

Esercizio 1



Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Per quattro Ingegneri si sono osservati gli anni trascorsi dalla Laurea X e il reddito annuale lordo Y misurato in euro. I risultati sono riassunti dalla seguente tabella:

Ingegnere	1	2	3	4
X	1	4	5	10
Y	15751,94	19625,36	21949,42	41316,55

- 1 Si determini la retta di regressione di Y su X
- 2 Se $Z = X^2$, la retta di regressione di Y su Z ha un migliore adattamento?
- 3 Si stimi il reddito annuale lordo di un Ingegnere laureato da 3 anni

Soluzione es.1

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

- (1) Ipotizziamo che valga un modello del tipo $Y = \beta_0 + \beta_1 x$. Per determinare la retta di regressione occorre calcolare le stime $\hat{\beta}_0$ e $\hat{\beta}_1$ di β_0 e β_1 rispettivamente. È noto che $\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}}$ e $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$; utilizzando queste formule e calcolando

$$s_{xy} = 1.2395 \cdot 10^5, \quad s_{xx} = 42 \quad s_{yy} = 3.8949 \cdot 10^8$$

si ottiene $\hat{\beta}_0 = 9904,87$ e $\hat{\beta}_1 = 2951,19$. L'espressione della retta di regressione (stimata) è dunque $\hat{y} = 9904,87 + 2951,19x$.

Soluzione es.1

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

- (2) Per confrontare due modelli di regressione utilizziamo il coefficiente di correlazione; è migliore il modello che presenta il coefficiente di correlazione più alto. Il coeff. per la regressione di Y su Z è dato dalla relazione

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}} = 0,9691$$

mentre, calcolando

$$s_{yz} = 1.5083 \cdot 10^6, \quad s_{zz} = 5841,$$

quello di Z su X è dato da

$$r_{zy} = \frac{s_{zy}}{\sqrt{s_{zz}s_{yy}}} = 1.$$

Il fatto che $r_{zy} = 1$ indica non solo che il secondo modello è preferibile al primo (ha un coeff. di correlaz. maggiore), ma anche che la relazione $Y = \hat{\gamma}_0 + \hat{\gamma}_1 z = 15493,69 + 258,23z$ è esatta.

Soluzione es.1

- (3) Utilizzando dunque il secondo modello prevediamo che un Ingegnere laureato da 3 anni guadagni
- $$Y = 15493,69 + 258,23 * 3^2 = 17817,76 \text{ euro.}$$

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Esercizio 2

I valori assunti da due grandezze X ed Y in 20 diversi casi forniscono per entrambe le grandezze media campionaria nulla e varianze campionarie $s_X^2 = 9$ e $s_Y^2 = 4$ rispettivamente.

- 1 Si determini la retta di regressione di Y su X sapendo che essa passa per il punto $(3, 1)$
- 2 Si calcoli il coefficiente r_{xy}^2
- 3 Si determini la retta di regressione di X su Y

Soluzione es.2

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

- (1) Detti $\hat{\beta}_0$ e $\hat{\beta}_1$ i coefficienti della retta di regressione, si deve avere che $1 = \hat{\beta}_0 + \hat{\beta}_1 3$. Questa equazione non è sufficiente da sola a determinare $\hat{\beta}_0$ e $\hat{\beta}_1$. Per il modo in cui viene costruita la retta dei minimi quadrati deve anche accadere che $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$. Dal testo sappiamo che $\bar{x} = 0$ e $\bar{y} = 0$. Risolvendo il sistema otteniamo $\hat{\beta}_0 = 0$ e $\hat{\beta}_1 = \frac{1}{3}$. Dunque la retta di regressione (stimata) ha equazione $y = \frac{1}{3}x$.

