Statistica Sociale

Soluzioni esempi di prova d'esame

Ho inserito per i vero/falso dei commenti utili come appunti per lo studio. Nel compito non è richiesto nessuno tipo di commento in questa parte. Negli esercizi ho aggiunto sia una spiegazione che un breve commento dei risultati.

Esame 1

Esercizio 1 — Domande Vero/Falso

Traccia. Rispondere Vero o Falso alle seguenti affermazioni (valgono 2 punti; le risposte errate non pesano sul punteggio finale):

1. La numerosità campionaria dipende dalla numerosità della popolazione indagata.
2. Un indice di variabilità assume valore minimo se e solo se tutte le unità assumono valore zero.
3. Nella suddivisione in classi, le classi possono avere ampiezza uguale o diversa.
4. La frequenza relativa è un valore compreso tra 0 e 1.
5. L'istogramma è il grafico adatto a rappresentare variabili qualitative ordinali.
6. Il grafico di dispersione è utile per avere evidenza di una relazione lineare tra due variabili.
7. Due variabili sono associate se le distribuzioni condizionate della tabella di contingenza sono uguali.
8. La media rappresenta sempre meglio della mediana l'ordine di grandezza di una variabile.
9. In un campione, per calcolare la varianza si divide la somma degli scarti dalla media per $n - 1$.
10. Il coefficiente di correlazione può essere superiore a 1.

"La numerosità campionaria dipende dalla numerosità della popolazione indagata."

Risposta: FALSO. La numerosità campionaria n non dipende dalla dimensione N della popolazione (salvo quando N è molto piccola).

"Un indice di variabilità assume il valore minimo se e solo se tutte le unità assumono il valore zero."

Risposta: FALSO. Un indice di variabilità (come la varianza o la deviazione standard) è minimo e vale 0, quando tutte le unità assumono lo *stesso* valore, non necessariamente zero. Se tutti i valori sono uguali a 5, la varianza è comunque 0 perché non esiste alcuna dispersione rispetto alla media.

"Nella suddivisione in classi, le classi possono avere ampiezza uguale o diversa."

Risposta: VERO. Non esiste alcun vincolo che imponga classi di uguale ampiezza. Classi di ampiezza diversa sono anzi spesso preferibili quando la distribuzione è molto asimmetrica: si usano classi strette dove i dati sono concentrati e classi più ampie dove sono rari. Fondamentale: le classi devono essere *esaustive* e *mutuamente esclusive*.

"La frequenza relativa è un valore compreso tra 0 e 1."

Risposta: VERO. La somma delle frequenze relative su tutte le modalità vale 1: $\sum_i f_{rel,i} = 1$.

"L'istogramma è il grafico adatto a rappresentare variabili qualitative ordinali."

Risposta: FALSO. L'istogramma è adatto a **variabili quantitative continue** raggruppate in classi. Per le variabili qualitative ordinali (es.: titolo di studio, giudizio) il grafico appropriato è il **grafico a barre**, con rettangoli separati per sottolineare la natura discreta delle modalità.

"Il grafico di dispersione è utile per avere evidenza di una relazione lineare tra due variabili."

Risposta: VERO. Il grafico di dispersione (*scatterplot*) riporta, in un piano cartesiano, le coppie (x_i, y_i) per ciascuna unità. Se i punti si dispongono attorno a una retta, si ha una evidenza visiva di una relazione lineare. È lo strumento esplorativo fondamentale prima di calcolare il coefficiente di correlazione (r).

"Due variabili sono associate se le distribuzioni condizionate della tabella di contingenza sono uguali."

Risposta: FALSO. È esattamente il contrario. Due variabili **non sono associate** (sono indipendenti) se le distribuzioni condizionate sono *uguali* tra loro e uguali alla distribuzione marginale. L'**associazione** si rileva quando le distribuzioni condizionate *differiscono*: se la percentuale di occupati varia significativamente al variare del titolo di studio, allora le due variabili sono associate.

"La media rappresenta sempre meglio della mediana l'ordine di grandezza di una variabile."

Risposta: FALSO. La media aritmetica è sensibile ai valori estremi: un singolo outlier può spostarla considerevolmente. La mediana è **robusta**: non risente dei valori anomali perché dipende solo dalla posizione centrale dei dati ordinati.

"In un campione, per calcolare la varianza si divide la somma degli scarti dalla media per $n-1$."

