

STATISTICA MEDICA

La statistica fu definita come la scienza deputata a raccogliere dati utili per meglio governare (come nel caso del censimento). Ad oggi, questa scienza viene usata in campo biomedico al fine di:

- **Scoprire le leggi che regolano i dati stessi;**
- **Operare il confronto fra dati in diverse condizioni sperimentali.**

La statistica si suddivide in due gruppi:

- **Statistica descrittiva:** riassume e descrive un insieme di dati relativi ad un campione.
- **Statistica inferenziale:** generalizza i risultati ottenuti su un campione all'intera popolazione.

NOZIONI

(1° Nozione) La variabile: la variabile è il dato che vogliamo rilevare in un campione o una popolazione. La variabile può essere:

- **Qualitativa:** variabile che non può essere contata ed ha carattere puramente descrittivo (come il colore degli occhi). Questa variabile può essere ulteriormente suddivisa in:

qualitativa nominale: usata per etichettare dei dati senza alcun contenuto quantitativo (come il colore dei capelli, la nazionalità, i nomi di persona)

qualitativa ordinale: basata su un principio di ordinamento (come un elenco in ordine crescente o decrescente).

- **Quantitativa:** variabile che può essere contata e ha carattere numerico (come indagine su un certo numero di persone). Questa variabile può essere ulteriormente suddivisa in:

quantitativa discreta: può assumere un numero finito o un'infinità numerabile di valori (X appartiene all'insieme dei numeri interi Z)

quantitativa continua: può assumere un'infinità più che numerabile di valori (X appartiene all'insieme dei numeri razionali Q)

- **Casuale:** variabile che può assumere valori diversi in base a fenomeni aleatori (come il lancio di un dado a sei facce che può assumere un valore numerico che va da 1 a 6)

- **Deterministica:** risultato di operazioni matematiche rispetto a variabili casuali (come il rilevamento della velocità che è s/t)

(2° Nozione) La popolazione: con il termine popolazione intendiamo un insieme di elementi che sono oggetto di indagine statistica. Si usa il termine "elementi" poiché non sempre si parla di persone, la popolazione infatti potrebbe essere rappresentata da cellule così come da oggetti non viventi. L'insieme ridotto di elementi estratto dalla popolazione è considerato rappresentativo della stessa è detto campione. La scelta del campione, ovvero il processo detto "campionamento", deve essere operato mediante criteri di casualità e di rappresentatività, usando metodologie statistiche di estrazione.

N = numerosità della popolazione

n = numerosità campione

Il campionamento può essere probabilistico o non probabilistico, ad oggi si preferisce il campionamento probabilistico. Le modalità di campionamento probabilistico sono:

- **Casuale semplice:** è il metodo di più facile realizzazione, avviene tramite estrazione da un'urna ed ogni individuo ha pari probabilità di essere estratto.

- **Sistematico:** dividendo la numerosità della popolazione (N) per la numerosità del campione (n) sarà possibile trovare il passo di campionamento (K)

($K = N/n$). Si estrarrà poi un numero nel passo di campionamento e la scelta delle unità successive avverrà sommando a tale numero il passo di campionamento stesso.

- **Stratificato:** è il metodo più usato; si dividono gli individui in strati o sottopopolazioni, si estrae un campione casuale semplice da ogni strato e infine si elabora una statistica dell'ampiezza richiesta.

