

STATISTICA (concetti principali)

La statistica studia i fenomeni collettivi

Definizione: un fenomeno si dice **collettivo** se è costituito da un insieme di fenomeni singoli tutti dello stesso tipo. Ogni singolo fenomeno di cui si compone un fenomeno collettivo prende il nome di **unità statistica**.

In statistica si utilizzano alcuni termini specifici che hanno particolari significati:

- ❑ **Popolazione:** l'insieme delle unità statistiche
- ❑ **Unità statistica** (u.s.) vedi sopra
- ❑ **Caratteri dell'unità:** caratteristica attribuita ad ogni unità e può essere di due tipi
- ❑ **Carattere qualitativo:** esprime una qualità (non numerico)
- ❑ **Carattere quantitativo:** esprime una quantità (numerico).

Se il carattere è qualitativo si parla di **serie statistiche**. Se il carattere è quantitativo si parla di **seriazioni statistiche**. Queste ultime possono essere continue (temperatura esterna di un corpo) o discrete (numero degli incidenti per mese).

L'indagine statistica: lo studio di un fenomeno statistico è complesso e va suddiviso in fasi dopo aver ben definito l'obiettivo. (Definire l'obiettivo significa anche stabilire un metodo per la rilevazione; quali dati rilevare, per quale fetta di popolazione, etc.)

Fasi dell'indagine

- 1 Fase. Raccolta:** può essere
- Campionaria
 - Totale

Se la raccolta dei dati è totale si parla di *Statistica descrittiva* (descrizione del fenomeno).

Se la raccolta dei dati è a campione, per poter estendere le considerazioni fatte per l'intera popolazione si parla di *Statistica inferente* (estensione alla popolazione di considerazioni fatte su un campione).

Il **METODO STATISTICO** per la raccolta dei dati può essere pubblici o privati. Il più importante in Italia è l'ISTAT.

2 Fase. Spoglio e Trascrizione

Questa fase spesso fatta da elaboratori automatici è lenta se fatta a mano e consiste in:

1. enumerazione dei dati (quanti dati sono stati trattati)
2. suddivisione dei dati in classi a seconda delle esigenze (noi abbiamo diviso in maschi e femmine).
3. trascrizione in tabelle (semplici, doppia entrata, composte).

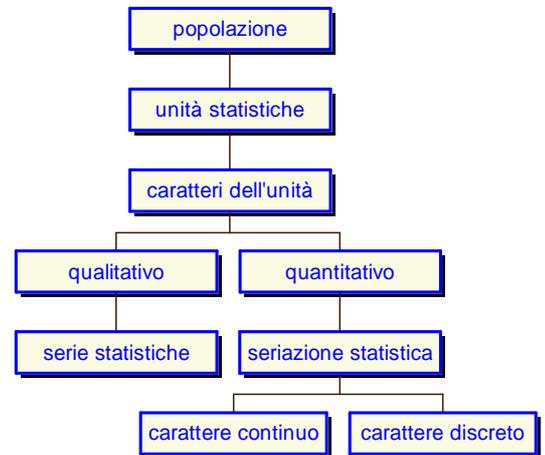
3 Fase. Elaborazione dei dati

Questa permette di fare una sintesi delle informazioni attraverso tabelle, grafici o considerazioni matematiche sui dati (medie, moda, ...).

In questa fase si può anche decidere di rappresentare i dati in modo sintetico cioè fare una sintesi delle informazioni.

Rappresentazioni grafiche

1. Grafici a nastri
2. grafici a settori circolari
3. istogrammi
4. poligoni di frequenza
5. diagrammi cartesiani
6. altre
 - ideogrammi
 - cartogrammi



Sintesi delle informazioni

Ordinare

mettere i dati in ordine logico (dal minore al maggiore se numerici)

Raggruppare

determinare le frequenze dei dati

Rappresentare graficamente

scegliere la rappresentazione più adatta

Calcolare i parametri caratteristici

per valutare il fenomeno

Rappresentazioni grafiche: consigli

1. Accompagnare i grafici con i dati
2. Citare la fonte dei dati
3. Titolare in modo chiaro
4. indicare i caratteri rappresentati
5. indicare l'unità di misura
6. utilizzare le griglie
7. rendere leggibili tutte le parti del grafico

Misure di tendenza centrale

Valori attorno ai quali si addensano i dati: centri della distribuzione

1. Media aritmetica
 - a. ponderata
2. media quadratica
3. moda
4. mediana

1. si dice media aritmetica di n dati la somma dei dati divisa per il loro numero

$$m = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- a. si dice media aritmetica ponderata di n dati x_i ognuno con frequenza nota f_i la somma dei prodotti dei dati per la rispettiva frequenza divisa per il loro numero complessivo

$$m_p = \frac{\sum x_i f_i}{\sum f_i} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{n}$$

2. si dice media quadratica di n dati x_i la radice quadrata della somma dei dati al quadrato diviso il numero totale dei dati

$$m_q = \sqrt{\frac{\sum x_i^2}{n}}$$

3. si definisce moda o termine modale di una distribuzione il termine al quale corrisponde la massima frequenza
4. si definisce mediana il termine che occupa la posizione centrale di una distribuzione quando i dati sono disposti in ordine crescente.