Risposta: VERO. La varianza **campionaria** corretta è:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Si divide per $n-1$ (anziché per n) per ottenere uno **stimatore non distorto** della varianza della popolazione s^2.

"Il coefficiente di correlazione può essere superiore a 1."

Risposta: FALSO. Il coefficiente di correlazione di Pearson è sempre compreso in $[-1, +1]$

Esercizio 2 — Risultati del test

Traccia. I risultati di un test sono riassunti nella tabella seguente (i punti assegnabili erano 100 e un/a concorrente poteva totalizzare 0):

Classi	[0–50)	[50–60)	[60–70)	[70–90)	[90–100)	[100]
Frequenze	7	12	27	15	3	2

a. Rappresentare i dati mediante il grafico più appropriato.

b. Calcolare la proporzione di concorrenti che superano il test, assumendo che il voto minimo sia 80.

c. Calcolare il voto medio, la classe mediana e quella modale.

Dati: $n = 7 + 12 + 27 + 15 + 3 + 2 = \mathbf{66}$.

a) Grafico appropriato.

Commento. Trattandosi di una variabile quantitativa continua con classi di *ampiezza diversa*, il grafico corretto è l'**istogramma delle densità di frequenza**.

Appunti. l'altezza di ogni rettangolo è la densità $d_i = f_i/a_i$. In questo modo l'area di ogni rettangolo rimane proporzionale alla frequenza anche con classi di diversa ampiezza.

Classe	f_i	Ampiezza a_i	Densita' $d_i = f_i/a_i$
[0–50)	7	50	$7/50 = 0{,}14$
[50–60)	12	10	$12/10 = 1{,}20$
[60–70)	27	10	$27/10 = \mathbf{2{,}70}$
[70–90)	15	20	$15/20 = 0{,}75$
[90–100)	3	10	$3/10 = 0{,}30$
[100]	2	1	$2/1 = 2{,}00$

b) Proporzione di concorrenti con voto ≥ 80.

Commento. La classe $[70, 90)$ ha ampiezza 20 e contiene 15 concorrenti. Ipotizzando distribuzione *uniforme* all'interno della classe, la quota in $[80, 90)$ (sottointervallo di ampiezza 10) è:

$$n_{[80,90)} = 15 \times \frac{10}{20} = 7{,}5$$

Il totale con voto ≥ 80 comprende anche le classi $[90, 100)$ e $[100]$:

$$n_{\geq 80} = 7{,}5 + 3 + 2 = 12{,}5$$

$$\text{voto} \geq 80 = \frac{12{,}5}{66} \approx \mathbf{0{,}189} \quad (18{,}9\%)$$

c) Media, classe mediana, classe modale.

Spiegazione. *Voto medio.* Per dati raggruppati in classi si usa il punto medio m_i di ciascuna classe come valore rappresentativo. Si costruisce la tavola dei prodotti $f_i \cdot m_i$:

Classe	f_i	Punto medio m_i	$f_i \cdot m_i$
[0–50)	7	25	$7 \times 25 = 175$
[50–60)	12	55	$12 \times 55 = 660$
[60–70)	27	65	$27 \times 65 = 1755$
[70–90)	15	80	$15 \times 80 = 1200$
[90–100)	3	95	$3 \times 95 = 285$
[100]	2	100	$2 \times 100 = 200$
Totale	**66**		**4275**

$$\bar{x} = \frac{\sum f_i \cdot m_i}{n} = \frac{4275}{66} \approx \mathbf{64{,}77}$$

Spiegazione. *Classe mediana.* La mediana è il valore che divide la distribuzione in due metà uguali. Con $n = 66$ osservazioni, la mediana e' la media del 33° e del 34° valore nell'ordinamento. Si costruisce la tabella delle frequenze cumulate F_i:

Classe	f_i	F_i	Contiene la mediana?
[0–50)	7	7	No: $F_1 = 7 < 33$
[50–60)	12	19	No: $F_2 = 19 < 33$
[60–70)	27	**46**	**Si**: $F_2 = 19 < 33 \leq 46 = F_3$
[70–90)	15	61	—
[90–100)	3	64	—
[100]	2	66	—

Commento. Sia il 33° che il 34° valore cadono nella classe $[60, 70)$.