Soluzione es.2

- (2) Notiamo che r_{xy}^2 si può scrivere come $r_{xy}^2 = \frac{\text{cov}(X, Y)^2}{\text{var}(X)\text{var}(Y)}$.
Non conoscendo i dati, non possiamo calcolare direttamente $\text{cov}(X, Y)$. Possiamo però ricavarla osservando che $\hat{\beta}_1 = \frac{\text{cov}(X, Y)}{\text{var}(X)}$ (la varianza campionaria di X è nota dal testo). Si ha dunque
 $\text{cov}(X, Y) = \hat{\beta}_1 \text{var}(X) = \frac{1}{3} * 9 = 3$. Da ciò ricaviamo infine $r_{xy}^2 = \frac{9}{9*4} = \frac{1}{4}$

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Soluzione es.2

- (3) Chiamiamo $\hat{\alpha}_0$ ed $\hat{\alpha}_1$ i coefficienti della retta di regressione di X su Y . Utilizzando quanto visto nei punti precedenti possiamo determinare $\hat{\alpha}_1$ come

$$\hat{\alpha}_1 = \frac{\text{cov}(Y, X)}{\text{var}(Y)} = \frac{\text{cov}(X, Y)}{\text{var}(Y)} = \frac{3}{4}.$$

Inoltre, dalla relazione $\bar{y} = \hat{\alpha}_0 + \hat{\alpha}_1 \bar{x}$ ricaviamo $\hat{\alpha}_0 = \bar{x} - \hat{\alpha}_1 \bar{y} = 0$. La retta di regressione di X su Y ha quindi equazione $\hat{x} = \frac{3}{4}y$

Esercizio 3

In una tabella sono raccolti i seguenti dati

X	2	3	4.5	7
Y	10	10	15	20

- 1 Calcolare la retta di regressione di Y su X
- 2 Calcolare il coefficiente di correlazione di X ed Y
- 3 Calcolare il coefficiente di correlazione di X e $Z = \ln Y$.
Questo modello è migliore del precedente?

Soluzione es.3

(1) Calcoliamo innanzitutto le medie campionarie di X e Y.

$$\bar{x} = \frac{2 + 3 + 4.5 + 7}{4} = 4.125$$

$$\bar{y} = \dots = 13.75$$

Calcoliamo ora i coefficienti della retta di regressione di Y su X:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \dots = 2.1586$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 4.8458$$

L'espressione della retta di regressione (stimata) è dunque $\hat{y} = 4.8458 + 2.1586x$.

Soluzione es.3

- (2) Siccome $s_{xx} = 14.1875$, $s_{yy} = 68.75$ e $s_{xy} = 30.625$, il coefficiente di correlazione di Y e X è dato da

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}} = 0.98059$$

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Soluzione es.3

- (3) Per confrontare due modelli di regressione utilizziamo il coefficiente di correlazione; è migliore il modello che presenta il coefficiente di correlazione più elevato. Facendo i conti, troviamo che $\bar{Z} = 2.58$ e il coefficiente di correlazione di Z e X è dato da

$$r_{zx} = \frac{s_{zx}}{\sqrt{s_{zz}s_{xx}}} = 0.972.$$

Siccome $r_{xz} < r_{xy}$ deduciamo che è migliore il modello in (1) di quello in (2).

Esercizio 4

Si vuole studiare la relazione tra le variabili Peso delle madri (X) e Peso dei figli (Y), entrambe misurate in Kg. Le osservazioni (x_i, y_i) per $i = 1, \dots, 12$ sono tali che $\sum_{i=1}^{12} x_i = 800$, $\sum_{i=1}^{12} y_i = 811$, $\sum_{i=1}^{12} x_i^2 = 53418$, $\sum_{i=1}^{12} y_i^2 = 54849$, $\sum_{i=1}^{12} x_i y_i = 54107$.

- ① Determinare la retta di regressione di Y su X.
- ② Proporre uno stimatore non distorto per la varianza σ^2 delle componenti di errore ε_i del modello di regressione. Supposte valide le ipotesi gaussiane:
- ③ verificare l'ipotesi nulla $H_0 : \beta_1 = 0$ contro $H_1 : \beta_1 \neq 0$ al livello del 2%.
- ④ verificare se la retta di regressione passa per l'origine al livello del 2%.

