $\beta_2^{(0)} = 0$, quindi

$$Z = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)}. \quad (3)$$

Sotto H_0 , in grandi campioni, Z ha approssimativamente distribuzione normale standard $N(0, 1)$.

3. Calcolo della statistica empirica

Dalla tabella:

$$\hat{\beta}_2 = 2.664, \quad SE(\hat{\beta}_2) = 1.132.$$

Sostituendo:

$$Z_{\text{emp}} = \frac{2.664}{1.132} \approx 2.35. \quad (4)$$

Questa è la Z-statistic mancante per β_2 .

4. Valori critici e decisione

Per un test a due code:

- livello di significatività 5%: valore critico $|z_{0.975}| \approx 1.96$;
- livello di significatività 1%: valore critico $|z_{0.995}| \approx 2.576$.

Confrontiamo il valore assoluto:

$$|Z_{\text{emp}}| \approx 2.35.$$

- Al 5%: $2.35 > 1.96 \Rightarrow$ **rigettiamo** H_0 .
- All'1%: $2.35 < 2.576 \Rightarrow$ **non rigettiamo** H_0 .

La conclusione è coerente con il p-value riportato (0.019):

- $0.019 < 0.05 \Rightarrow$ rigetto a livello 5%;
- $0.019 > 0.01 \Rightarrow$ non rigetto a livello 1%.

Conclusione punto 1. Il coefficiente β_2 è statisticamente significativo al 5%, ma non è significativo all'1%.

Punto 2: numero di parametri significativi all'1%

Vogliamo sapere quanti parametri (inclusa l'intercetta) risultano significativi al livello di significatività $\alpha = 0.01$, cioè al 99% di confidence level, usando test a due code.

1. Regola di decisione all'1%

Per ciascun coefficiente (inclusa l'intercetta) consideriamo:

$$H_0 : \beta_j = 0, \quad H_1 : \beta_j \neq 0.$$

La statistica è:

$$Z_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}.$$

Per un test a due code al livello $\alpha = 0.01$:

$$\text{rigetto di } H_0 \iff |Z_j| > 2.576.$$

2. Z-statistics (calcolate dove mancano)

- Intercetta: $|Z_\alpha| = |5.439| = 5.439$.
- β_1 : $|Z_1| = |-1.203| = 1.203$.
- β_2 : già calcolato al punto 1:

$$Z_2 = \frac{2.664}{1.132} \approx 2.35.$$

- β_3 :

$$Z_3 = \frac{3.623}{1.187} \approx 3.05.$$

- β_4 : $|Z_4| = |4.398| = 4.398$.
- β_5 : $|Z_5| = |-0.399| = 0.399$.

3. Confronto con la soglia 2.576

- α : $5.439 > 2.576 \Rightarrow$ significativo all'1%.
- β_1 : $1.203 < 2.576 \Rightarrow$ non significativo all'1%.
- β_2 : $2.35 < 2.576 \Rightarrow$ non significativo all'1%.
- β_3 : $3.05 > 2.576 \Rightarrow$ significativo all'1%.
- β_4 : $4.398 > 2.576 \Rightarrow$ significativo all'1%.
- β_5 : $0.399 < 2.576 \Rightarrow$ non significativo all'1%.

Conclusione punto 2. I parametri significativi all'1% (inclusa l'intercetta) sono α , β_3 e β_4 . In totale:

3 parametri significativi all'1%.

Punto 3: vincolo $\beta_3 - \beta_2 = \beta_4$ in forma $R\beta = r$

1. Scrittura del vincolo in forma standard

Il vincolo è:

$$\beta_3 - \beta_2 = \beta_4.$$

Portando tutto a sinistra:

$$\beta_3 - \beta_2 - \beta_4 = 0.$$

2. Vettore dei parametri

Raggruppiamo tutti i parametri in un unico vettore colonna:

$$\beta = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}.$$

3. Matrice R e vettore r

Vogliamo un'espressione del tipo:

$$R\beta = r.$$

Poiché c'è un solo vincolo, R è una matrice 1×6 e r è uno scalare.

