

Statistica:

Test di ipotesi:

- Ipotesi di ricerca
- Disegni di ricerca
- Test di ipotesi
- Evidenza empirica

Ipotesi di ricerca:

l'idea di ricerca che ha sempre dei passi ben precisi perché prima di qualsiasi ricerca io devo assicurarmi che certi lavori certe idee non siano già state pubblicate.

quesito di ricerca è un concetto in cui ancora non ho stabilito che cosa voglio dimostrare.

l'ipotesi di ricerca è un concetto meno creativo. abbiamo visto che il passo successivo è il disegno di ricerca.

Disegni di ricerca:

2 famiglie:

analisi di differenze tra gruppi: lavoro su un unico parametro, stessa variabile. Prima, dopo o tra due gruppi.

analisi di differenze tra variabili: guardo se esistono relazioni tra variabili, eseguo studi correlazionali.

Test di ipotesi:

è la pura statistica che si basa su un confronto tra gruppi o su relazioni casuali.

Evidenza empirica:

Si installa in un sistema ed è importante assicurarsi che sia vero.

Metodi di statistica medica:

- Estrarre dalla popolazione in studio un campione rappresentativo (Campionamento)
- Trarre dal campione una serie di informazioni (Rilevazione dei dati)
- Descrivere le informazioni rilevate sul campione (Statistica Descrittiva)
- Approssimare la distribuzione che le informazioni potrebbero assumere nella popolazione di riferimento (Curve di Distribuzione)
- Generalizzare, sulla base delle stime effettuate sul campione, quanto accade sulla popolazione da cui il campione è stato estratto (Statistica Inferenziale)

Campionamento:

- Popolazione obiettivo (o popolazione target)
 - Popolazione campionata (o base di campionamento) dalla quale estraggo il campione
 - Il campione
- Errore di selezione o bias di selezione

Tipi di campionamento:

- Campionamenti probabilistici (è il vero campione)
 - Casuale: ogni soggetto ha la stessa probabilità di essere estratto
 - Casuale semplice: viene estratto tra tutti i campioni della stessa grandezza, hanno stessa probabilità di essere estratti senza nessun limite
 - Sistemático: decido che ogni 10 elementi ne estraggo uno, ma ha limiti. Rischio di trovarne molti con la stessa caratteristica

- Stratificato: non hanno la stessa probabilità di essere estratti, seleziono io, limito l'applicazione della ricerca (per sesso, età, ecc)
- A grappolo: scelgo es provincia di Bologna, poi divido per aziende, ecc. usato molto per indagini inquinamento, non viene usato in ambito medico

- Campionamenti non probabilistici
 - A quote
 - A Valanga: scelgo io la popolazione
 - A scelta ragionata: scelgo io il gruppo di persone per convenienza
 - Accidentale

La taglia di un campione è la grandezza in termini numerici, quante persone sono in un campione = N
n = taglia del campione

Statistica descrittiva: i tipi di variabili sono:

- quantitative: hanno unità di misura. Le misure fatte su queste variabili danno informazioni sulla loro grandezza
 - Variabili continue
 - Variabili discrete
- qualitative si esprimono i termini categorici

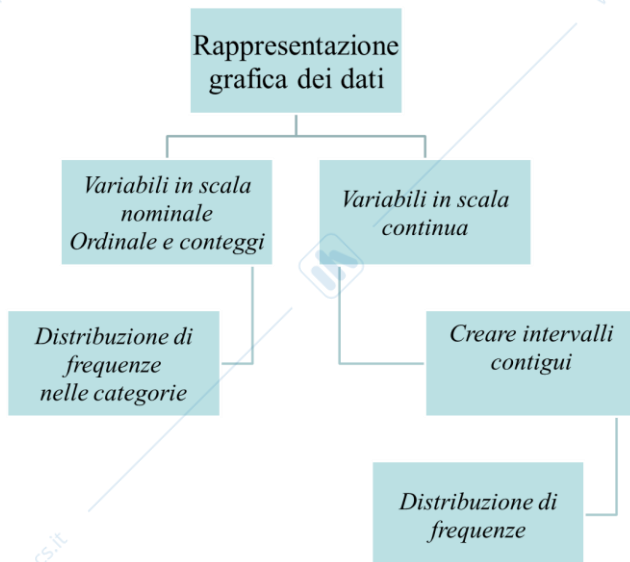
Scale di misura:

