

STATISTICA

Nasce dalla parola “*stato*”, “*statista*” (chi si occupa delle questioni di stato).

Insieme dei principi e dei metodi per la raccolta, elaborazione, utilizzazione e interpretazione di informazioni riguardanti fenomeni collettivi. In termini più generali la statistica ha come scopo la **conoscenza quantitativa dei fenomeni collettivi**.

Esistono due tipi di statistica:

1. Statistica **descrittiva**: fornisce gli strumenti per rappresentare, sintetizzare ed interpretare il modo in cui un fenomeno di interesse si è manifestato nel collettivo osservato.

Non fornisce strumenti per estendere i risultati ad una popolazione più ampia.

- collezionare (sondaggio);
- presentare (tabelle);
- sintetizzare (media).

Le componenti di base sono:

1. **Variabile**: informazione di interesse;
2. **Unità statistica**: entità elementare a cui si riferiscono le informazioni;
3. **Modalità del carattere**: modo in cui il carattere si manifesta;
4. **Campione collettivo o statistico**.

2. Statistica **inferenziale**: permette di estendere i risultati e le interpretazioni all'intera popolazione.

Si tratta di un ragionamento induttivo: si traggono conclusioni sulla base di un campione.

Si basa sulla **teoria della probabilità** (ha una componente di incertezza).

Le fasi di studio sono:

1. **Rilevazione obiettivi** (per predisporre il tipo di rilevazione);
2. **Pianificazione**;
3. **Conduzione studio e raccolta dati**;
4. **Analisi statistica**;
5. **Interpretazione e uso dei risultati**.

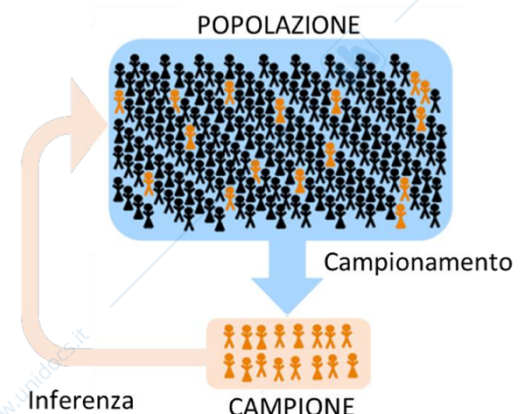
Un corretto disegno permette sia la generalizzazione sia la corretta interpretazione

È fondamentale sapere come viene reclutato il campione, poiché esso deve essere rappresentativo.

La popolazione è l'insieme di tutte le unità che costituiscono il gruppo che si è interessati ad analizzare. I dati vengono raccolti in indagini totali.

Il campione è il sottoinsieme delle unità osservate nella popolazione.

I dati vengono raccolti in indagini campionarie.



Perché fare indagini campionarie?

1. Troppo costoso intervistare tutta la popolazione e la elaborazione dei dati richiederebbe troppo tempo;
2. Si commetterebbero troppi errori di natura non campionaria;
3. L'indagine completa è impossibile in pratica (es. controllo le componenti elettroniche – non possono testare tutte le lampadine);
4. L'indagine completa è impossibile in linea di principio (es. le misure – non posso misurare esattamente un oggetto).

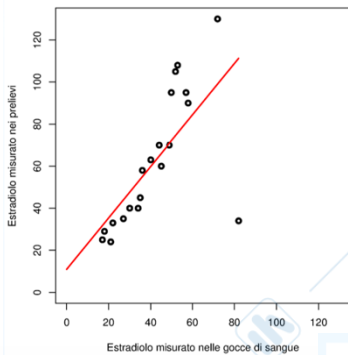
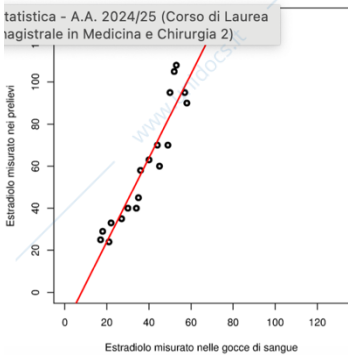
Perché la statistica serve a un clinico?

1. Rappresentare le info in possesso (es. caratteristiche di un paziente);
2. Identificare un quadro diagnostico;
3. Sapere come trarre conclusioni relative a una popolazione;
4. Avere previsioni affidabili.

