Soluzioni Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

A.A.: 2020/2021
Docente: Gianluca Orlando
Appello: VIII Appello
Data: 19/11/2021

Esercizio 1. Dei provini cilindrici di calcestruzzo mostrano le seguenti resistenze a compressione (in MPa):

17	21	21	26	24	19	14	19	23	27	31	31	26	26	22	27

Sapendo che il campione è estratto da una popolazione normale con media incognita μ e varianza nota $\sigma^2 = 4$, con che livelli di significatività si può accettare l'ipotesi $\mu = 25$?

Soluzione. Il campione estratto $X_1, \dots, X_n$ ha ampiezza $n = 16$. Ogni X_i ha distribuzione normale con media μ e varianza $\sigma^2 = 4$. Vogliamo testare l'ipotesi

$$H_0 : \mu = 25\,, \quad H_1 : \mu \neq 25\,.$$

Utilizziamo la statistica

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\,,$$

dove $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ è la media campionaria. Poiché ogni X_i è distribuita in modo normale, Z è distribuita come una normale standard $\mathcal{N}(0,1)$ (nonostante $n < 30$).
Denotiamo con z_α i quantili gaussiani, ovvero i numeri reali tali che

$$\alpha = P(Z > z_\alpha) = P(Z < -z_\alpha)\,.$$

Osserviamo che, per simmetria della Gaussiana,

$$\alpha = 2\frac{\alpha}{2} = 2P(Z > z_{\alpha/2}) = P(Z < -z_{\alpha/2}) + P(Z > z_{\alpha/2}) = P(|Z| > z_{\alpha/2})\,.$$

La traccia chiede, in pratica, di calcolare il p-value del test. Ricordiamone la definizione.
Il livello di significatività α è la probabilità di commettere un errore del primo tipo. Si può accettare l'ipotesi nulla con livello di significatività α se si verifica la seguente condizione: assumendo vera l'ipotesi $\mu = 25$, la realizzazione del campione osservata $X_1, \dots, X_n$ verifica

$$\Big|\frac{\overline{X} - 25}{2/\sqrt{16}}\Big| \leq z_{\alpha/2}\,,$$

(infatti, la probabilità che questo non accada, assumendo $\mu = 25$, è α). Il più piccolo $z_{\alpha/2}$ che verifica questa condizione è proprio quello ottenuto dall'osservazione

$$\Big|\frac{\overline{X} - 25}{2/\sqrt{16}}\Big| = 3.25\,.$$

Quindi il più grande livello di significatività α per cui si può accettare l'ipotesi nulla è

$$\begin{aligned} 2P(Z > 3.25) &= 2(1 - P(Z \leq 3.25)) \sim 2(1 - 0.9994) \\ &= 0.0012 = 0.12\% \,. \end{aligned}$$

Questo è il p-value del test. Si può accettare il test per $\alpha \leq 0.12\%$.

Esercizio 2. Consideriamo una moneta che è truccata con il 50% di probabilità. Se è truccata, esce sempre testa.

1. Osserviamo l'esito di un lancio: esce testa. Conoscendo questo esito, con che probabilità la moneta è effettivamente truccata?

2. Osserviamo l'esito di 3 lanci consecutivi indipendenti: esce testa in tutti e 3 i lanci. Conoscendo questo esito, con che probabilità la moneta è effettivamente truccata?

Soluzione. Descriviamo in ognuno dei due casi lo spazio campione come l'insieme delle $(n+1)$-ple composte da

$$(\tau, x_1, \ldots, x_n)\,,$$

dove $\tau = 0$ se la moneta è truccata, $\tau = 1$ se la moneta non è truccata e ogni x_i è l'esito dell'i-esimo lancio della moneta T o C.
1. Lo spazio campione è dato da

$$\begin{aligned} \Omega &= \{(\tau, x_1) : \sigma \in \{0,1\}\,, x_1 \in \{T, C\}\} \\ &= \{(0,T),\ (0,C),\ (1,T),\ (1,C)\}\,. \end{aligned}$$

L'evento a priori "la moneta è truccata" è

$$A = \{(0,T),\ (0,C)\}.$$

L'evento "esce testa" è

$$B = \{(0,T), (1,T)\}\,.$$

Il problema chiede di calcolare $\mathbf{P}(A|B)$. Per il teorema di Bayes abbiamo che:

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B|A)\mathbf{P}(A) + \mathbf{P}(B|\Omega \setminus A)\mathbf{P}(\Omega \setminus A)}$$

e tutte le probabilità nel membro a destra sono note. Infatti, a priori, la moneta è truccata con il 50% di probabilità, quindi

$$\mathbf{P}(A) = \frac{1}{2}\,, \quad \mathbf{P}(\Omega \setminus A) = \frac{1}{2}\,.$$

e

$$\mathbf{P}(B|A) = \mathbf{P}(\text{"esce testa se la moneta è truccata"}) = 1$$

$$\mathbf{P}(B|\Omega \setminus A) = \mathbf{P}(\text{"esce testa se la moneta non è truccata"}) = \frac{1}{2}\,.$$