Scala	Tipologia
Nominale	Qualitativa : Solo categorie, i dati non possono essere disposti in alcun ordine,
Ordinale	Qualitativa : Le categorie sono ordinate, ma non possono essere calcolate le differenze tra loro
Numerica	discreta Quantitativa : Conteggi
	continua Quantitativa : Variabile che può essere misurata, forniscono informazioni sulla loro grandezza

La scala numerica crescente è la più sicura.
(in quella nominale non ci sono relazioni matematiche tra le categorie)

Analisi esplorativa dei dati: (statistica descrittiva)

1. **Rappresentazione** grafica
2. **Centro** un valore rappresentativo medio
3. **Variazione** una misura di quanto i dati siano dispersi al loro interno
4. **Valori estremi o outlier** valori che si discostano dalla maggioranza dei dati



Metodi per la descrizione delle variabili:

- Variabili Qualitative

Rappresentazioni grafiche

Istogrammi

Torte

Mappatura

Tabelle di frequenza

Frequenza assoluta

Frequenza relativa

Frequenza cumulativa

Frequenza valida

Tavole di contingenza

Frequenze assolute

Frequenze relative al totale di riga

Frequenze relative al totale di colonna

Frequenze relative al totale di generale

Indicatori per la scala nominale ordinale:

Frequenza assoluta (f.a):

numero di casi che presentano una certa caratteristica

Frequenza relativa (f.r):

numero di casi che presentano una certa caratteristica diviso per l'effettivo (somma delle frequenze).

Frequenza cumulativa

Somma della frequenza (assoluta o relativa) di ogni modalità alle modalità precedenti molto utile con variabili qualitative in scala ordinale

Frequenza valida (f.%):

numero di casi che presentano una certa caratteristica diviso per l'effettivo e rapportato a 100. Usato per fare confronti nella letteratura.

Istogramma: per descrivere

Ogiva: rappresentare velocemente situazioni sommate tra loro. Unire categorie.

Tavole di contingenza: mettere a confronto due variabili in scala. Meglio mettere in percentuale il campione piuttosto che esprimere con numero assoluto.

- Variabili Quantitative

Rappresentazioni grafiche (sono le stesse delle variabili qualitative)

- Istogrammi
- Torte
- Mappatura

Indicatori di posizione

- Media
- Mediana
- Moda
- Media Pesata
- Quintili

Indicatori di dispersione

- Scarto semplice
- Deviazione Standard
- Varianza
- Coefficiente di variazione
- Intervallo interquartile
- Errore standard

Per prima cosa devo creare degli intervalli e posso usare due regole:

Regola di Sturges: tutti gli intervalli sono della stessa ampiezza, una volta determinati ho fatto le categorie e devo verificare la frequenza dentro le categorie.

Densità di frequenza: intervalli di ampiezze diverse: frequenza della categoria/frequenza di quell'intervallo.

Intervalli stessa ampiezza:

calcolo quante classi posso fare con il mio campione, quindi qual è la taglia del campione. Con la regola di Sturges si costruisce tutto in funzione del campione, in base al n di campioni sappiamo quante categorie. Successivamente si calcola l'ampiezza della categoria.

Abbiamo classi diverse, non possiamo più usare la frequenza ma la densità di frequenza che è: numero di soggetti con una data caratteristica/ampiezza di quella classe.

Misure di tendenza centrale:

rappresenta e presenta il campione, la sua caratteristica:

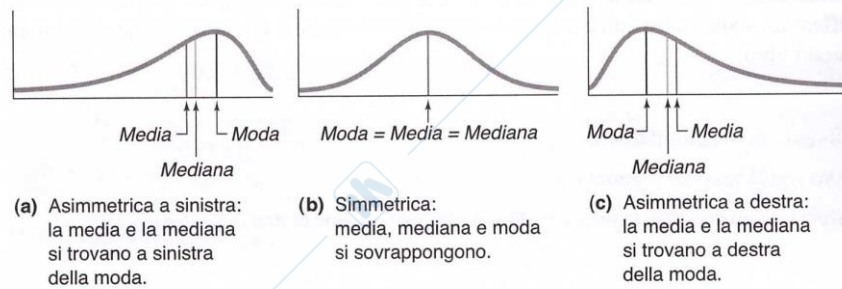
Media: quando si tratta della popolazione campionaria è indicata con: μ . Quando si parla di campione è indicata con \bar{X} (con barra sopra). Risente molto degli estremi.