Nell'approccio più moderno alla statistica si dice che la statistica è lo studio della variabilità: ogni fenomeno si manifesta con un maggiore o minore grado di variabilità.

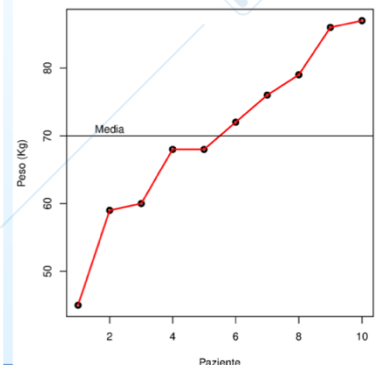
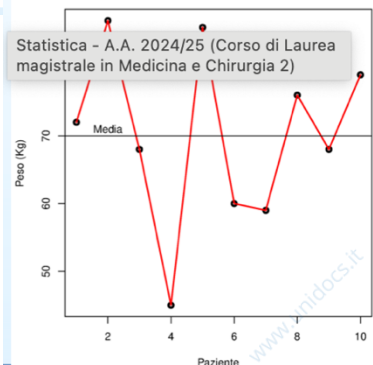
POSSIBILI ERRORI:

- Due gruppi non sono comparabili e non è possibile valutare l'effetto della variabile di interesse escludendo effetti di confondimento;
- Si seleziona un campione che non è rappresentativo della popolazione. Quindi, i risultati non sono generalizzabili alla popolazione target.;
- Non si considera un'ampia dimensione del campione, per cui se non si trovano risultati evidenti non si può discriminare se l'intervento studiato non sia efficace o i dati a disposizione non siano suff. a mostrare l'effetto dell'intervento nonostante la variabilità dei soggetti;
- Riportare nel database valori in unità di misura diverse/ sbagliare a riportare i valori o indurre il soggetto a rispondere in un certo modo.



Un solo dato sbagliato cambia il disegno finale

Idea distorta a causa della mancata casualità



LE VARIABILI

Noi vogliamo descrivere, sintetizzare ed estrapolare il massimo di informazione. Terminata la fase di acquisizione dei dati vediamo come rappresentarli e sintetizzarli:

1. il primo risultato della rilevazione sono i dati grezzi, ovvero distribuzione del collettivo secondo i caratteri considerati: distribuzione unitaria.

Le variabili si classificano in quattro tipi, a seconda dei valori che assumono:

1. **Categorica**: se i valori indicano l'appartenenza a un gruppo, le categorie devono essere mutualmente esclusive (un soggetto in una e una sola categoria):
 - **Nominali**: categorie non soggette ad alcun tipo di ordinamento, possono essere anche simboli (es. presenza di febbre, gruppo sanguigno);
 - **Ordinali**: presentano in via naturale un ordine (es. liv istruzione o soddisfazione).
2. **Numerica**: i valori rappresentano delle quantità che vengono misurate o conteggiate.
 - **Discrete**: si possono contare, valore intero ma senza unità di misura (es. n. figli o n. interventi subiti);
 - **Continue**: da $-\infty$ a $+\infty$ e possono essere espresse in unità di misura (es. peso, età).

DISTRIBUZIONE DI FREQUENZA

È una tabella contenente le categorie che si osservano nei dati e le corrispondenti frequenze con cui i dati appartengono alle categorie.

- È un modo per riassumere i dati;
- Condensa i dati grezzi in forma più utile;
- Consente una veloce interpretazione.

Funzione che associa ad una variabile statistica X_i la sua **frequenza n_i** = numero di volte in cui la variabile si presenta.

“**i**” = valore della variabile associato alla generica 1-esima posizione nella successione ordinata.

$$\text{Frequenza relativa } (p_i) = \frac{\text{Frequenza}}{\text{n. totale di osservazioni}} = \frac{n_i}{n}$$

Se mettiamo in relazione la frequenza con il totale otteniamo la **frequenza relativa p_i** la quale è il rapporto tra frequenza assoluta e il numero totale di osservazioni.