- **A più stadi:** è un metodo di campionamento nel quale la popolazione viene suddivisa in strati, solo alcuni dei quali vengono estratti a caso e all'interno di questi strati estratti si estrae a sua volta un campione casuale secondo un ulteriore piano di campionamento

(Esempio: al fine di ridurre i costi di formazione del personale, si decide di non effettuare un'indagine in tutti i comuni di una regione, ma di scegliere a caso alcuni di questi. In questo modo si riduce eventualmente pure il costo dovuto agli spostamenti del personale di rilevazione. La scelta casuale dei comuni può avvenire con un qualsiasi piano di campionamento, per esempio con probabilità di estrazione proporzionale alla dimensione del comune. I comuni formano così il primo stadio. Una volta scelti i comuni, si estrae in ciascuno di loro un campione di famiglie, p.es. seguendo un piano di campionamento semplice. Le famiglie formano così il secondo stadio. All'interno delle famiglie si decide di intervistare soltanto una persona scelta a caso. Il campionamento delle persone forma così il terzo stadio (anche se non sempre possibile, è raccomandabile, durante un'indagine statistica, conoscere la popolazione e le sue caratteristiche).

Le modalità di campionamento non probabilistico sono invece:

- **Campionamento di convenienza:** si intervistano le unità più accessibili.
- **Campionamento per quote:** si divide la popolazione in più sottogruppi fissando le quote da intervistare per ciascun sottogruppo.
- **Campionamento a valanga:** si sceglie la prima unità, ad essa viene chiesto di suggerirne un'altra e così via.

(3° Nozione) Precisione e Accuratezza.

- **Precisione:** si riferisce a quanto si avvicinano tra di loro misure indipendenti.
- **Accuratezza:** si riferisce a quanto siamo vicini al valore reale con una data misura.

PIANO DI RACCOLTA DEI DATI

Il piano di raccolta dei dati richiede variabili pertinenti ed unità statistiche pertinenti.

1° Esempio: studio della mortalità infantile. // **variabile pertinente:** età in giorni.

2° Esempio: malattia alla prostata, i soggetti studiati sono gli uomini. // **unità statistica pertinente:** uomo (non verrà di certo considerata una donna).

ANALISI DEI DATI

Come abbiamo già accennato, la statistica descrittiva riassume e descrive un insieme di dati relativi ad un campione. Gli strumenti atti a sintetizzare i dati raccolti sono:

- 1) Distribuzione di frequenza.
- 2) Grafici.
- 3) Indici di posizione.

1) DISTRIBUZIONE DI FREQUENZA

Comprende 5 tipi di frequenze principali:

- **Frequenza assoluta:** la frequenza assoluta (n_i) è il numero di volte che un certo valore si verifica in un insieme di dati. Se un insieme di dati è composto da 50 persone, di cui 30 sono femmine, la frequenza assoluta del genere femminile è 30.
- **Frequenza relativa:** la frequenza relativa è la proporzione in cui un evento si verifica rispetto al totale degli eventi. La frequenza relativa (p_i) si calcola dividendo la frequenza assoluta (n_i) del valore per il totale dei dati (n).
- **Frequenza cumulata:** la frequenza cumulata è pari alla somma della sua frequenza assoluta e di quelle delle modalità che la precedono. Naturalmente, affinché il calcolo della frequenza cumulata abbia un senso, è necessario che nella distribuzione sia possibile definire qualche tipo di ordinamento.
- **Frequenza cumulata relativa:** la frequenza cumulata relativa è il rapporto tra la sua frequenza cumulata e la somma di tutte le frequenze assolute presenti nel campione.
- **Frequenza percentuale:** la frequenza percentuale è uguale al prodotto della frequenza relativa moltiplicata per 100.

2) GRAFICI

- **Diagramma a barre:** il diagramma a barre, o ortogramma, è un grafico utilizzato per rappresentare caratteri qualitativi. Può essere di due tipi: a colonne (verticale), oppure a nastri (orizzontale). Il diagramma consiste in una successione di rettangoli equidistanti, tanti quanti sono gli attributi del carattere, la cui altezza o lunghezza è proporzionale alla frequenza dell'attributo corrispondente.