I coefficienti della combinazione lineare $\beta_3 - \beta_2 - \beta_4$ vengono posti in corrispondenza di $(\alpha, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$:

$$R = (0 \quad 0 \quad -1 \quad 1 \quad -1 \quad 0), \quad r = (0).$$

4. Forma F (Wald) della statistica test

Vogliamo testare:

$$H_0 : R\beta = r \quad \text{vs} \quad H_1 : R\beta \neq r.$$

La statistica di Wald in forma F è:

$$F = \frac{1}{q} (R\hat{\beta} - r)' \left[R \widehat{\text{Var}}(\hat{\beta}) R' \right]^{-1} (R\hat{\beta} - r), \quad (5)$$

dove:

$$\widehat{\text{Var}}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1},$$

e $q = \text{rank}(R)$ è il numero di vincoli (qui $q = 1$).

Sotto H_0 :

$$F \sim F_{q, T-k},$$

dove k è il numero di parametri nel modello non ristretto.

Gradi di libertà. Poiché c'è un solo vincolo lineare indipendente, il primo grado di libertà (al numeratore) è:

$$q = 1.$$

Punto 4: modello sotto la restrizione e F-test via RSS

1. Modello ristretto

Il modello originario è:

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \beta_5 x_{5,t} + \varepsilon_t.$$

La restrizione è:

$$\beta_3 - \beta_2 = \beta_4 \implies \beta_4 = \beta_3 - \beta_2.$$

Sostituendo β_4 nel modello:

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + (\beta_3 - \beta_2) x_{4,t} + \beta_5 x_{5,t} + \varepsilon_t \quad (6)$$

$$= \alpha + \beta_1 x_{1,t} + \beta_2 (x_{2,t} - x_{4,t}) + \beta_3 (x_{3,t} + x_{4,t}) + \beta_5 x_{5,t} + \varepsilon_t. \quad (7)$$

Questo è il **modello ristretto**: i parametri liberi sono $\alpha, \beta_1, \beta_2, \beta_3, \beta_5$ (5 parametri invece dei 6 originari).

2. Statistica F tramite confronto RSS

Definiamo:

- RSS_U : somma dei residui al quadrato del modello *non ristretto*;
- RSS_R : somma dei residui al quadrato del modello *ristretto*.

Per testare $H_0 : \beta_3 - \beta_2 = \beta_4$ si usa:

$$F = \frac{(RSS_R - RSS_U)/q}{RSS_U/(T - k)}, \quad (8)$$

dove:

- $q = 1$ è il numero di vincoli lineari indipendenti;
- $k = 6$ è il numero di parametri nel modello non ristretto (intercetta + 5 coefficienti);
- T è il numero di osservazioni.

Sotto H_0 :

$$F \sim F_{q, T-k} = F_{1, T-6}.$$

Punto 5: calcolo della F-statistic e decisione al 5%

Sono forniti:

$$RSS_U = 543,2, \quad RSS_R = 552,12, \quad T = 240.$$

Ricordiamo:

$$q = 1, \quad k = 6 \quad (\text{parametri nel modello non ristretto}).$$

I gradi di libertà del denominatore sono:

$$T - k = 240 - 6 = 234.$$

1. Calcolo della statistica F

Usiamo la formula:

$$F = \frac{(RSS_R - RSS_U)/q}{RSS_U/(T - k)}.$$

Calcoliamo passo-passo:

1. Differenza tra RSS ristretto e non ristretto:

$$RSS_R - RSS_U = 552.12 - 543.2 = 8.92.$$

2. Termine al denominatore:

$$\frac{RSS_U}{T - k} = \frac{543.2}{234} \approx 2.321.$$

3. Statistica F:

$$F_{\text{emp}} = \frac{8.92}{2.321} \approx 3.84.$$

2. Confronto con il valore critico al 5%

Il testo fornisce i valori critici:

$$F(1, 234) = 3.88, \quad F(2, 234) = 3.03, \quad F(3, 234) = 2.64.$$

Nel nostro caso il test ha distribuzione $F_{1,234}$, quindi il valore critico al 5% è:

$$F_{1,234}^{(5\%)} = 3.88.$$

Confrontiamo:

$$F_{\text{emp}} \approx 3.84 < 3.88.$$

La statistica empirica non supera il valore critico: la statistica cade nella regione di non-rifiuto.

Conclusione punto 5. Al livello di significatività del 5%, **non rigettiamo** l'ipotesi nulla $H_0 : \beta_3 - \beta_2 = \beta_4$. I dati non forniscono evidenza sufficiente per rifiutare la restrizione al 5%.