

La statistica descrittiva

Delle seguenti indagine statistiche individua la popolazione, il carattere oggetto di studio e le possibili modalità di tale carattere.

1 ESERCIZIO SVOLTO

Indagine: utilizzo del tempo libero da parte degli studenti di una scuola

- il carattere studiato è: utilizzo del tempo libero
- la popolazione è: l'insieme di tutti gli studenti della scuola;
- le modalità con cui si può presentare il carattere sono: lettura, sport, cinema e teatro, visite ai musei, shopping.....

2 Indagine: numero di figli nelle famiglie italiane.

3 Indagine: il grado di istruzione dei lavoratori di una data azienda.

4 Indagine: il peso corporeo (in kg) di un gruppo di bambini di una seconda classe di una scuola elementare.

5 Indagine: le attività sportive preferite dai ragazzi di età compresa fra i 13 e i 18 anni.

6 ESERCIZIO SVOLTO

Le distribuzioni di frequenza. I dati raccolti in un'indagine devono essere organizzati per poterli analizzare. Si ricorre di solito alle distribuzioni di frequenza, cioè tabelle nelle quali vengono riportate le modalità di un carattere e le relative frequenze. Consideriamo il seguente esempio.

Da un'indagine condotta su un certo numero di alunni che frequentano la scuola superiore si è rilevato che 300 alunni leggono 1 solo libro al mese, 150 alunni ne leggono 2 soli al mese, 50 alunni ne leggono più di 2 al mese, 100 alunni non leggono libri.

Costruiamo la tabella delle **frequenze assolute** e **relative**.

Ricordiamo che la **frequenza relativa** si ottiene dividendo ciascuna frequenza assoluta per il totale delle osservazioni; la **frequenza relativa percentuale**, invece, si ottiene moltiplicando e dividendo per 100 la frequenza relativa.

Ad esempio, data la frequenza relativa 0,32, otteniamo la frequenza relativa percentuale nel seguente modo:

$$0,32 \cdot \frac{100}{100} = \frac{32}{100} = 32\%$$

La somma di tutte le frequenze relative è sempre uguale ad 1, mentre la somma delle frequenze percentuali è sempre uguale a 100%.

La tabella delle frequenze relativa all'esempio è la seguente:

Libri letti al mese	Fr. assoluta	Fr. relativa	Fr. relativa perc.
1	300	$\frac{300}{600} = 0,5$	50%
2	150	$\frac{150}{600} = 0,25$	25%
> 2	50	$\frac{50}{600} = 0,08$	8%
0	100	$\frac{100}{600} = 0,17$	17%
Totale	600	1	100%

7 Da un'indagine condotta sugli studenti di una scuola si è rilevato che 382 alunni hanno fatto meno di 3 giorni di assenza, 224 alunni da 3 ad 8 giorni di assenza, 103 alunni hanno fatto da 9 a 15 giorni di assenza, 82 alunni hanno fatto da 16 a 30 giorni di assenza e che 15 alunni hanno fatto oltre 30 giorni di assenza. Organizza i dati in una tabella delle frequenze e determina le frequenze relative e le frequenze relative percentuali.

8 Da un'indagine condotta fra un certo numero di studenti iscritti ad una facoltà universitaria si è rilevato che 300 provengono da un liceo scientifico, 180 da un liceo classico, 80 da un istituto tecnico, 40 da altri tipi di scuola. Organizza i dati in una tabella delle frequenze e determina le frequenze relative e le frequenze relative percentuali.

9 Dai registri dei matrimoni di un grande comune risulta che in un certo anno si sono sposate 2350 coppie; analizzando l'età delle spose, si è visto che 915 erano di età compresa fra i 18 e i 26 anni, 1230 di età compresa tra 27 e 32, le rimanenti di età superiore ai 32 anni. Organizza i dati in una tabella delle frequenze e determina le frequenze relative e le frequenze relative percentuali.

10 ESERCIZIO SVOLTO

La media aritmetica, la moda e la mediana (di un insieme di dati). Per comprendere a fondo un fenomeno statistico evidenziandone le caratteristiche essenziali servono degli indici di sintesi dei dati. Tali indici, ricordiamo, sono la **media**, la **moda** e la **mediana**.

Ad esempio, dati i seguenti valori:

4 5 9 10 2 4 5 6 7 4

calcoliamo la media aritmetica, la moda e la mediana.