Segue che

$$\mathbf{P}(A|B) = \frac{1 \cdot \frac{1}{2}}{1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}} = \frac{2}{3}\,.$$

Come soluzione alternativa, possiamo anche usare semplicemente la definizione di probabilità condizionata:

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(0,T)}{\mathbf{P}(B)}$$

e restano quindi calcolare le probabilità $\mathbf{P}(0,T)$ e $\mathbf{P}(1,T)$.
Se la moneta è truccata, esce sempre testa, ovvero, dalla definizione di probabilità condizionata

$$\mathbf{P}((0,T)|A) = 1 \implies \frac{\mathbf{P}(0,T)}{\mathbf{P}(A)} = 1 \implies \mathbf{P}(0,T) = \mathbf{P}(A) = \frac{1}{2}.$$

Se la moneta non è truccata, la probabilità che esca testa è $\frac{1}{2}$, quindi

$$\mathbf{P}((1,T)|\Omega \setminus A) = \frac{1}{2} \implies \frac{\mathbf{P}(1,T)}{\mathbf{P}(\Omega \setminus A)} = \frac{1}{2} \implies \mathbf{P}(1,T) = \frac{1}{2}\mathbf{P}(\Omega \setminus A) = \frac{1}{4}.$$

Segue che

$$\mathbf{P}(B) = \mathbf{P}(0,T) + \mathbf{P}(1,T) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$$

Concludiamo che

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(0,T)}{\mathbf{P}(B)} = \frac{1/2}{3/4} = \frac{2}{3}.$$

La soluzione è la stessa.
2. Lo spazio campione è dato da

$$\Omega = \{(\tau, x_1, x_2, x_3) : \sigma \in \{0,1\}, x_i \in \{T, C\} \text{ per } i = 1,2,3\}.$$

L'evento a priori "la moneta è truccata" è

$$A = \{(0, x_1, x_2, x_3) \ : \ x_i \in \{T, C\} \text{ per } i = 1,2,3\}.$$

L'evento "esce testa in tutti e 3 i lanci" è

$$B = \{(0,T,T,T), (1,T,T,T)\}.$$

Il problema chiede di calcolare $\mathbf{P}(A|B)$. Per il teorema di Bayes: Per il teorema di Bayes abbiamo che:

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B|A)\mathbf{P}(A) + \mathbf{P}(B|\Omega \setminus A)\mathbf{P}(\Omega \setminus A)}$$

e tutte le probabilità nel membro a destra sono note. Infatti, a priori, la moneta è truccata con il 50% di probabilità, quindi

$$\mathbf{P}(A) = \frac{1}{2}, \quad \mathbf{P}(\Omega \setminus A) = \frac{1}{2}.$$

e

$$\mathbf{P}(B|A) = \mathbf{P}(\text{“esce tre volte testa se la moneta è truccata”}) = 1$$

$$\mathbf{P}(B|\Omega \setminus A) = \mathbf{P}(\text{“esce tre volte testa se la moneta non è truccata”}) = \frac{1}{8}.$$

Segue che

$$\mathbf{P}(A|B) = \frac{1 \cdot \frac{1}{2}}{1 \cdot \frac{1}{2} + \frac{1}{8} \cdot \frac{1}{2}} = \frac{8}{9}.$$

Come soluzione alternativa, possiamo usare direttamente la definizione di probabilità condizionata:

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(0, T, T, T)}{\mathbf{P}(B)} .$$

Restano da calcolare le probabilità $\mathbf{P}(0, T, T, T)$ e $\mathbf{P}(1, T, T, T)$.
A priori, la moneta è truccata con il 50% di probabilità, quindi

$$\mathbf{P}(A) = \frac{1}{2}, \quad \mathbf{P}(\Omega \setminus A) = \frac{1}{2} .$$

Se la moneta è truccata esce sempre testa, quindi

$$\mathbf{P}((0, T, T, T)|A) = 1 \implies \frac{\mathbf{P}(0, T, T, T)}{\mathbf{P}(A)} = 1 \implies \mathbf{P}(0, T, T, T) = \mathbf{P}(A) = \frac{1}{2} .$$

Se la moneta non è truccata, la probabilità che esca testa tre volte consecutivamente è

$$\mathbf{P}((1, T, T, T)|\Omega \setminus A) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8} \implies \frac{\mathbf{P}(1, T, T, T)}{\mathbf{P}(\Omega \setminus A)} = \frac{1}{8}$$
$$\implies \mathbf{P}(1, T, T, T) = \frac{1}{8}\mathbf{P}(\Omega \setminus A) = \frac{1}{8}\frac{1}{2} = \frac{1}{16} .$$

Otteniamo che

$$\mathbf{P}(B) = \mathbf{P}(0, T, T, T) + \mathbf{P}(1, T, T, T) = \frac{1}{2} + \frac{1}{16} = \frac{9}{16} .$$

Concludiamo che

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(0, T, T, T)}{\mathbf{P}(B)} = \frac{1/2}{9/16} = \frac{8}{9} .$$

La soluzione è la stessa.