Nota: La mediana e la moda sono altre due misure dell'indice di posizione. Per calcolare la mediana bisogna dapprima riordinare i valori x in ordine numerico; se n è dispari, essa è il valore centrale di x ; se n è pari, è la media delle due x centrali. La moda invece è il valore di x che ricorre più frequentemente. Se due o più valori distinti di x ricorrono con la stessa frequenza, ma non ce n'è alcuno che abbia una frequenza maggiore, si può dire che l'insieme degli x non ammette moda, o equivalentemente che è bimodale, e le due mode sono allora i due valori di x più frequenti. (encarta)

Altre misure: Misure di variabilità

Misura la variazione dei dati in assoluto o rispetto la media

Quando in un fenomeno statistico si ha una variabilità dei dati, le misure di tendenza centrale non sono sufficienti a dare una descrizione precisa del fenomeno. È bene allora prendere in considerazione anche altri fattori quindi altre "misure" che mi consentano di chiarire quanto i dati differiscono tra loro e se sono o no molto concentrati attorno ai valori centrali della serie. Tali misure sono assolute (tipo il campo di variazione) o dipendono dalla media. Per meglio comprenderle è bene definire lo **scarto** che è la differenza fra il dato e la media dei dati.

La misura degli scarti può servire a vedere se i dati sono rappresentati dalla propria media (scarti piccoli = valori vicini alla media); ma tale misura cambia se viene fatta in valore assoluto o utilizzando il quadrato degli scarti.

Di seguito sono riportate le più importanti misure di variabilità:

1. **campo di variazione (Assoluto)**

$$\omega_a = x_{\max} - x_{\min}$$

è la differenza fra i valori estremi: massimo – minimo
corrisponde all'intervallo all'interno del quale ho i dati

2. **scarto semplice medio**

è la media degli scarti (se è zero non è detto che i dati siano tutti pari alla media perché)

3. scarto semplice medio assoluto

è la media dei valori assoluti degli scarti (garantisce che tutti i valori diversi da m siano conteggiati infatti in valore assoluto gli scarti non si possono

mi dice quanto, in media un dato differisce dalla media in positivo o negativo

4. varianza σ^2

è la media dei quadrati degli scarti (ancora più precisa di quella precedente perché amplifica gli errori grandi e minimizza quelli piccoli con l'elevamento al quadrato, mi segnala se ci sono valori "sballati" cioè molto distanti dalla media) è il quadrato dello scarto quadratico medio

$$\text{var} = \frac{\sum (x_i - m)^2}{n} = \frac{\sum (x_i - m)^2 f_i}{\sum f_i}$$

5. scarto quadratico medio σ

è la radice quadrata della varianza

6. coefficiente di variazione

è il rapporto fra lo scarto quadratico medio e la media (è un numero puro, cioè non dipende dall'unità di misura scelta) permette di confrontare più distribuzioni statistiche.

Esempio se le misure fatte da Tizio sono in cm il suo σ sarà in cm, Caio misura in mm il suo σ sarà in mm; ma se vogliono confrontare le proprie serie per vedere la migliore (con dati più vicini alla media) lo potranno fare calcolando il coefficiente di variazione che sarà cm/cm = mm/mm = numero puro.

Schema per il calcolo di σ

esercitazione di laboratorio

<i>valori</i>	<i>frequenze</i>	<i>xi*fi</i>	<i>xi-M</i>	<i>(xi-M)^2</i>	<i>(xi-M)^2*fi</i>
11	1	11	-5,3	28,09	28,09
12	5	60	-4,3	18,49	92,45
13	2	26	-3,3	10,89	21,78
15	5	75	-1,3	1,69	8,45
17	7	119	0,7	0,49	3,43
18	3	54	1,7	2,89	8,67
20	3	60	3,7	13,69	41,07
21	4	84	4,7	22,09	88,36
somme	30	489	-3,4	98,32	292,3
	media pond	16,3		varianza	9,743
				scarto quadratico medio	3,121