

ESERCIZI DI RIEPILOGO 1

ESERCIZIO 1

La tabella seguente contiene la distribuzione di frequenza della variabile $X = \text{“età”}$ (misurata in anni) per un campione casuale di bambini:

x_i	f_i
4.6	8
3.2	3
5.4	6
2.6	2
5.2	5

- Calcolare la media e lo scarto quadratico medio delle età.
- In un campione di adulti l'età media è pari a 37 e lo scarto quadratico medio a 6. Si confronti la variabilità delle età nei due gruppi.

ESERCIZIO 2

La tabella contiene i dati storici relativi al numero di persone (in milioni) impiegate nei principali settori produttivi in Italia:

Anno	Agricoltura	Industria	servizi
1951	8,2	6,3	5,0
1971	3,2	8,3	7,2
1991	1,8	6,9	12,8

Proporre una rappresentazione grafica adeguata a mostrare il cambiamento della forza lavoro in Italia dopo il 1950.

ESERCIZIO 3

L'altezza media di un campione di giocatori di basket è pari a 195 mentre la varianza è pari a 9.

- a) Sapendo che la distribuzione delle altezze ha forma approssimativamente simmetrica e campanulare qual è la percentuale di giocatori con altezza compresa tra 186 e 204 ?
- b) Si verifichi che il valore trovato al punto precedente soddisfa la disuguaglianza di Tcebiscev.

ESERCIZIO 4

Le mele di un'azienda ortofrutticola sono etichettate da un bollino codificato che può dare diritto a premi in denaro di diversa entità. L'azienda ha reso noto che il premio medio è pari a 20 euro e lo scarto quadratico medio dei premi è pari a 4 euro.

- a) A vostro parere la forma della distribuzione dei premi è simmetrica o asimmetrica a destra/sinistra ?
- b) Tenendo conto della risposta data al punto a) cosa è possibile dire circa la percentuale di premi nell'intervallo $[12,28]$?
- c) Sempre tenendo conto della risposta data al punto a) cosa è possibile dire sulla frequenza dei premi con valore superiore a 28 euro?

ESERCIZIO 5

Il seguente diagramma ramo – foglia contiene dati campionari (ramo = unità)

3		0, 1
4		5, 8, 8
5		0, 3, 4, 5, 7, 8, 9
6		1, 4, 7, 9
7		3, 6, 9
8		0, 3, 7

- a. Ricavare i cinque numeri di sintesi e disegnare il box plot.
- b. Trovare l'ottavo decile.
- c. Trovare il 92° percentile.

ESERCIZIO 6

La tabella seguente mostra la distribuzione del reddito di un campione di 700 famiglie (valori espressi in migliaia di euro):

classi di reddito	Numero di famiglie
0 - 6	12
6 - 10	40
10 - 20	250
20 - 30	160
30 - 40	180
40 - 50	30
50 - 60	20
60 - 100	8
Totale	700

- rappresentare graficamente la variabile commentarne la forma.
- calcolare il reddito mediano e quello medio. Dal confronto tra i due valori cosa potete dire sull'asimmetria della distribuzione?
- qual è la frequenza relativa di coloro che hanno un reddito minore di 15 000 euro? Evidenziare tale proporzione sul grafico ricavato al punto a)

ESERCIZIO 7

La varianza può essere calcolata con una formula detta "ridotta". Si scriva la suddetta formula e la si utilizzi per calcolare la varianza della popolazione seguente:

1, 3, 6, 3, 1, 1.

ESERCIZIO 8

In una classe di 15 alunni l'età media è pari a 18 e la varianza dell'età è pari a zero. Si scriva la distribuzione di frequenza dell'età degli alunni.

ESERCIZIO 9

Quale informazione aggiuntiva fornisce il coefficiente di correlazione rispetto alla covarianza? Si spieghi il motivo di questa differenza.

ESERCIZIO 10

La covarianza di due variabili X e Y può essere calcolata attraverso una formula più semplice. Si scriva la formula semplificata e la si utilizzi per calcolare la covarianza delle seguenti coppie (x,y) di valori:

(1,2), (3,4), (6,2), (3,2), (1,8), (1,9).

ESERCIZIO 11

Da un'indagine sulla nazionalità (X) degli studenti dei corsi “graduate” di una università americana è emersa la distribuzione di frequenza riportata nella tabella sotto. Si sa inoltre che sono di sesso maschile il 50% degli studenti italiani e tedeschi, il 65% degli studenti cinesi e il 25% degli altri studenti.