Esercizio 3. Siano $\lambda, \mu \in \mathbb{R}$ due parametri e sia (X, Y) una variabile aleatoria bidimensionale discreta con le seguenti probabilità congiunte:

X \ Y	0	1	2
0	λ	1/8	1/16
1	1/4	μ	1/4

1. Calcolare $\mathrm{Cov}(X, Y)$.
2. Per quali valori di λ e μ le variabili aleatorie X e Y sono indipendenti?

Soluzione. Innanzitutto troviamo la condizione soddisfatta da λ e μ affinché le probabilità congiunte sommino a 1:

$$1 = \lambda + \frac{1}{8} + \frac{1}{16} + \frac{1}{4} + \mu + \frac{1}{4} = \lambda + \mu + \frac{11}{16} \implies \lambda + \mu = 1 - \frac{11}{16} = \frac{5}{16}$$
$$\implies \lambda = \frac{5}{16} - \mu .$$

1. Ricordiamo che $\mathrm{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$. Calcoliamo questi valori attesi:

$$\begin{aligned}\mathbf{E}(XY) &= \sum_{i=0}^{1}\sum_{j=0}^{2} x_i y_j \mathbf{P}(X = x_i, Y = y_j) = \\ &= 0\cdot 0\cdot\lambda + 0\cdot 1\cdot\frac{1}{8} + 0\cdot 2\cdot\frac{1}{16} + 1\cdot 0\cdot\frac{1}{4} + 1\cdot 1\cdot\mu + 1\cdot 2\cdot\frac{1}{4} = \\ &= \mu + \frac{1}{2}\,,\end{aligned}$$

$$\mathbf{E}(X) = \sum_{i=0}^{1} x_i\mathbf{P}(X = x_i) = 0\cdot\mathbf{P}(X=0) + 1\cdot\mathbf{P}(X=1) = \frac{1}{4} + \mu + \frac{1}{4} = \mu + \frac{1}{2}\,,$$

$$\begin{aligned}\mathbf{E}(Y) &= \sum_{j=0}^{2} y_j\mathbf{P}(Y = y_j) = 0\cdot\mathbf{P}(Y=0) + 1\cdot\mathbf{P}(Y=1) + 2\cdot\mathbf{P}(Y=2) \\ &= \frac{1}{8} + \mu + 2\Big(\frac{1}{16} + \frac{1}{4}\Big) = \mu + \frac{3}{4}\,.\end{aligned}$$

Quindi

$$\mathrm{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = \mu + \frac{1}{2} - \Big(\mu + \frac{1}{2}\Big)\Big(\mu + \frac{3}{4}\Big) = \Big(\mu + \frac{1}{2}\Big)\Big(\frac{1}{4} - \mu\Big)\,.$$

2. Ricordiamo che se X e Y sono indipendenti, allora $\mathrm{Cov}(X,Y) = 0$. Questo è verificato per $\mu = -\frac{1}{2}$ e $\mu = \frac{1}{4}$. Il primo valore non è ammissibile perché negativo. Il secondo dà, insieme alla condizione trovata all'inizio dell'esercizio, $\lambda = \frac{5}{16} - \mu = \frac{5}{16} - \frac{1}{4} = \frac{1}{16}$. La tabella dei valori è in questo caso

X \ Y	0	1	2	
0	1/16	1/8	1/16	1/4
1	1/4	1/4	1/4	3/4
	5/16	3/8	5/16	1

Tuttavia le due variabili non sono indipendenti anche in questo caso poiché, ad esempio,

$$\mathbf{P}(X=0)\mathbf{P}(Y=0) = \frac{1}{4}\frac{5}{16} = \frac{5}{64} \neq \frac{1}{16} = \mathbf{P}(X=0, Y=0)\,.$$

Si ricordi che (X, Y indipendenti) implica soltanto ($\mathrm{Cov}(X,Y) = 0$), non sono proposizioni equivalenti!

Esercizio 4. Si consideri il seguente campione di dati:

28	35	55	15	40	35	29	29	65	29	33	31	19	37

Trovare eventuali dati anomali e sospetti e tracciare un box-plot.