- la **media aritmetica** di n numeri è il rapporto fra la loro somma ed n ; nel nostro caso, quindi, abbiamo che la media è: $\frac{4 + 5 + 9 + 10 + 2 + 4 + 5 + 6 + 7 + 4}{10} = 5,6$
- la **moda** è il valore al quale corrisponde la massima frequenza; la moda della nostra distribuzione è, quindi, 4
- la **mediana** è il valore centrale di una serie di n dati (ordinati in senso crescente o decrescente) se n è dispari, altrimenti è uguale alla semisomma dei due valori centrali se n è pari. Riscriviamo i dati della distribuzione in ordine crescente:

2 4 4 4 5 5 6 7 9 10

La mediana è quindi: $\frac{5 + 5}{2} = 5$.

11 Calcola la media aritmetica, la mediana e la moda dei seguenti gruppi di dati statistici:

- a. 4; 8; 6; 12; 15; 16; 10; 7; 12;
 b. 12; 15; 8; 6; 9; 15; 10; 8;
 c. 3; 0,5; 1,25; 4; 2; 3; 1,5; 0,125.

12 Calcola la media aritmetica, la mediana e la moda dei seguenti gruppi di dati statistici:

- a. 17; 16; 18; 20; 25; 30; 24; 26; 30;
 b. 44; 43; 58; 53; 55; 56; 53; 52; 48; 46; 49;
 c. 3; 5; 2; 5; 7; 8; 9; 10; 9; 3; 5; 8; 10.

13 Da un gruppo di amici si è rilevato il numero di scarpe; i dati sono i seguenti:

35 35 36 36 36 37 38 39 39 39 39 40 40 40 40 40

Calcola media, moda e mediana.

14 Dieci misurazioni successive del peso di una certa massa hanno condotto ai seguenti risultati (espressi in grammi):

15 14,3 14,5 14,2 15,1 14,3 14,7 14,8 14,3 14,2

Calcola media, moda e mediana.

15 ESERCIZIO SVOLTO

La media, moda e mediana (di una distribuzione di frequenza). Quando i dati si presentano con una certa frequenza è utile costruire le tabelle di frequenza e da esse calcolare la media, moda e mediana. Vediamo il seguente esempio.

Da un'indagine su un gruppo di 11 ragazzi relativamente al loro peso corporeo sono emersi i seguenti risultati (in kg):

60 53 55 71 83 65 48 48 50 52 52

Organizziamo i dati in una tabella delle frequenze dove aggiungiamo una terza colonna in cui calcolare il prodotto di ciascuna modalità con la sua frequenza.

Peso x	Frequenza assoluta f	Prodotto $x \cdot f$
48	2	$48 \cdot 2 = 96$
50	1	$50 \cdot 1 = 50$
52	2	$52 \cdot 2 = 104$
53	1	$53 \cdot 1 = 53$
55	1	$55 \cdot 1 = 55$
60	1	$60 \cdot 1 = 60$
65	1	$65 \cdot 1 = 65$
71	1	$71 \cdot 1 = 71$
83	1	$83 \cdot 1 = 83$
TOTALE	11	637

- per calcolare la media (detta **ponderata** perché ogni dato si presenta con una certa frequenza) basta dividere il totale ottenuto nella terza colonna per la somma delle frequenze.

La media è, quindi, uguale a: $\frac{637}{11} = 57,9$.

- per calcolare la moda dobbiamo individuare il peso a cui corrisponde la massima frequenza. Dalla tabella notiamo che i valori che presentano frequenza maggiore sono due, il **48** e il **52**. Questa distribuzione ha due valori modali e si dice anche **bimodale**.
- per calcolare la mediana osserviamo che le modalità sono già in ordine crescente (dal peso più basso al peso più alto); per calcolare il valore centrale della distribuzione dobbiamo:
 - costruire la **colonna delle frequenze cumulate** che si ottiene sommando alla frequenza di una certa osservazione tutte le frequenze delle osservazioni precedenti;
 - cercare il valore a cui corrisponde la metà della frequenza complessiva.

Peso x	Frequenza f	Frequenze cumulate
48	2	2
50	1	3
52	2	5
53	1	6
55	1	7
60	1	8
65	1	9
71	1	10
83	1	11
TOTALE	11	

Dal momento che il totale delle frequenze è 11 che è un numero dispari, il valore mediano sarà quello che occupa il posto centrale e che si trova quindi nella sesta posizione. Guardando la colonna delle frequenze cumulate vediamo che tale posizione (6) corrisponde al valore **53** che, quindi, rappresenta il valore cercato.