<i>Nazionalità</i>	<i>p_i</i>
Italia	0.20
Germania	0.10
Cina	0.65
Altro	0.05

- a) si rappresenti la variabile “Nazionalità” mediante un opportuno diagramma a barre.

- b) si ricavi la distribuzione di frequenza congiunta delle variabili X =”Nazionalità” e Y =”Sesso”.
- c) Quale percentuale degli studenti è di sesso maschile e non proviene dalla Cina?
- d) si rappresenti graficamente la variabile Y mediante un diagramma a torta.
- e) Si proponga un grafico che mostri la distribuzione del sesso al variare del paese di provenienza. Che indicazione è possibile trarre circa la dipendenza tra le due variabili?

Risultati

1. a) Calcolo media e scarto quadratico medio:

$$\bar{x} = \frac{1}{n} \sum_i x_i f_i$$

$$= \frac{1}{24} (4.6*8 + 3.2*3 + 5.4*6 + 2.6*2 + 5.2*5) = 4.583$$

$$s_x^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 f_i$$

$$= \frac{1}{23} \left((4.6 - \bar{x})^2 * 8 + (3.2 - \bar{x})^2 * 3 + (5.4 - \bar{x})^2 * 6 + (2.6 - \bar{x})^2 * 2 + (5.2 - \bar{x})^2 * 5 \right) = 0.84$$

$$s_x = \sqrt{0.84} = 0.92$$

- b) Per comparare la variabilità calcolo il coefficiente di variazione nei due gruppi (adulti e bambini):

$$CV_{età\ bambini} = \frac{s_x}{|\bar{x}|} = 0.20 \quad CV_{età\ adulti} = \frac{6}{37} = 0.16$$

dal confronto posso concludere che l'età è meno variabile tra gli adulti.

2. Un grafico adeguato è un grafico a barre sovrapposte che mostri, per ogni anno, la proporzione dei lavoratori nei vari settori.

3. a) L'intervallo è $(\mu - 3\sigma, \mu + 3\sigma)$ e la forma della distribuzione è campanulare quindi posso usare la regola empirica e concludere che circa il 99.73% dei giocatori ha un'altezza compresa tra 186 e 204 cm.

- c) Utilizzando la disuguaglianza di Tchebicheff avremmo concluso che *almeno* l'88.9% per cento dei giocatori ha un'altezza compresa tra 186 e 204cm. Essendo $99.73 > 88.9$ la disuguaglianza risulta verificata.

4. ...

5. a) I cinque numeri di sintesi sono:

min=minimo delle osservazioni=3,

Q1 si trova nella posizione $0.25(n+1)=6$ quindi $Q1=5$

Q2 si trova nella posizione $0.5(n+1)=11,5$ quindi $Q2=5.85$

Q3 si trova nella posizione $0.75(n+1)=17$ quindi $Q3=7.3$

Max=massimo delle osservazioni= 8.7

b) L'ottavo decile si trova nella posizione $0.8(n+1)$ quindi ...

c) Il novantaduesimo percentile occupa la posizione ...

6. a) Istogramma, utilizzando la densità di frequenza (...)

b) La variabile "reddito" è continua per intervalli. Per calcolare la **media** calcoliamo i punti medi degli intervalli e poi utilizziamo la formula per variabili discrete:

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_i x_i^m f_i \\ &= \frac{1}{700} (3*12+8*40+15*250+25*160+35*180+45*30 + 55*20 + 80*8) \\ &= 24.99\end{aligned}$$

Per calcolare la **mediana** : passo 1) identifico l'intervallo in cui cade la mediana, si tratta di [20,30] dal momento che la frequenza cumulata è inferiore a 350 nell'estremo inferiore e superiore a 350 nell'estremo superiore; passo 2) la mediana è il valore di x che soddisfa:

$$0.5 = F(20) + (x-20) * (\text{densità intervallo in cui cade mediana})$$

(...)

c) procedere in modo simile al calcolo della mediana.

7. Per la formula ridotta si veda l'appendice al capitolo 5. La varianza della popolazione è 4.68.

8. ...
9. Fornisce l'informazione circa *l'intensità* del legame lineare tra le due variabili.
10. Per la formula ridotta si veda l'appendice al capitolo 5. La covarianza nel campione dato è -0.58.
11. Si veda l'esercitazione 2