Commento. *Classe modale.*Con classi di ampiezza *diversa* non si può identificare la classe modale guardando la frequenza assoluta più alta (sarebbe fuorviante), ma occorre usare la **densità di frequenza massima**. Confrontando le densità già calcolate al punto a):

Classe	Densità' d_i	Classe modale?
[0–50)	0,14	No
[50–60)	1,20	No
[60–70)	**2,70**	**Si** (massima)
[70–90)	0,75	No
[90–100)	0,30	No
[100]	2,00	No

Commento. Classe modale: [**60**–**70**)

Esercizio 3 — Occupazione e titolo di studio

Traccia. Nella tabella sono riportate le frequenze percentuali di un gruppo di 300 donne secondo il titolo di studio e l'occupazione (Si/No):

Occupate	**Laurea**	**Diploma**	**Lic. media**
Si	45	35	20
No	30	20	50

Sapendo che le occupate rappresentano i due terzi del totale:

a. Indicare se ci sono più occupate tra le laureate o tra le diplomate.

b. Si può dire che esiste un'associazione tra occupazione e titolo di studio? Motivare la risposta.

c. Calcolare la percentuale totale di donne non occupate.

Spiegazione. Le righe della tabella fornita sono **distribuzioni condizionate**: le percentuali nelle righe "Si" e "No" sono calcolate rispettivamente sul totale delle occupate e delle non occupate.

Totale donne: $N = 300$.Poiché' le occupate sono i $\frac{2}{3}$ del totale:

$$N_{\text{occ}} = \frac{2}{3} \times 300 = \mathbf{200}, \qquad N_{\text{non occ}} = \frac{1}{3} \times 300 = \mathbf{100}$$

Le frequenze assolute si ottengono applicando le percentuali ai rispettivi totali:

	Laurea	**Diploma**	**Lic. media**	**Totale**
Occupate (%)	45	35	20	100
Non occupate (%)	30	20	50	100
Occupate (n)	$200 \times 0{,}45 = 90$	$200 \times 0{,}35 = 70$	$200 \times 0{,}20 = 40$	200
Non occupate (n)	$100 \times 0{,}30 = 30$	$100 \times 0{,}20 = 20$	$100 \times 0{,}50 = 50$	100
Totale	**120**	**90**	**90**	**300**

a) Più occupate tra laureate o diplomate?

	Laureate	**Diplomate**	**Lic. media**
Occupate (n)	90	70	40
Totale (n)	120	90	90
Tasso di occupazione	$90/120 = \mathbf{75{,}0\%}$	$70/90 = \mathbf{77{,}8\%}$	$40/90 = 44{,}4\%$

Commento. Qui abbiamo due risposte corrette: **In numero assoluto**: più occupate tra le **laureate** ($90 > 70$), perché sono anche in totale di più.

In proporzione: la quota è più alta tra le **diplomate** ($77{,}8\% > 75{,}0\%$).

Spiegazione. In statistica sociale il confronto proporzionale è più significativo: **le diplomate hanno un tasso di occupazione leggermente superiore a quello delle laureate**; le licenziate medie mostrano il tasso più basso (44,4%).

b) Associazione tra occupazione e titolo di studio?

Si confrontano le distribuzioni condizionate di "occupazione" per ciascun titolo di studio:

	Laureate	**Diplomate**	**Lic. media**
Occupate	$90/120 = 75{,}0\%$	$70/90 = 77{,}8\%$	$40/90 = 44{,}4\%$
Non occupate	$30/120 = 25{,}0\%$	$20/90 = 22{,}2\%$	$50/90 = 55{,}6\%$
Totale	100%	100%	100%

Spiegazione. Se le due variabili fossero indipendenti, le distribuzioni condizionate sarebbero tutte uguali. Invece variano sensibilmente: dal 44,4% delle licenziate medie al 77,8% delle diplomate.

Commento. Le distribuzioni condizionate *differiscono* tra i gruppi. **Esiste associazione** tra titolo di studio e occupazione.

Commento. **c) Percentuale totale di non occupate.**

$$\frac{N_{\text{non occ}}}{N} = \frac{100}{300} = \frac{1}{3}$$

Percentuale non occupate = **33,3**%

Esame 2

Esercizio 1 — Domande Vero/Falso

Traccia. Rispondere Vero o Falso alle seguenti affermazioni (valgono 2 punti; le risposte errate non pesano sul punteggio finale):