- 16** Al termine dell'Esame di Stato e dopo la pubblicazione degli esiti, una scuola effettua una indagine per determinare come si sono distribuiti i voti superiori ai 95/100. I risultati sono indicati nella seguente tabella:

Voto in centesimi	96	97	98	99	100
N. diplomati	12	5	8	2	9

Determina media, moda e mediana della distribuzione.

- 17** La distribuzione del reddito in un gruppo di persone è così ripartita:

Reddito (in migliaia di euro)	10	20	30	40	50
Frequenza	22	34	64	55	25

Determina media, moda e mediana della distribuzione.

- 18** La seguente tabella indica la distribuzione dei voti presi da una studentessa universitaria nel corso di laurea in ingegneria:

Voti	18	25	26	27	28	30
Frequenza	1	5	4	8	3	2

Determina media, moda e mediana della distribuzione.

19 ESERCIZIO SVOLTO

Scarto quadratico medio e varianza. Per valutare correttamente un fenomeno a volte non bastano gli indici introdotti.

Per esempio confrontiamo i voti ottenuti da due studenti nelle cinque verifiche di matematica del secondo quadrimestre:

studente A: 7 7 5 8 5

studente B: 7 6 6 7 6

La media aritmetica dello studente A è: $\frac{7+7+5+8+5}{5} = 6,4$

La media aritmetica dello studente B è: $\frac{7+6+6+7+6}{5} = 6,4$

Le due medie coincidono, ma le situazioni sono completamente diverse. Lo studente A ha ricevuto due insufficienze mentre lo studente B non ha mai ricevuto insufficienze e le sue valutazioni risultano abbastanza omogenee.

Possiamo anche dire che lo studente A presenta una variabilità maggiore dello studente B, cioè un andamento meno costante nelle valutazioni.

L'indice che misura questa maggiore o minore variabilità dei dati rispetto alla media aritmetica è lo **scarto quadratico medio** che si indica con la lettera σ .

Ricordiamo che per calcolare lo scarto quadratico medio di una distribuzione di n dati che ammettono media aritmetica M , si può utilizzare la seguente formula:

$$\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - M^2}$$

Il quadrato dello scarto quadratico medio, cioè σ^2 , è detto **varianza**, quindi

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - M^2$$

Ritornando all'esempio, calcoliamo lo scarto quadratico medio delle distribuzioni di voti dei due studenti:

$$\sigma_A = \sqrt{\frac{7^2 + 7^2 + 5^2 + 8^2 + 5^2}{5} - 6,4^2} = 1,2 \qquad \sigma_B = \sqrt{\frac{7^2 + 6^2 + 6^2 + 7^2 + 6^2}{5} - 6,4^2} = 0,49$$

Le valutazioni dello studente B, come avevamo già anticipato, presentano una minor variabilità dato che $\sigma_B < \sigma_A$.

Calcola lo scarto quadratico medio e la varianza delle seguenti sequenze di numeri.

20 10 12 15 18 24 32

21 2 5 6 8 9 12

22 5 7 12 -4 -1 15

- 23** Due gruppi di studenti effettuano 5 misurazioni relative alla lunghezza (in cm) di un oggetto e riportano i dati in una tabella. Quale gruppo presenta la minor variabilità fra i dati?

GRUPPO A		GRUPPO B	
misurazioni	valore	misurazioni	valore
prima	7,8	prima	7,3
seconda	7,7	seconda	7,3
terza	7,3	terza	7,5
quarta	7,5	quarta	7,9
quinta	7,7	quinta	7,4

Risultati di alcuni esercizi.

11. a. 10, 10, 12; b. 10,375, 9,5, 8 e 15; c. 1,921, 1,75, 3

12. a. 22,8, 24, 30; b. 50,6, 52, 53; c. 6,46, 7, 5

13. 38, 40, 39

14. 14,5; 14,3; 14,4

16. 97,75; 96; 98

17. 31,35; 30; 30

18. 26; 27; 27

20. $\sigma = 7,52$; $\sigma^2 = 56,58$

21. $\sigma = 3,16$; $\sigma^2 = 10$

22. $\sigma = 6,67$; $\sigma^2 = 44,55$

23. Gruppo A