

Esercizi di statistica descrittiva

a cura di Giovanni M. Marchetti

1 marzo 2019

Questi esercizi hanno la struttura degli esercizi dell'esame scritto di Statistica, cioè sono di 3 tipi: vero/falso, scelta multipla, risposta aperta.

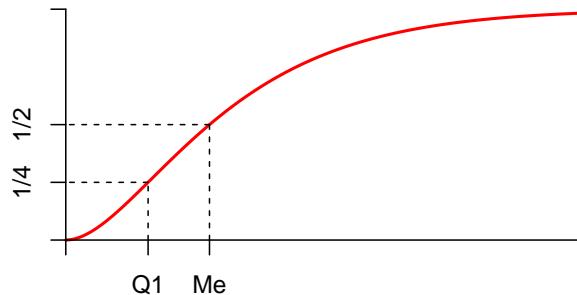
1. Quale delle seguenti affermazioni descrive uno svantaggio nell'utilizzo del campo di variazione come misura di variabilità o di dispersione?
 - (a) Non è misurato nella stessa unità di misura dei dati.
 - (b) È molto influenzato dai valori anomali.
 - (c) Produce una misura della variabilità troppo elevata.
 - (d) Si può calcolare solo per variabili continue.

Soluzione.

- (a) Falso: è $x_{\max} - x_{\min}$ e quindi è misurato nella stessa unità di misura dei dati.
 - (b) Vero: perché ovviamente basta un solo valore o molto grande o molto piccolo per farlo cambiare a differenza degli altri indici di variabilità.
 - (c) Falso: è una frase senza senso.
 - (d) Falso: anche la varianza si può calcolare solo per variabili continue e non per questo è uno svantaggio.
2. Il primo quartile è sempre minore o uguale alla mediana,

- (a) Solo se la media è maggiore di zero
 - (b) Solo se la moda è maggiore della mediana
 - (c) SI è vero
 - (d) NO è falso

Soluzione. Per definizione $P(X \leq Q_1) = 0.25$ mentre $P(X \leq Me) = 0.5$, come si vede dalla funzione di ripartizione.



Quindi è chiaro che Q_1 non può essere più grande di Me altrimenti la probabilità di essere inferiore alla mediana sarebbe minore di 0.25.

3. Quale delle seguenti affermazioni è vera?

- (a) La covarianza è sempre maggiore del coefficiente di correlazione.
- (b) La covarianza può essere uguale al coefficiente di correlazione.
- (c) Né la covarianza né il coefficiente di correlazione possono essere pari a zero.
- (d) Il coefficiente di correlazione è sempre più grande della covarianza.

Soluzione.

- (a) Falso. La covarianza cambia al variare dell'unità di misura e quindi si può rendere più piccola del coefficiente di correlazione.
 - (b) Vero. Perché il coefficiente di correlazione è la covarianza fra le due variabili standardizzate.
 - (c) Falso. Evidentemente altrimenti non esisterebbe il concetto di incorrelazione.
 - (d) Falso. La covarianza cambia al variare dell'unità di misura e quindi si può rendere più grande del coefficiente di correlazione.
4. Il coefficiente di correlazione lineare è sempre compreso nell'intervallo $[0, 1]$. Vero o Falso?

Soluzione. Falso. È sempre compreso in $[-1, 1]$.

5. Il diagramma a barre è comunemente utilizzato per descrivere dati qualitativi. Vero o Falso?

Soluzione. Vero.

6. Se due insiemi di dati hanno lo stesso campo di variazione, allora:

- (a) I due insiemi hanno la stessa varianza.
- (b) La distanza tra il valore più piccolo e il valore più grande nei due insiemi ha lo stesso valore.
- (c) I due insiemi hanno lo stesso scarto interquartile.
- (d) il valore più piccolo e il valore più grande nei due insiemi sono gli stessi.

Soluzione. A) Falso in generale. Per esempio $x_1 = 0, x_2 = 1$ ha campo di variazione 1. Anche $y_1 = 1, y_2 = 1.5, y_3 = 2$ ha campo di variazione 1. Tuttavia $\text{var}(x) = 1/2$ e $\text{var}(y) = 1/4$. B) Vero per definizione. C) Falso come per A). D) Falso: non è necessario che $x_{\max} = y_{\max}$ e $x_{\min} = y_{\min}$ per ottenere lo stesso campo di variazione.

7. La differenza interquartile corrisponde al campo di variazione del 50% centrale dei dati. Vero o Falso?

Soluzione. Vero: la differenza interquartile è $Q_3 - Q_1$ e

$$P(Q_1 \leq X \leq Q_3) = 0.5$$

Quindi è il campo di variazione del 50% centrale dei dati.

8. In una ricerca di marketing, ai consumatori è stato dato uno fra 4 marche di detersivo per lavastoviglie ed è stato chiesto di usarlo per un mese. Allo scadere del periodo è stato chiesto un giudizio sul detersivo utilizzato in termini di qualità complessiva. I risultati sono i seguenti.