1. Le unità di una popolazione statistica sono necessariamente delle persone fisiche.
2. Poiché non è noto chi risponderà, il campionamento volontario è un campionamento probabilistico.
3. Le modalità di una variabile nominale per due unità diverse possono essere una maggiore dell'altra.
4. Le frequenze percentuali sono preferibili alle frequenze relative per fare confronti in campioni di numerosità diversa.
5. Nel grafico a barre, i rettangoli non possono essere distanziati.
6. Se il coefficiente di correlazione tra X e Y è pari a 0, X e Y sono indipendenti.
7. Il primo quartile è uguale al 25-esimo percentile.
8. La mediana non è influenzata dalla presenza di un valore molto piccolo o molto grande rispetto agli altri.
9. Se X e Y sono indipendenti, le distribuzioni condizionate di Y rispetto a X sono uguali alla marginale di Y.
10. Il principio della suddivisione in classi si può applicare anche alle variabili qualitative.

"Le unità di una popolazione statistica sono necessariamente delle persone fisiche."

Risposta: FALSO. Le **unità statistiche** sono gli oggetti elementari su cui vengono rilevate le caratteristiche di interesse. Possono essere persone fisiche, ma anche imprese, comuni, famiglie, atti giudiziari, animali, transazioni, ecc. Fondamentale: ciò che le definisce è il fatto di essere *distinte*, *identificabili* e *omogenee* rispetto allo scopo dell'indagine .

"Poichè non è noto chi risponderà, il campionamento volontario è un campionamento probabilistico."

Risposta: FALSO. Il campionamento **probabilistico** richiede che ogni unità della popolazione abbia una probabilità *nota e positiva* di essere estratta.Nel campionamento **volontario** (es.: sondaggi online aperti a tutti), le unità si autoselezionano: non c'è un meccanismo di estrazione casuale controllato dal ricercatore.

"Le modalità di una variabile nominale per due unità diverse possono essere una maggiore dell'altra."

Risposta: FALSO. Una variabile **nominale** classifica le unità in categorie *senza* alcun ordinamento. Le modalità sono solo etichette (es.: colore degli occhi). Non ha senso dire che "blu è maggiore di verde".

"Le frequenze percentuali sono preferibili alle frequenze relative per fare confronti in campioni di numerosita' diversa."

Risposta: FALSO. Frequenze percentuali e frequenze relative sono *perfettamente equivalenti*: $f\% = f_{rel} \times 100$. Entrambe normalizzano le frequenze assolute rispetto alla numerosità totale, rendendole confrontabili tra campioni di dimensioni diverse.

"Nel grafico a barre, i rettangoli non possono essere distanziati."

Risposta: FALSO. Nel **grafico a barre** i rettangoli sono *tipicamente separati* tra loro da uno spazio, per sottolineare la natura discreta o categoriale della variabile (qualitative o quantitative discrete). Al contrario, nell'**istogramma** i rettangoli sono adiacenti perché rappresentano una variabile quantitativa continua con classi contigue.

"Se il coefficiente di correlazione tra X e Y è pari a 0, X e Y sono indipendenti."

Risposta: FALSO. $r = 0$ indica assenza di relazione **lineare** tra X e Y, ma non implica indipendenza statistica.

"Il primo quartile è uguale al 25-esimo percentile."

Risposta: VERO. Il **primo quartile** Q_1 è per definizione il valore sotto al quale cade il 25% delle osservazioni, che corrisponde al **25° percentile** P_{25}.

"La mediana non è influenzata dalla presenza di un valore molto piccolo o molto grande rispetto agli altri."

Risposta: VERO. La mediana è una misura **robusta** agli outlier: dipende solo dalla posizione centrale dei dati ordinati, non dai valori numerici degli estremi. Questa proprietà la rende preferibile alla media in presenza di distribuzioni asimmetriche o valori anomali.

"Se X e Y sono indipendenti, le distribuzioni condizionate di Y rispetto a X sono uguali alla marginale di Y."

Risposta: VERO. Questa è la **definizione operativa di indipendenza statistica**: X e Y sono indipendenti se e solo se conoscere il valore di X non modifica la distribuzione di Y. In una tabella di contingenza, l'indipendenza si manifesta quando le distribuzioni condizionate di Y al variare delle modalità di X sono tutte uguali alla distribuzione marginale di Y.

"Il principio della suddivisione in classi si può applicare anche alle variabili qualitative."

Risposta: FALSO. La suddivisione in classi si applica a **variabili quantitative continue**, dove i valori possibili sono infiniti e ha senso definire intervalli numerici. Per le variabili qualitative le modalità sono già categorie discrete e non esiste una scala numerica su cui costruire intervalli.