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Soluzione es.4

- (1) Per calcolare la retta di regressione incominciamo a ricavare s_{xx} e s_{xy} dai nostri dati.

$$\bar{x} = \frac{1}{n} \sum_i x_i = 66.\bar{6}$$

$$\bar{y} = \frac{1}{n} \sum_i y_i = 67.6$$

$$s_{xx} = \sum_i x_i^2 - n\bar{x}^2 = 85.73$$

$$s_{xy} = \sum_i x_i y_i - n\bar{x}\bar{y} = 40.\bar{3}$$

quindi

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = 0.476$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 35$$

Soluzione es.4

- (2) Uno stimatore non distorto della varianza degli errori della regressione è dato da

$$\begin{aligned}\hat{\sigma}^2 &= \frac{SS_E}{n-2} = \frac{\sum_i (y_i - \hat{y}_i)^2}{n-2} = \frac{\sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2} \\ &= \dots \simeq 2\end{aligned}$$

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Soluzione es.4

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

(3) Vogliamo testare l'ipotesi

$$H_0 : \beta_1 = 0 = \beta_{1,0}$$

contro

$$H_1 : \beta_1 \neq 0$$

Utilizziamo quindi la statistica test

$$\frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t_{n-2}$$

accettando l'ipotesi nulla al livello di significatività α se

$$-t_{\alpha/2; n-2} \leq \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \leq t_{\alpha/2; n-2}. \text{ Siccome}$$

$$\frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = 3.116 > t_{\alpha/2; n-2} = 2.764, \text{ allora rifiutiamo } H_0$$

al livello del 2%.

Soluzione es.4

(4) Vogliamo testare l'ipotesi

$$H_0 : \beta_0 = 0 = \beta_{0,0}$$

contro

$$H_1 : \beta_0 \neq 0$$

Utilizziamo quindi la statistica test

$$\frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}} \sim t_{n-2}$$

accettando l'ipotesi nulla al livello di significatività α se

$$-t_{\alpha/2;n-2} \leq \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}} \leq t_{\alpha/2;n-2}.$$

Soluzione es.4

(4) Siccome $\frac{\bar{x}^2}{S_{xx}} = 51.84$, $\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right] = 103.852$,

$\frac{\hat{\beta}_0}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}} = 3.434 > t_{\alpha/2; n-2} = 2.764$, allora rifiutiamo
 H_0 al livello del 2%.

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Esercizio 5

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8



Giant sequoia, Sequoia & King's Canyon NP (CA), USA

Nel corso di uno studio naturalistico sono stati raccolti dati su 8 esemplari di una certa specie di albero: sono state misurate le altezze x_i (in metri) e i pesi y_i (in Kg.).

Nell'intento di evidenziare una relazione tra le grandezze si ipotizza un legame lineare:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

sotto le ipotesi gaussiane (cioè $\varepsilon \sim N(0; \sigma^2)$).

Esercizio 5



Giant sequoia, Sequoia & King's Canyon NP (CA), USA

Dopo l'esecuzione della regressione, si sono trovati i seguenti valori per le stime dei parametri e la somma dei residui al quadrato:

$$\hat{\beta}_0 = 10.2; \hat{\beta}_1 = 25.3; SS_E = 172.5$$

con stime degli errori standard di $\hat{\beta}_0$ e $\hat{\beta}_1$ date da

$$se(\hat{\beta}_0) = 5.7859; se(\hat{\beta}_1) = 0.8088$$

Esercizio 5



Giant sequoia, Sequoia & King's Canyon NP (CA), USA

- 1 Calcolare un intervallo di confidenza al livello 80% per β_0 e per β_1
- 2 verificare se la retta di regressione passa per l'origine al livello 20% ed al livello 10%
- 3 calcolare gli intervalli di confidenza al 90% per la previsione media e per la previsione del peso di un albero alto 10 metri.