- **Diagramma in coordinate cartesiane:** la rappresentazione grafica in questione si avvale di un riferimento in coordinate cartesiane. Gli assi si chiamano rispettivamente, asse delle ascisse, che si indica con x , e asse delle ordinate, che si indica con y . Il punto O è denominato origine degli assi. Con l'utilizzo dei diagrammi cartesiani si pone in evidenza il comportamento di un fenomeno in funzione di un altro. Ciascun dato statistico è sempre individuato da una coppia di numeri e cioè dal suo valore e dal tempo cui si riferisce. Questa coppia di numeri definisce dunque sul piano cartesiano una successione di punti, i quali, se uniti con una linea, che potrà essere una curva, una retta o, più spesso, una spezzata, daranno la rappresentazione del comportamento del fenomeno.

- **Istogramma:** rappresenta i dati statistici mediante superfici rettangolari. Si avvalgono anch'essi del sistema di assi cartesiani. Sull'asse delle ascisse si pongono degli intervalli, ciascuno dei quali ha la sua ampiezza. Gli istogrammi si prestano assai bene per le rappresentazioni di caratteri quantitativi continui.

- **Diagramma circolare:** il diagramma circolare è una rappresentazione utilizzata per i caratteri qualitativi e mostra le proporzioni delle parti rispetto all'insieme. Nel diagramma circolare, l'area del cerchio esprime la frequenza totale, mentre i settori circolari esprimono le quote riferibili agli attributi del carattere.

3) INDICI DI POSIZIONE

- **Moda:** è definita come il valore che ha la frequenza più alta. Non si calcola ma si individua; possono esistere più mode per una stessa distribuzione.
- **Mediana:** rappresenta quel valore che ripartisce la distribuzione in due metà. La prima metà conterrà tutti i valori inferiori o uguali alla mediana, mentre la seconda metà conterrà tutti i valori maggiori o uguali alla mediana. Non si calcola ma si individua; quello che possiamo calcolare è la posizione del suo valore.
- **Media:** è quel valore che corrisponde alla somma di tutti i valori diviso il numero dei valori stessi. E' simmetrica perché moda, mediana e media coincidono; la somma degli scarti della media è uguale a 0.
- **Media aritmetica ponderata:** a differenza della media aritmetica, per il calcolo della media ponderata è necessario, oltre a conoscere i numeri di cui si vuole calcolare la media, avere anche i valori dei pesi di ogni singolo valore. Il peso ci fornisce informazioni su quanto ogni numero sia "rilevante" ai fini della media.
- **Media geometrica:** è utilizzata per valori molto distanti fra di loro o potenze di 10.

INDICI DI DISPERSIONE

- **Varianza:** la varianza è anche conosciuta come deviazione standard quadratica ed è indicata con la lettera greca sigma al quadrato σ^2 . La varianza è la media aritmetica dei quadrati delle differenze tra ogni valore della distribuzione e un valore medio preso come riferimento.
Varianza popolazione: devianza/n osservazioni.
Varianza campione: devianza/n-1.

(devianza = differenza tra il valore osservato di una variabile e qualche altro valore, spesso la media di quella variabile).

- **Deviazione standard:** la deviazione standard è un indice statistico fondamentale che esprime la tendenza dei dati a variare. Supponiamo ad esempio di voler analizzare il listino prezzi di due fornitori concorrenti. Il fornitore con una deviazione standard più bassa avrà dei prezzi con meno oscillazioni, e quindi, i dati, varieranno molto poco.

- **Campo di variazione:** è la differenza fra valore massimo e minimo.
 $R = x_{\max} - x_{\min}$.

INDICI DI FORMA

- **Asimmetria:** una distribuzione sarà asimmetrica quando parte sinistra e destra di quest'ultima non saranno speculari rispetto all'asse centrale.

- **Tendenza centrale:** la tendenza centrale di un insieme di dati indica dove questi sono posizionati o concentrati.

- **Curtosi:** la curtosi è l'altezza massima raggiunta nella curva di frequenze di una distribuzione statistica. Troviamo 3 tipi di curve:

1) **Leptocurtosi:** curva di frequenza con una curtosi molto alta. Gli elementi della distribuzione sono concentrati nelle immediate vicinanze della media aritmetica; la variabilità è bassa.

