

La statistica

E' la disciplina che si occupa della
raccolta di dati
qualitativi/quantitativi relativi a
diversi fenomeni collettivi, della
loro elaborazione e del loro utilizzo
ai fini di conoscenza oppure di
azione

Si chiama **carattere** la proprietà che è oggetto di uno studio in un'indagine statistica

I caratteri si possono suddividere in:

- **qualitativi**: espressi da un nome o da un aggettivo, come ad esempio i tipi di programmi che si preferisce guardare in TV;
- **quantitativi**: espressi da numeri e divisi in **caratteri discreti** e **caratteri continui**.

In ognuna delle due tipologie si possono distinguere:

- **caratteri sconnessi**: tra le cui modalità si possono solo stabilire relazioni di uguaglianza o diversità, ad esempio il luogo di nascita;
- **caratteri ordinati**: tra le cui modalità si può stabilire una relazione di ordine, ad esempio l'altezza di una persona.

Si chiama **popolazione** 'insieme degli individui a cui ci rivolgiamo per prendere informazioni circa il carattere da indagare.

Ogni elemento della popolazione prende il nome di **unità statistica**.

Si chiama **modalità** ogni possibile valore con cui un carattere può presentarsi.

La raccolta dei dati può avvenire mediante:

Censimento: interessa l'intera popolazione di una nazione.

Rilevamento totale: si utilizza quando la popolazione statistica è formata da un numero ridotto di elementi

Rilevamento per campione: Si utilizza quando la popolazione statistica assume dimensioni consistenti

La **frequenza assoluta** di un dato statistico è il numero di volte in cui tale dato si presenta nell'indagine statistica.

Consideriamo i seguenti dati relativi alla massa di 20 alunni di una scuola.

| | | | | | |
|----------|-------|----------|-------|---------------|--------------------|
| Stefano | 48 kg | Andrea | 47 kg | Massa (in kg) | Frequenza assoluta |
| Anna | 37 kg | Giulia | 42 kg | | |
| Luca | 39 kg | Giorgio | 50 kg | 43 | 5 |
| Marco | 47 kg | Claudio | 43 kg | 50 | 4 |
| Paolo | 43 kg | Mara | 39 kg | 47 | 3 |
| | | | | 39 | 3 |
| Giovanni | 50 kg | Nicola | 47 kg | 42 | 2 |
| Matteo | 39 kg | Nora | 43 kg | 37 | 2 |
| Silvio | 43 kg | Pietro | 50 kg | 48 | 1 |
| Sara | 37 kg | Simone | 43 kg | | |
| Enrico | 50 kg | Giovanna | 42 kg | | |
| | | | | Totale | 20 |

La **frequenza relativa** di un dato statistico è data dal rapporto fra la sua frequenza assoluta e il totale delle osservazioni; la **frequenza relativa percentuale** è data dal prodotto fra la sua frequenza relativa e 100.

Relativamente all'esempio precedente:

| Peso (in kg) | Frequenza assoluta | Frequenza relativa | Frequenza relativa % |
|---------------|--------------------|--------------------|-------------------------|
| 43 | 5 | $5 : 20 = 0,25$ | $0,25 \cdot 100 = 25\%$ |
| 50 | 4 | $4 : 20 = 0,20$ | $0,20 \cdot 100 = 20\%$ |
| 47 | 3 | $3 : 20 = 0,15$ | $0,15 \cdot 100 = 15\%$ |
| 39 | 3 | $3 : 20 = 0,15$ | $0,15 \cdot 100 = 15\%$ |
| 42 | 2 | $2 : 20 = 0,10$ | $0,10 \cdot 100 = 10\%$ |
| 37 | 2 | $2 : 20 = 0,10$ | $0,10 \cdot 100 = 10\%$ |
| 48 | 1 | $1 : 20 = 0,05$ | $0,05 \cdot 100 = 5\%$ |
| Totale | 20 | 1 | 100% |

Quando si devono fare confronti fra diverse situazioni, è necessario introdurre un valore di sintesi che caratterizzi il fenomeno studiato.

Tale valore viene detto **indice di posizione** e può essere definito in diversi modi, in relazione cioè agli obiettivi che si pone l'indagine.

I più importanti indici di posizione sono:

- la media aritmetica
- la mediana
- la moda

Si chiama **media aritmetica** di un insieme di dati statistici numerici il valore che si ottiene addizionando tutti i dati e dividendo la somma ottenuta per il numero di dati:

$$\text{media aritmetica} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

ESEMPIO

Se un alunno ha i seguenti voti:

Italiano: 6 Storia: 4 Scienze: 8 Geografia: 5 Matematica: 7 Inglese: 9

la media aritmetica dei voti si calcola nel seguente modo:

$$\frac{6 + 4 + 8 + 5 + 7 + 9}{6} = \frac{39}{6} = 6,5$$

Si chiama **mediana** di un insieme di dati statistici il valore che nella serie di dati, disposti in ordine crescente (o decrescente), occupa il posto centrale.

ESEMPIO

Data la distribuzione:

2 4 5 6 9 11 14

Il termine centrale è 6, pertanto la mediana è 6

Se la serie è costituita da un numero pari di elementi allora la mediana è la media aritmetica dei due elementi centrali.

ESEMPIO

Data la distribuzione:

2 4 5 6 9 11 14 20

Il termine centrale cade nell'intervallo tra 6 e 9, pertanto:

$$\mathbf{Mediana} = (6 + 9) : 2 = 7,5$$

Si chiama **moda** di un insieme di dati statistici il valore che si presenta con maggiore frequenza.

ESEMPIO

La moda della distribuzione

BLU VERDE ROSSO BIANCO BLU GIALLO GIALLO BLU

è BLU

Può anche capitare che in un'indagine statistica ci siano più di un valore modale. Ad esempio la moda della distribuzione

(18) 10 (15) 12 (18) (15) 20

è

MODA = 15 ; 18

Se sono presenti due valori modali, la distribuzione viene detta **bimodale**.