Soluzione es.5

- (1) Gli intervalli di confidenza per i parametri β_0 e β_1 della regressione sono della forma:

$$\hat{\beta}_0 - t_{\alpha/2; n-2} \text{se}(\hat{\beta}_0) \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2; n-2} \text{se}(\hat{\beta}_0)$$

$$\hat{\beta}_1 - t_{\alpha/2; n-2} \text{se}(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2; n-2} \text{se}(\hat{\beta}_1)$$

Avendo quindi $\alpha/2 = 0.1$; $n = 8$; $t_{\alpha/2; 6} = 1.440$,
otteniamo quindi:

$$\begin{aligned} 1.87 &\leq \beta_0 \leq 18.53 \\ 23.65 &\leq \beta_1 \leq 26.95 \end{aligned}$$

Soluzione es.5

(2) Vogliamo testare l'ipotesi

$$H_0 : \beta_0 = 0 = \beta_{0,0}$$

contro

$$H_1 : \beta_0 \neq 0$$

Siccome $\beta_{0,0} = 0$ non appartiene all'intervallo di confidenza trovato sopra, rifiutiamo H_0 al livello di significatività 20%. Mentre l'intervallo di confidenza al livello 90% è, sapendo che in questo caso

$$t_{\alpha/2, n-2} 1.9432,$$

$$[-1.043, 21.413];$$

in questo caso $\beta_0 = 0$ appartiene all'intervallo, pertanto l'ipotesi nulla è accettata.

Soluzione es.5

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

- (3) Dai dati ricaviamo immediatamente
 $\hat{\sigma}^2 = SS_E / (n - 2) \approx 21.5625$. Ricordando che

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}$$

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

da cui $s_{xx} \approx 39.9622$ e

$$\bar{x}^2 = \left(\frac{se(\hat{\beta}_0)^2}{\hat{\sigma}^2} - \frac{1}{n} \right) S_{xx} \approx 47.0549$$

ma, essendo per la natura del problema $\bar{x} \geq 0$, si ha
 $\bar{x} \approx 6.8597$.

(3) Infine

$$\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) \approx 9.1464$$

$$\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) \approx 30.7089$$

da cui, calcolando come al punto precedente
 $t_{\alpha/2, n-2} = 1.9432$ (con $\alpha = 0.1$), i due intervalli sono,
 rispettivamente,

$$[260.1757, 266.2243] \quad [257.6584, 268.7416].$$

Esercizio 6

In una certa comunità si registrano mensilmente il consumo di gelati X misurato in Kg. e il numero di casi di allergia al polline Y. I dati raccolti nell'ultimo anno forniscono le seguenti informazioni:

$$\bar{X} = 120; \quad S_X = 100$$

$$\bar{Y} = 22; \quad S_Y = 5$$

Il coefficiente di correlazione vale 0.93.

- 1 Si determini la retta di regressione di Y su X e se ne disegni il grafico
- 2 Si stimi (puntualmente) il numero di casi di allergia in un mese in cui il consumo di gelati è pari a 300 Kg. La previsione ottenuta è buona?

Soluzione es.6

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

- 1 Per calcolare la retta di regressione incominciamo a ricavare s_{xx} e s_{xy} dai nostri dati. Siccome $S_X^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$ e analogamente per S_Y^2 , otteniamo:

$$s_{xx} = (n-1) s_X^2 = 1100$$

$$s_{yy} = \dots = 55$$

Siccome

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = r_{xy} \frac{s_{yy}}{s_{xx}} = 0.0465$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 16.42$$

- 2 Dalla retta di regressione otteniamo

$$\hat{y}(300) = 16.42 + 0.0465 \cdot 300 = 30.37,$$

quindi il numero stimato di casi di allergia in un mese in cui il consumo di gelati è pari a 300 Kg è 30.37

Esercizio 7

Da un'elaborazione preliminare sui seguenti dati:

x_i	y_i
2.1	0.8518
3.2	2.3551
-1.2	-8.7368
-3.4	-11.2042
2.3	0.8329
2.4	-1.1961
1.7	2.3834
-0.9	-9.3468
-0.8	-6.1546
1.9	-1.9388

Esercizio 7

risultano le seguenti deviazioni standard campionarie $\sigma_X = 2.1427$, $\sigma_Y = 5.1807$ ed il coefficiente di correlazione $\rho_{XY} = 0.9421$. Si esegue una regressione lineare supponendo $Y = \beta_0 + \beta_1 \cdot x + \epsilon$ e che valgano le ipotesi gaussiane.