2) **Mesocurtosi:** curva di frequenza con una curtosi mediamente alta; la variabilità è media.

3) **Platicurtosi:** curva di frequenza con una curtosi molto bassa; la variabilità è alta.

QUANTILI

I quantili sono utilizzati in statistica per frazionare in N parti uguali un insieme di dati numerici disposti in ordine progressivo crescente o decrescente. I quantili vengono inoltre usati quando non si conosce la forma della distribuzione e quando quest'ultima è asimmetrica. I principali quantili sono i decili e i percentili.

- **Decili:** decili sono quantili che suddividono la seriazione in dieci parti uguali. Ogni decile rappresenta $1/10$ della seriazione.

- **Percentili:** i percentili sono quantili che suddividono la seriazione in cento parti uguali. Ogni percentile identifica $1/100$ della seriazione.

IPOTESI STATISTICHE

- **H₀**: è un'affermazione riguardo alla popolazione che si assume essere vera fino a che non ci sia una prova evidente del contrario.

- **H₁**: indica l'ipotesi alternativa.

Si può rifiutare un' ipotesi solo quando si palesa essere falsa con almeno una probabilità pari ad $1-\alpha$ (in genere, 95%). La decisione che viene presa è infatti sempre soggetta ad errori, che in statistica sono distinti in:

- **Errore di I tipo (alpha)**: rifiuti l'ipotesi nulla che in realtà è vera. (come condannare una persona che in realtà è innocente).

- **Errore di II tipo (beta)**: accetti l'ipotesi nulla che in realtà è falsa. (come giudicare innocente una persona che in realtà è colpevole).

TEST E PARAMETRI

Parametri principali:

- **Sigma (σ)**.

- **Media (m)**.

- **Test parametrici**: test T- student (per il calcolo di media e deviazione standard).

- **Test non parametrici**: test chi - quadro e test di fisher; tutti i tipi di test, parametrici o meno, verificano un'ipotesi.

TEST T-STUDENT

Il test T-Student comprende un insieme di tecniche parametriche che permettono di confrontare le medie di due gruppi per capire se la differenza osservata è reale o dovuta al caso. Quanto deve essere numeroso il campione per effettuare il test? il dataset dovrebbe contenere almeno una decina di unità statistiche per ogni gruppo; se sono di più il risultato sarà più facilmente generalizzabile. Per poter utilizzare questo test, ci sono tre controlli da eseguire:

1) La variabile dipendente deve essere quantitativa.

2) La variabile indipendente deve essere qualitativa e deve avere solo due modalità (come il genere; maschio o femmina).

3) Le unità statistiche devono essere tra loro indipendenti.

MEDICINA BASATA SULL' EVIDENZA (EBM)

La medicina basata sull'evidenza consiste nell'uso cosciente e giudizioso delle migliori evidenze biomediche al momento disponibili, così da prendere le decisioni più adeguate per l'assistenza del singolo paziente. Il paziente è il centro della medicina fondata sulle prove; quest'ultimo quindi deve sapere perché gli vengono consigliati determinati esami e cure rispetto ad altre.

L'approccio EBM si può sintetizzare come un modello a 5 fasi:

1. porre domande cliniche cui si può rispondere.
2. cercare le prove.
3. valutare criticamente le prove per la loro validità e utilità.
4. prendere una decisione, integrando la propria competenza clinica con i valori del paziente.
5. valutare ciò che si è fatto.

LO STUDIO IN CIECO

Qualsiasi tecnica o metodica, per essere validata dalla comunità scientifica deve essere quasi sempre supportata da uno studio scientifico in cieco.

Questo perché esistono dei fenomeni non controllabili come **l'effetto placebo** che rischia di falsare i risultati. Gli studi in cieco mirano proprio a limitare questi effetti consci o inconsci di aspettativa, evitando di mettere a conoscenza di determinate informazioni le persone coinvolte nella ricerca.