Marca	Giudizio					Totale
	Mediocre	Medio	Discreto	Buono		
A	5	17	11	10	43	
B	14	26	8	18	66	
C	10	23	11	17	61	
D	11	19	7	5	42	
Totale	40	85	37	50	212	

Qual è la proporzione di consumatori che ha valutato il proprio detersivo discreto o buono?

- (a) 0.39
- (b) 0.49
- (c) 0.29
- (d) 0.41

Soluzione.

Casi possibili 212. Casi favorevoli: $37 + 50 = 87$. Quindi la proporzione è $87/212 = 0.410$.

9. Fra i consumatori che hanno valutato il prodotto A, qual è la proporzione che lo ha giudicato mediocre?

- (a) 0.189
- (b) 0.116
- (c) 0.125
- (d) 0.024

Soluzione. Casi possibili: 43. Casi favorevoli: 5. Quindi la proporzione è $0.1163 \simeq 0.116$.

10. L'operazione di standardizzazione di una variabile statistica consiste nel dividere gli scarti dalla media aritmetica per la varianza. Vero o Falso?

Soluzione. Falso: $z = (x - \mu)/\sigma$ NON $z = (x - \mu)/\sigma^2$.

- 20.** Per tutti gli individui di un collettivo viene rilevato il codice di avviamento postale del comune di residenza. Quale delle seguenti definizioni è appropriata per questi dati?

- (a) Dati quantitativi
- (b) Dati qualitativi
- (c) Serie storica
- (d) Dati numerici

Soluzione

- (a) Falso: non si possono eseguire operazioni algebriche con un senso. Solo verificare se una altro codice è uguale o diverso.

- (b) Vero.
(c) Falso. Non ci sono riferimenti temporali.
(d) Falso: vedi A).
11. Se $P(A) = 0.25$ e $P(B) = 0.75$, allora gli eventi A e B sono collettivamente esaustivi. Vero o Falso?
- Soluzione.** Falso. Esempio: in un mazzo di carte sia A “pesco una carta di cuori” e B “pesco una carta non di picche”. A e B non sono esaustivi.
12. Si considerino due investimenti possibili aventi lo stesso tasso di rendimento atteso. Negli ultimi mesi l’investimento A ha avuto un prezzo medio di chiusura di 14.00 e una deviazione standard di 4.00. L’investimento B un prezzo medio di chiusura di 58.00 e una deviazione standard di 15.00. Il valore di mercato dell’investimento A ha una variabilità relativa maggiore di quella dell’investimento B . Vero o Falso?
- Soluzione.** Per valutare la variabilità relativa si usa il coefficiente di variazione $CV = \sigma/\mu$. Quindi
- $$CV_A = 4/140.286, \quad CV_B = 15/58 = 0.259$$
- Quindi $CV_A > CV_B$. Vero.
13. Considera la seguente distribuzione di frequenza.

Modalità	Frequenza	Freq. % Cumulate
fino a 500	1	4
500 - 1000	*	64
1000 - 3000	4	80
3000 - 4000	3	92
4000 - 5000	1	96
più di 5000	1	100

Qual è la frequenza mancante nella posizione occupata dall’asterisco?

- (a) 15
(b) 16
(c) 3
(d) 25

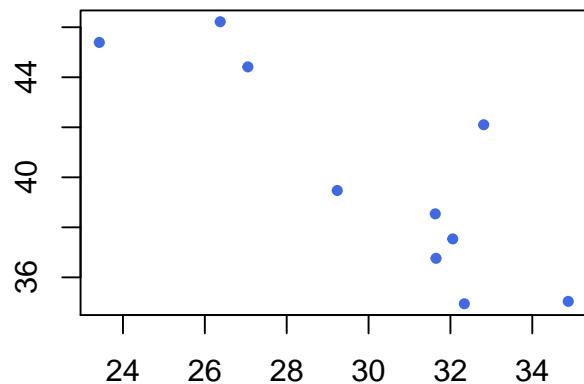
Soluzione.

La frequenza relativa della classe $(500, 1000]$ è $0.64 - 0.04 = 0.6$. La frequenza assoluta della stessa classe è x e la frequenza totale è $10 + x$. Quindi

$$0.6 = x/(10 + x)$$

da cui si ottiene l’equazione $0.6(10 + x) = x$ che risolta dà $x = 15$.

14. Sia dato lo scatter seguente.



Quale tra i valori seguenti potrebbe essere il coefficiente di correlazione per le due variabili?

- (a) -0.8
- (b) 0.0
- (c) -0.3
- (d) -1.

Soluzione. È chiaro che il coefficiente di correlazione deve essere negativo. Ovviamente si scartano la B) (incorelazione) e la D) (perfetto allineamento). Tra A) e C) si sceglie A) perché l'associazione lineare è più vicina a -1 che a 0. Qui ci vuole un po' d'occhio.

15. La varianza di $X + c$, dove c è una costante, è pari alla varianza di X più la costante. Vero o Falso?

Soluzione. Falso: $\text{var}(X + c) = \text{var}(X)$.