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Esercizio 7

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

- 1 Stimare i coefficienti β_0 e β_1 .
- 2 Calcolare un intervallo di confidenza per β_0 al 95%.
- 3 Si valuti l'opportunità di aggiungere il nuovo regressore x^2

$$Y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \epsilon$$

con un test al 5% sapendo che per il modello in questione $SS_E = 25.639$.

- 4 Si considera il seguente modello

$$Y = \beta_0 + \beta_1 \cdot x^3 + \epsilon;$$

quanto dovrebbe valere la somma dei quadrati residua affinché sia preferibile questo modello a quello di regressione semplice?

Soluzione es.7

Anche in questo esercizio i dati sono ridondanti in quanto tutti i coefficienti che si servono possono essere ricavati dai dati in tabella.

(1) Poniamo

$$s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2,$$

$$s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

dove $\bar{x} = 0.73$ e $\bar{y} = -3.2154$;

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Soluzione es.7

(1) allora

$$s_{xx} = \sigma_x^2(n-1) = 9 \cdot 2.1427^2 = 41.3205,$$

$$s_{yy} = \sigma_y^2(n-1) = 9 \cdot 5.1807^2 = 241.5569,$$

$$s_{xy} = \rho_{xy} \sqrt{s_{xx}s_{yy}} = 0.9421 \cdot \sqrt{41.3205 \cdot 241.5569} = 94.12$$

Da cui facilmente

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = 2.2778,$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = -3.2154 - 2.2778 \cdot 0.73 = -4.8782.$$

(2) Considerando la matrice

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{k,1} \\ 1 & x_{1,2} & \cdots & x_{k,2} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1,n} & \cdots & x_{k,n} \end{pmatrix}$$

si calcola facilmente

$$(X^t \cdot X)^{-1}_{1,1} = \frac{1}{n} + \frac{\bar{x}^2}{s_{xx}} = \frac{1}{10} + \frac{0.73^2}{41.3205} = 0.1129.$$

Ricordiamo che l'errore quadratico medio (o media quadratica dei residui) è

$$\hat{\sigma}^2 = MS_E := \frac{1}{n-p} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

(in questo caso $n = 10$ e $p = k + 1 = 2$).

Soluzione es.7

(2) Nel caso di regressione semplice

$$\hat{\sigma}^2 = \left(SS_T - \hat{\beta}_1 s_{xy} \right) \frac{1}{n-2} = \left(s_{yy} - \frac{s_{xy}^2}{s_{xx}} \right) \frac{1}{n-2} = 3.3953$$

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Soluzione es.7

(2) Allora l'intervallo di confidenza a livello α è $[\hat{\beta}_o^-, \hat{\beta}_o^+]$ dove

$$\hat{\beta}_o^\pm := \hat{\beta}_0 \pm \text{se}(\hat{\beta}_0) \cdot t_{\frac{\alpha}{2}, n-p}$$

e

$$\text{se}(\hat{\beta}_0) := \sqrt{\hat{\sigma}^2 \cdot (X^t \cdot X)^{-1}_{1,1}} = \sqrt{3.3953 \cdot 0.1129} = 1.4277,$$

quindi, essendo $t_{0.025, 8} = 2.306$, si ha che l'intervallo cercato è $[-8.1705, -1.5859]$.