Soluzione. Ordiniamo i dati

15	19	28	29	29	29	31	33	35	35	37	40	55	65

Il primo quartile Q_1 è il numero tale che $1/4$ dei dati del campione è a sinistra di Q_1. Poiché $(14+1)/4 = 15/4 = 3 + 3/4$ non è intero, si ottiene Q_1 tramite la formula[1] $Q_1 = (1-3/4) \cdot 28 + 3/4 \cdot 29 = 28.75$
Poiché $15/2 = 7 + 1/2$ si ottiene Q_2 come media dei valori ai posti 7 e 8, cioè $Q_2 = \frac{31+33}{2} = 32$.
Infine, poiché $15\frac{3}{4} = 11{+}1/4$ si ottiene Q_3 tramite la formula $Q_3 = (1{-}1/4){\cdot}37{+}1/4{\cdot}40 = 37.75$.

L'interquartile range è dato da $IQR = Q_3 - Q_1 = 9$.
I dati anomali cadono negli intevalli

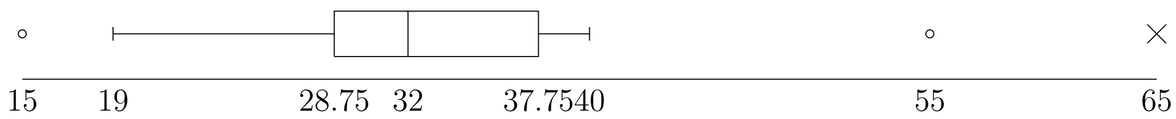
$$(-\infty, Q_1 - 3IQR] \cup [Q_3 + 3IQR, +\infty) = (-\infty, 1.75] \cup [64.75, +\infty)$$

quindi 65 è un dato anomalo.
I dati sospetti cadono negli intervalli

$$(Q_1 - 3IQR, Q_1 - 1.5IQR] \cup [Q_3 + 1.5IQR, Q_3 + 3IQR) = (1.75, 15.25] \cup [51.25, 64.75)$$

quindi 15 e 55 sono dati sospetti.

Segue il box plot.



Esercizio 5. Siano $a \in (0,1)$ e $b \in \mathbb{R}$ due parametri e sia X una variabile aleatoria distribuita con una densità di probabilità $f\colon \mathbb{R} \to \mathbb{R}$ definita da

$$f(x) := \begin{cases} b\,a^x & \text{per } x \geq 0 \\ 0 & \text{altrimenti.} \end{cases}$$

Trovare a e b tali che $\mathbf{E}(X) = 2$.

Soluzione. Poiché $a \in (0,1)$, abbiamo $\ln a < 0$. Tenendo conto di questo, possiamo imporre che la densità di probabilità abbia integrale 1:

$$1 = \int_{\mathbb{R}} f(x)\,\mathrm{d}x = \int_0^{+\infty} ba^x\,\mathrm{d}x = \int_0^{+\infty} be^{x\ln a}\,\mathrm{d}x = \frac{b}{\ln a} e^{x\ln a}\Big|_0^{+\infty} = -\frac{b}{\ln a} \implies b = -\ln a\,.$$

Imponiamo che il valore atteso sia 2. Integrando per parti[2]:

$$\begin{aligned} 2 = \mathbf{E}(X) &= \int_{\mathbb{R}} xf(x)\,\mathrm{d}x = \int_0^{+\infty} bxa^x\,\mathrm{d}x = \int_0^{+\infty} bxe^{x\ln a}\,\mathrm{d}x \\ &= \frac{b}{\ln a} xe^{x\ln a}\Big|_0^{+\infty} - \int_0^{+\infty} \frac{b}{\ln a} e^{x\ln a}\,\mathrm{d}x = -\frac{b}{|\ln a|^2} e^{x\ln a}\Big|_0^{+\infty} \\ &= \frac{b}{|\ln a|^2} \end{aligned}$$

[1] è accettata anche la soluzione in cui Q_1 e Q_3 sono calcolate con la media dei due valori.
[2] $(\frac{b}{\ln a} xe^{x\ln a})' = \frac{b}{\ln a} e^{x\ln a} + bxe^{x\ln a}$

da cui ricaviamo

$$b = 2|\ln a|^2 \,.$$

Risulta quindi

$$-\ln a = 2|\ln a|^2 \implies |\ln a| = 2|\ln a|^2 \implies |\ln a| = \frac{1}{2} \implies -\ln a = \frac{1}{2} \implies a = e^{-1/2},$$

$$b = -\ln a = -\ln e^{-1/2} = \frac{1}{2} \,.$$

In conclusione,

$$f(x) := \begin{cases} \frac{1}{2}e^{-x/2} & \text{per } x \geq 0 \\ 0 & \text{altrimenti.} \end{cases}$$