Mediana: valori in ordine crescente, la mediana divide l'insieme in due parti uguali. Se il numero è dispari: $(n+1)/2$. Se è pari, al centro ho due valori e la mediana è la media dei due valori. Il limite è che può avere un valore che non appartiene all'insieme.

Moda: il valore che si presenta con la frequenza più alta. Non è unica, può anche non esserci la moda.

Moda pesata: i dati possiedono diversi gradi di importanza.

Mediante i tre indicatori



a = Asimmetria negativa (valori più bassi nella coda)

c = Asimmetria positiva (valori più alti nella coda)

Misure di variabilità o dispersione:

- **Intervallo di variazione (o range):**

- Il **campo di variazione** di una distribuzione è la differenza fra il valore massimo e il valore minimo rilevati

- **Intervallo** valore minimo e valore massimo

NB l'utilità è limitata tiene conto solo di 2 valori dell'insieme dei dati

(meno interpreta la variabilità, non so i valori come si posizionano rispetto agli altri)

- **Scarto semplice: IMP:**

relazione tra ogni singolo valore e il valore medio.

• Scarto semplice della popolazione

$$- SS_i = X_i - \mu$$

• Scarto semplice del campione

$$ss_i = X_i - \bar{X}$$

- **Varianza:** valore medio di tutti gli scarti.

Varianza di popolazione

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

Varianza nel campione

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

n-1= gradi di libertà.

È al quadrato per evitare il problema di avere valori negativi.

- **Deviazione standard:**

è la radice quadrata della varianza. Quando elevo al quadrato anche l'unità di misura è elevata al quadrato, quindi avrei due unità di misure un po' diverse. È l'escamotage che permette di associare due valori con stessa unità di misura.

- Coefficiente di variazione

$$CV = \frac{\sigma}{\mu} * 100$$

$$CV = \frac{s}{\bar{X}} * 100$$

- La DS è utile per indicare la dispersione di un solo insieme di dati
- la CV serve per confrontare la dispersione di due insiemi di dati con unità di misura diverse (es altezza e peso)
- La CV si utilizza anche per la stessa variabile ma in situazioni diverse (es altezze bambini e adulti)

Se è maggiore del 30% è distribuita normalmente

Misure di posizione:

- **Mediana** è il valore centrale di un insieme di dati ordinati
- **Percentili** dividono i dati ordinati in 100 gruppi comprendenti ciascuno 1% dei valori e vengono indicati con $P_1 P_2 P_{99}$
- **Quartili** dividono i dati ordinati in quattro parti uguali comprendenti ciascuno 25% dei valori
 - Q_1 separa il 25% inferiore dei dati dal 75% superiore dei dati (cioè il 25% dei dati è minore o uguale a Q_1)
 - Q_2 separa il 50% inferiore dei dati dal 50 superiore dei dati MEDIANA
 - Q_3 separa il 75% inferiore dei dati dal 25% superiore dei dati

$$Q_1 = \frac{n+1}{4}$$

$$Q_2 = \frac{2*(n+1)}{4} = M$$

$$Q_3 = \frac{3*(n+1)}{4}$$

Range o scarto interquartile:

differenza tra il valore che chiude il terzo quartile e il primo.

L'intervallo interquartile è un indicatore di dispersione che valuta quanto ampio sia l'intervallo di misure all'interno del quale sono compresi il 50% dei soggetti con misure più vicine alla mediana.

Quando media e deviazione standard diversi = grande dispersione.

Outlier:

Outlier è un valore estremo che dobbiamo analizzare per capire se è un errore o un dato reale.

Può avere un influenza notevole sulla media

Può avere un influenza notevole sulla DS

Può avere un influenza notevole sulla scala con cui è disegnato un istogramma.