ID	UNITÀ OSPEDALIERA
1	Cardiologia
2	Cure intensive
3	Emergenza
4	Emergenza
5	Chirurgia
6	Cardiologia
7	Chirurgia
8	Chirurgia
9	Emergenza
10	Cardiologia
11	Cardiologia
12	Cardiologia
13	Cure intensive
14	Chirurgia
15	Cardiologia
16	Chirurgia
17	Emergenza
18	Cure intensive
19	Chirurgia
20	Chirurgia

UNITÀ OSPEDALIERA (i)	Frequenza (n_i)	Frequenza relativa ($p_i = n_i/n$)
Cardiologia	6	6/20 = 0.3
Chirurgia	7	7/20 = 0.35
Cure intensive	3	3/20 = 0.15
Emergenza	4	4/20 = 0.2
Totale	20	1.00

Frequenza relativa cumulativa: frequenza relativa fino a quella classe inclusa.

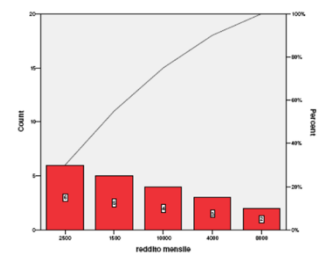
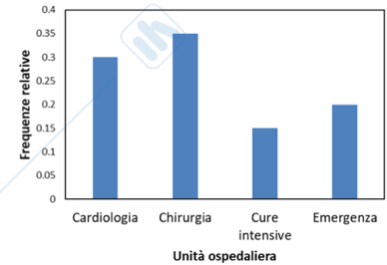
GRAFICI DI DISTRIBUZIONE DI FREQUENZA

DIAGRAMMI A BARRE (variabili qualitative)

Ogni barra è associata a una categoria della variabile considerata, tutte le barre hanno stessa larghezza e l'altezza delle barre è proporzionale alle frequenze delle categorie.

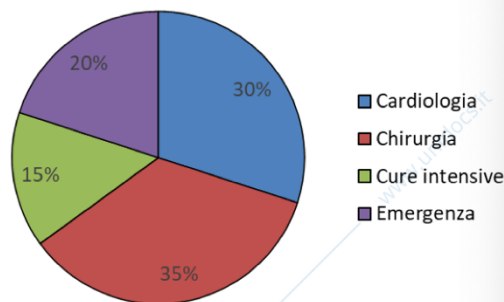
Ne esistono due tipi:

- Rappresentazione **quantitativa**: sull'asse dell'ascisse si indica il valore della variabile e sulle ordinate la frequenza;
- Rappresentazione **qualitativa**: "a barre di Pareto" sull'ascisse si segnano i valori della variabile in ordine decrescente di importanza in termini di frequenza.



DIAGRAMMI A TORTA (variabili qualitative)

La torta è divisa in tante fette quante le categorie della variabile categorica considerata, l'ampiezza di ciascuna fetta è proporzionale alla frequenza della categoria corrispondente.



VARIABILI NUMERICHE

Le osservazioni possono assumere molti valori diversi tra loro: si suddivide l'insieme dei valori che la variabile può assumere in intervalli, detti **classi**.

Quando le classi hanno la stessa ampiezza è possibile calcolare l'ampiezza dell'intervallo.

$$\text{Ampiezza dell'intervallo} = \frac{\text{valore massimo} - \text{valore minimo}}{\text{numero di classi}}$$

(ogni osservazione deve poter essere assegnata ad una ed una sola classe)

ISTOGRAMMA (*variabili numeriche*)

Gli estremi degli intervalli sono rappresentati sull'asse orizzontale; l'asse verticale rappresenta la frequenza (relativa o assoluta) delle classi.

!! Numero delle classi:

- TROPPE: distribuzione frastaglia e diva indicazione di come la frequenza cambia nelle classi;
- POCHE : comprime troppo la variazione è oscura e importanti andamenti.

OGIVE (*variabili numeriche*)

Linea spezzata, che rappresenta frequenze cumulative e che unisce i punti che hanno per ascisse gli estremi degli intervalli e per ordinate le corrispondenti frequenze relative cumulative.

Utile per capire quante sono le frequenze superiori o inferiori a un certo valore e fare dei confronti tra le frequenze cumulative di gruppi diversi.

Forma della distribuzione:

- **Simmetrica**: osservazioni bilanciate, distribuite in modo regolare attorno a centro;
- **Asimmetrica**: osservazioni non distribuite in modo simmetrico rispetto al centro (può essere con coda a sinistra o a destra).