Esercizio 2 — Spesa sanitaria

Traccia. In un'indagine sulla spesa sanitaria si sono rilevate le spese straordinarie (per emergenze). La tabella riporta, per alcune classi di età dei rispondenti, la numerosità delle classi e le frequenze condizionate per la variabile dicotomica "nessuna spesa straordinaria" (Spesa $= 0$) e "spesa straordinaria positiva" (Spesa > 0):

Classi di età	40–50	50–60	60–70	70–80	80–90
Rispondenti	3000	2100	2500	1800	600
Spesa $= 0$	0,71	0,67	0,58	0,50	0,44
Spesa > 0	?	?	?	?	?

a. Costruire la tabella delle frequenze assolute congiunte per le due variabili.

b. Rappresentare graficamente i dati in modo da confrontare la distribuzione per età secondo la presenza o assenza di spesa.

c. Valutare se vi à associazione tra le due variabili.

a) Frequenze assolute congiunte.

Le frequenze condizionate di Spesa> 0 si ottengono come complemento a 1 di quelle di Spesa$= 0$. Le frequenze assolute si calcolano moltiplicando per la numerosità della classe:

Classe eta'	n_i	**Spesa** $= 0$	**Spesa** > 0
40–50	3000	$3000 \times 0{,}71 = 2130$	$3000 \times 0{,}29 = 870$
50–60	2100	$2100 \times 0{,}67 = 1407$	$2100 \times 0{,}33 = 693$
60–70	2500	$2500 \times 0{,}58 = 1450$	$2500 \times 0{,}42 = 1050$
70–80	1800	$1800 \times 0{,}50 = 900$	$1800 \times 0{,}50 = 900$
80–90	600	$600 \times 0{,}44 = 264$	$600 \times 0{,}56 = 336$
Totale	**10000**	**6151**	**3849**

b) Rappresentazione grafica.

Commento. Per confrontare la distribuzione della spesa tra classi di età, il grafico più appropriato è un **grafico a barre sovrapposte**: sull'asse x le classi di età, sull'asse y le frequenze condizionate percentuali.

c) Associazione tra età e spesa.

Commento. Per valutare l'associazione si confrontano le distribuzioni condizionate (già fornite nel testo) con la distribuzione **marginale** di Spesa. Se le due variabili fossero indipendenti, le proporzioni condizionate sarebbero uguali alla marginale per ogni classe di età.

Distribuzione marginale di Spesa (calcolata sui totali):

$$p(\text{Spesa} = 0) = \frac{6151}{10000} = 0{,}615, \qquad p(\text{Spesa} > 0) = \frac{3849}{10000} = 0{,}385$$

Confronto tra proporzione condizionata di Spesa> 0 e proporzione marginale 0,385:

Classe età	$p(\text{Spesa} > 0 \mid \text{eta'})$	Marginale	Differenza
40–50	0,29	0,385	$-0{,}095$
50–60	0,33	0,385	$-0{,}055$
60–70	0,42	0,385	$+0{,}035$
70–80	0,50	0,385	$+0{,}115$
80–90	0,56	0,385	$+0{,}175$

Le differenze non sono nulle e crescono sistematicamente con l'età.

> Commento. Le distribuzioni condizionate *differiscono* dalla marginale e cambiano sistematicamente con l'età. **Esiste associazione**: la proprobabilità sostenere spese sanitarie straordinarie cresce con l'avanzare dell'età 29% nella classe 40–50 a 56% nella classe 80–90).

Esercizio 3 — Prezzi nei negozi

> **Traccia.** I prezzi (in euro) di un determinato prodotto, rilevati in diversi negozi di Trieste sono:
>
> $$118,\ 122,\ 115,\ 118,\ 113,\ 115,\ 117,\ 122,\ 119,\ 119$$
>
> a. Rappresentare i dati in modo da evidenziare tendenza centrale e variabilità dei prezzi osservati.
>
> b. Calcolare la media.
>
> c. Ha senso trovare il prezzo modale?

I dati grezzi sono: $118, 122, 115, 118, 113, 115, 117, 122, 119, 119$.

Passo 1: ordinamento dei dati.

Posizione	1	2	3	4	5	6	7	8	9	10
Valore	113	115	115	117	118	118	119	119	122	122

$n = 10$ valori.

a) Grafico.