Per graficare il dato senza avere deformazione del risultato: BoxPlot.

Evidenza: valore minimo, valore massimo, Q_1 , Q_2 , Q_3 .

Leggo i risultati in modo molto immediato.

Quando la mediana è molto vicina al Q_1 = asimmetria positiva

Quando la mediana è molto vicina al Q_3 = asimmetria negativa

Valore atteso:

Uno dei concetti più importanti nella terminologia statistica è il valore atteso

IL VALORE CHE POTREMMO ASPETTARCI DI RILEVARE IN UNA SITUAZIONE DI NORMALITA'

Un valore al quale si fa riferimento, ma non è detto che corrisponda.

Importante: verificare i parametri perché possono influenzare il risultato

ESEMPIO DI ESERCIZI:

E' data la distribuzione di 1863 pazienti secondo il numero di cicli di terapia:

Numero di cicli terapia	Numero di pazienti
1	332
2	440
3	412
4	401
5	177
6	101

Calcolare per ogni classe della distribuzione:

- le frequenze relative (serve per il mio studio, quella percentuale per confronto con la letteratura)
- le frequenze cumulate crescenti
- le frequenze cumulate crescenti percentuali

variabile in scala numerale ordinale: possiamo parlare solo di frequenza

	B	C	D	E	F	G
	Num. Comp.	Num. famiglie	Freq.rel	$X \leq x_j$	Freq.cum	Freq.rel.cum
4	1	332	0,18	≤ 1	332	0,18
5	2	440	0,24	≤ 2	772	0,41
6	3	412	0,22	≤ 3	1184	0,64
7	4	401	0,22	≤ 4	1585	0,85
8	5	177	0,10	≤ 5	1762	0,95
9	6	101	0,05	≤ 6	1863	1,00
10		1863	1,00			

ESERCIZIO 2:

Si rileva il peso di 20 pazienti adulti ricoverati in un reparto di medicina interna per verificare se esiste la necessità di dotare il reparto di materassi rinforzati.

(Si deciderà per l'acquisto di un materasso se almeno il 15% di pazienti pesa più di 90 kg).

Peso (kg) :

80 83 92 65 75 83 90 80 86 86 92 100 75 68 83 92 72 75 83 75

Utilizzando la funzione FREQUENZA() evidenzia :

Le frequenze assolute

Le frequenze relative percentuali

Le frequenze cumulate percentuali

Cumulativa fino a 90, se quel che rimane è più o meno 25%.

ESERCIZIO 3:

Le donne in gravidanza (entro il 4o mese) che vengono seguite in un centro dietologico pesano rispettivamente (pesi in kg): 64.3; 65.2; 70.0; 54.5; 58.8; 81.5; 61.0; 62.0. Qual'è la media? e la mediana? I dati suggeriscono una forte asimmetria della distribuzione del Peso?

Prima metto in ordine i dati, calcolo media, mediana e deviazione standard.

Asimmetria (calcola i quartili)

PROBABILITA'

- La teoria della probabilità è alla base dell'inferenza statistica
- Evento *raccolta di esiti di un esperimento*
- Evento elementare *evento che non può essere suddiviso in eventi più semplici*
- Evento raro *bassa probabilità che si verifichi*
- Spazio campionario *composto da tutti gli eventi elementari*

Probabilità oggettiva:

P Classica a priori

Tutti gli eventi hanno la stessa probabilità di verificarsi

$$P(E) = m / N$$

L'evento E si può verificare in N modi mutuamente esclusivi e m sono quelli che possiedono la caratteristica E, semplicemente facce di un dado 6(N) ogni faccia ha solo 1 caratteristica (m)

Gli eventi debbono essere equiprobabili

P con frequenza relativa

$$P(E) = m / n$$

Ripetendo un processo n volte e un evento E si verifica m volte praticamente cerco la frequenza relativa di successo di E