Soluzione es.7

- (3) Nel modello con regressore aggiunto, si tratta di testare l'ipotesi nulla

$$H_0 : \beta_2 = 0.$$

Si calcola

$$\begin{pmatrix} 10 & 7.3 & 46.65 \\ 7.3 & 46.65 & 37.519 \\ 46.65 & 37.519 & 343.6245 \end{pmatrix}$$

(3) da cui

$$\begin{aligned}\hat{\beta} &= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^t \cdot X)^{-1} \cdot X^t \cdot Y \\ &= \begin{pmatrix} 0.2815 & -0.0146 & -0.0366 \\ -0.0146 & 0.0243 & -0.0007 \\ -0.0366 & -0.0007 & 0.008 \end{pmatrix} \cdot \begin{pmatrix} -32.1541 \\ 70.6524 \\ -128.3333 \end{pmatrix} \\ &= \begin{pmatrix} -5.3827 \\ 2.2687 \\ 0.1096 \end{pmatrix}.\end{aligned}$$

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

(3) Pertanto

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} = \frac{25.639}{10-3} = 3.6627,$$

e quindi

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 (X^t \cdot X)^{-1}_{3,3}} = 35.4767.$$

Lo stimatore che si utilizza è

$$T_0 := \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{0.1096}{35.4767} = 0.0031$$

con regione di accettazione

$$|T_0| \leq t_{\alpha/2, n-p} \equiv t_{0.025, 7} = 2.3646.$$

L'ipotesi nulla è pertanto accettata e non è opportuno aggiungere il regressore x^2 .

Soluzione es.7

- (4) Il secondo modello, con k_2 regressori, risulta migliore del primo, con k_1 regressori, se e solo se

$$\frac{SS_{E_2}}{n - k_2 - 1} < \frac{SS_{E_1}}{n - k_1 - 1}.$$

Nel caso in questione si ha $n = 10$, $k_1 = k_2 = 1$ ed $SS_{E_1} = 27.1627$ per cui il modello con regressore x^3 sarà migliore di quello con x se e solo se

$$SS_{E_2} < 27.1627.$$

Esercizio 8

Ad un gruppo di 4 uomini vengono misurati peso X , altezza Y e circonferenza toracica Z ; il risultato è riassunto nella seguente tabella:

X (in Kg)	Y (in cm)	Z (in cm)
93	185	99
78	183	103
76	178	95
77	174	92

Esercizio 8

Eseguendo una regressione di X rispetto ad Y e Z si ottiene

$$SS_E = \sum_{i=1}^n (\hat{x}_i - x_i)^2 = 22.56232409,$$

$$SS_R = \sum_{i=1}^n (\hat{x}_i - \bar{x})^2 = 171.4376759,$$

$$SS_T = \sum_{i=1}^n (x_i - \bar{x})^2 = 194.$$

- 1 Eseguire un test per la significatività della regressione al 10%.
- 2 Determinare il P-value del test.

Soluzione es.8

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8

Osserviamo che i dati del problema sono ridondanti e che gli “errori quadratici” si ricavano banalmente dai dati in tabella. D’altro canto conosciuti gli “errori quadratici”, non siamo più interessati ai dati della tabella se non per conoscere l’ampiezza del campione $n = 4$ ed il numero di regressori $k = 2$.

Soluzione es.8

(1) Si consideri l'ipotesi nulla

$$H_0 : \beta_0 = \beta_1 = \beta_2 = 0.$$

Lo stimatore è

$$F_0 := \frac{SS_R/k}{SS_E/(n-p)}$$

(dove $p = k + 1$) con la regione di accettazione a livello α

$$f_0 < f_{\alpha,k,n-p}.$$

Eseguendo i calcoli si ottiene

$$f_0 = \frac{171.4376759/2}{22.56232409/(4-3)} = \frac{85.7188}{22.56232409} = 3.7992$$

e $f_{0.1,2,1} = 49.5$. Pertanto accetto H_0 al livello del 5% pertanto la regressione non è significativa.

Soluzione es.8

(2) Se utilizziamo le tabelle dei quantili otteniamo

$$f_{0.25,2,1} = 7.5 > 3.7992$$

da cui $\bar{\alpha} > 0.25$. Utilizzando un calcolatore si ottiene

$$\bar{\alpha} = f_{.,2,1}^{-1}(3.7992) = 1 - F_{f_{2,1}}(3.7992) = 0.341.$$

Esercizio 1

Esercizio 2

Esercizio 3

Esercizio 4

Esercizio 5

Esercizio 6

Esercizio 7

Esercizio 8