STUDIO IN CIECO SEMPLICE

La prima tipologia di studio scientifico che andiamo ad analizzare è quello in cieco semplice, prendiamo come esempio lo studio di un farmaco:

- i pazienti vengono divisi in due gruppi
- un gruppo prende il farmaco
- l'altro gruppo prende una pastiglia senza il principio attivo
- tutti i partecipanti appartenenti ai due gruppi sono convinti di prendere il farmaco col principio attivo
- il medico che dà il farmaco e chi analizza i risultati conoscono quali pazienti appartengono ai diversi gruppi

Confrontando i risultati dei due gruppi si può capire quale sia la reale efficacia del principio attivo, escludendo di fatto parte dell'effetto placebo.

STUDIO IN DOPPIO CIECO

Riprendendo l'esempio di prima, lo studio in doppio cieco possiede le stesse caratteristiche dello studio in cieco semplice, solo che in questo caso anche il medico non è a conoscenza di quale paziente appartiene a quale gruppo.

Dal suo punto di vista ogni persona sta assumendo lo stesso farmaco, senza sapere chi sta prendendo un placebo. Questo permette di escludere che le convinzioni del medico possano influenzare il paziente convincendo maggiormente della bontà della cura il gruppo che sta prendendo il farmaco vero. Lo studio in doppio cieco, quando si parla di medicina e psicologia è l'unica via per valutare nel modo corretto i risultati di un esperimento. A oggi, solo dopo aver ottenuto dei risultati validati da questo metodo di ricerca, i farmaci e le terapie possono essere ritenuti efficaci dalla comunità scientifica.

VALUTAZIONE DELLE PROVE

I livelli di evidenza sono classificati da 1 a 5 in ordine decrescente di qualità e consistono nei seguenti:

- **Livello 1** (la massima qualità): meta-analisi, studi controllati e randomizzati di alta qualità.
- **Livello 2**: studi di coorte.
- **Livello 3**: revisione tramite studi caso-controllo.
- **Livello 4**: studi di coorte e studi caso-controllo di qualità minore.
- **Livello 5**: opinione di esperti non basata su una valutazione critica, ma su concetti di fisiologia e ricerca di laboratorio.

META - ANALISI

La metanalisi è una tecnica clinico-statistica quantitativa che permette di combinare i dati di più studi condotti su di uno stesso argomento, generando un unico dato conclusivo per rispondere a uno specifico quesito clinico. La metanalisi risulta utile quando esiste incertezza nella valutazione di efficacia di un trattamento, o perché i risultati dei singoli studi non sono univoci, oppure perché i singoli studi sono effettuati su pochi pazienti e - considerati singolarmente - sono scarsamente affidabili. La combinazione dei dati diminuisce l'imprecisione dei risultati dei singoli studi.

STUDIO DI COORTE

Uno studio di coorte è un disegno di studio non sperimentale ma osservazionale (i ricercatori non effettuano nessun intervento) che segue un gruppo di persone (una coorte) per un periodo di tempo di anni o anche decenni. Le coorti sono gruppi di individui selezionati sulla base di determinate caratteristiche condivise e suddivisi in sottogruppi in base alle loro differenze.

Uno studio di coorte può essere utilizzato per determinare la prognosi di una malattia. Un gruppo di pazienti a cui è stato diagnosticato uno stadio precoce della malattia viene sottoposto a ripetuti controlli nel tempo per valutare l'incidenza (nuovi casi all'anno) e il decorso temporale dei diversi esiti.

OSSERVAZIONE SISTEMATICA

Condurre un'osservazione sistematica significa rivolgere l'attenzione al comportamento messo in atto in una precisa situazione, registrando fedelmente le informazioni ricercate. Le fasi sono:

- Scelta del problema da studiare.
- Situazione in cui osservare.
- Spaccato di realtà da inquadrare.
- Strumenti e procedure da utilizzare.