Commento. Il grafico più appropriato è il **boxplot** (diagramma a scatola e baffi), che sintetizza in un unico grafico cinque statistiche: minimo, Q_1, mediana Q_2, Q_3 e massimo. Permette di valutare contemporaneamente tendenza centrale, dispersione e asimmetria della distribuzione.

b) Media

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{113 + 115 + 115 + 117 + 118 + 118 + 119 + 119 + 122 + 122}{10}$$

Sommiamo passo per passo:

$$113 + 115 = 228; \quad 228 + 115 = 343; \quad 343 + 117 = 460; \quad 460 + 118 = 578;$$
$$578 + 118 = 696; \quad 696 + 119 = 815; \quad 815 + 119 = 934;$$
$$934 + 122 = 1056; \quad 1056 + 122 = \mathbf{1178}$$

$$\bar{x} = \frac{1178}{10} = \mathbf{117{,}8} \text{ €}$$

Calcolo dei quartili.
Primo quartile (Q_1): mediana dei valori a **sinistra** della mediana

$$Q_1 = \frac{x_{(2)} + x_{(3)}}{2} = \frac{115 + 115}{2} = \mathbf{115} \text{ €}$$

Terzo quartile (Q_3): mediana dei valori a **destra** della mediana

$$Q_3 = \frac{x_{(7)} + x_{(8)}}{2} = \frac{119 + 119}{2} = \mathbf{119} \text{ €}$$

Calcolo della varianza e deviazione standard.

Costruiamo la tavola degli scarti dalla media $\bar{x} = 117{,}8$:

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	113	$113 - 117{,}8 = -4{,}8$	$(-4{,}8)^2 = 23{,}04$
2	115	$115 - 117{,}8 = -2{,}8$	$(-2{,}8)^2 = 7{,}84$
3	115	$115 - 117{,}8 = -2{,}8$	$(-2{,}8)^2 = 7{,}84$
4	117	$117 - 117{,}8 = -0{,}8$	$(-0{,}8)^2 = 0{,}64$
5	118	$118 - 117{,}8 = +0{,}2$	$(+0{,}2)^2 = 0{,}04$
6	118	$118 - 117{,}8 = +0{,}2$	$(+0{,}2)^2 = 0{,}04$
7	119	$119 - 117{,}8 = +1{,}2$	$(+1{,}2)^2 = 1{,}44$
8	119	$119 - 117{,}8 = +1{,}2$	$(+1{,}2)^2 = 1{,}44$
9	122	$122 - 117{,}8 = +4{,}2$	$(+4{,}2)^2 = 17{,}64$
10	122	$122 - 117{,}8 = +4{,}2$	$(+4{,}2)^2 = 17{,}64$
		Somma	**77,60**

Verifica: Fondamentale: la somma degli scarti deve essere zero: $(-4{,}8) + (-2{,}8) + (-2{,}8) + (-0{,}8) + (0{,}2) + (0{,}2) + (1{,}2) + (1{,}2) + (4{,}2) + (4{,}2) = 0$. ✓

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{77{,}60}{9} \approx \mathbf{8{,}622} \qquad s = \sqrt{8{,}622} \approx \mathbf{2{,}937} \text{ €}$$

Possibile commento. La deviazione standard $s \approx 2{,}94$ € indica che i valori si discostano in media di circa **3 euro** dalla media $\bar{x} = 117{,}8$ €: la variabilità è contenuta, coerentemente con l'intervallo interquartile IQR $= 4$ €.

Riepilogo degli indici:

Statistica	**Valore**
Minimo	113 €
Q_1	115 €
Mediana Q_2	118 €
Q_3	119 €
Massimo	122 €
Media $\bar{x}$	117,8 €
IQR	4 €
Dev. standard s	$\approx$ 2,94 €

Possibile commento. Media (117,8) e mediana (118) sono molto vicine: distribuzione approssimativamente simmetrica.

c) Ha senso trovare il prezzo modale?

Contiamo le frequenze di ogni valore distinto:

Prezzo	Frequenza assoluta	Note
113	1	
115	**2**	moda
117	1	
118	**2**	moda
119	**2**	moda
122	**2**	moda
Totale	10	

I valori 115, 118, 119 e 122 compaiono ciascuno **due volte**: la distribuzione è **quadrimodale**.

> Commento.
> **No**, non ha senso identificare un unico prezzo modale: la presenza di quattro mode con la stessa frequenza massima non fornisce informazioni utili sulla tendenza centrale. Quindi è preferibile usare la mediana (118 €) o la media (117,8 €).