Gli eventi non debbono essere equiprobabili

- La probabilità di un evento impossibile è 0
- La probabilità di un evento che si verifichi sicuramente è 1
- Per ogni evento A la P(A) è compresa tra 0 e 1 $0 \leq P(A) \leq 1$
- Nozioni di base:
 - Regola della somma $P(A \text{ o } B) = P(A) + P(B) - P(A \text{ e } B)$ probabilità in cui sia A sia B si verificano quando gli eventi non sono indipendenti. stabilire se A e B sono disgiunti
 - Eventi complementari $P(\bar{A})$
 - Regola del prodotto $P(A \text{ e } B) = P(A) * P(B)$ (indipendenti) prodotto probabilità

$$P(A \text{ e } B) = P(A) * P(B | A) \text{ (dipendenti)}$$

Regola del prodotto: l'evento A e B sono dipendenti

Probabilità condizionata

- $P(B | A)$ la probabilità che si verifichi B alla luce del fatto che si sia già verificato A
- Eventi dipendenti:
 - $P(A \text{ e } B) = P(A) * P(B | A)$
- Eventi indipendenti $P(B | A) = P(B)$
- $P(A \text{ e } B) = P(A) * P(B)$

Probabilità di "almeno uno"

- Almeno uno significa uno o più
- Il complementare di "almeno uno" è "nessuno"
- $P(\bar{A}) = P(\bar{A}) * P(\bar{A}) * \dots * P(\bar{A})$
- $P(A) = 1 - P(\bar{A})$

Probabilità condizionata

$$P(B | A) = \frac{P(A \cap B)}{P(A)}$$

Tabella 3.1 Risultati del test di gravidanza.

	Test positivo (gravidanza)	Test negativo (no gravidanza)
La donna è incinta	80	5
La donna non è incinta	3	11

a) P che la donna sia positiva al test dato che è incinta
 $P(\text{positiva} | \text{incinta}) = P(\text{incinta e positiva}) / P(\text{incinta})$

a) P che la donna sia incinta dato che è positiva al test
 $P(\text{incinta} | \text{positiva}) = P(\text{positiva e incinta}) / P(\text{positiva})$

Variabili aleatorie:

- Una *variabile aleatoria* è una variabile (x) che assume un singolo valore numerico, determinato in base al caso, per ciascuno degli esiti dell'esperimento
 - Una *distribuzione di probabilità* è un grafico, una tabella o una formula che fornisce la probabilità di ciascun valore della variabile aleatoria
- Proprietà
- $\sum P(x) = 1$, la somma di tutte le probabilità vale 1.
 - $0 \leq P(x) \leq 1$

Se la somma è maggiore di uno, la variabile aleatoria assume contemporaneamente due valori e si sommano due volte.

Distribuzione di probabilità:

La relazione tra una variabile aleatoria e la probabilità che questa si verifichi possiamo sintetizzarla con una funzione detta distribuzione di probabilità cioè **curva di distribuzione**

- Per ogni variabile è importante definire la curva di distribuzione più appropriata*
- Ci permette di calcolare la probabilità di un evento cioè che probabilità ho quando la variabile aleatoria assume un determinato valore*

Distribuzione con frequenza relativa:

Progetto alimentare con un ricorso al programma di assistenza di alimenti negli ultimi 12 mesi

Gli eventi sono mutuamente esclusivi quindi tutte le probabilità sono comprese tra 0 e 1 e

La probabilità coincide con la frequenza relativa

N. Volte ricorso al programma	Frequenza /P(x)	P(x) %
1	62/ 0.2088	20.88
2	47/ 0.1582	15.82
3	39/0.1313	13.13
4	39/0.1313	13.13
5	58/0.1953	19.53
6	37/0.1246	12.46
7	4/0.0135	1.35
8	11/0.0370	3.7
Totale	297/ 1	100

È la più limitante, devo essere nella stessa identica situazione, anche il campione deve essere simile al campione su cui ottengo i risultati.

Frequenza cumulata:

E' utile per rispondere a domande :

Qual è la probabilità che una famiglia scelta a caso abbia fatto ricorso al programma un numero di volte minore o uguale a due?

N.Volte ricorso al programma	Frequenza cumulata P(x) cumulata	P(x)%
1	0.2088	20,9
2	0.3670	36,7
3	0.4983	49,8
4	0.6296	62,9
5	0.8249	82,5
6	0.9495	95,0
7	0.9630	96,3
8	1.0000	100
Totale		

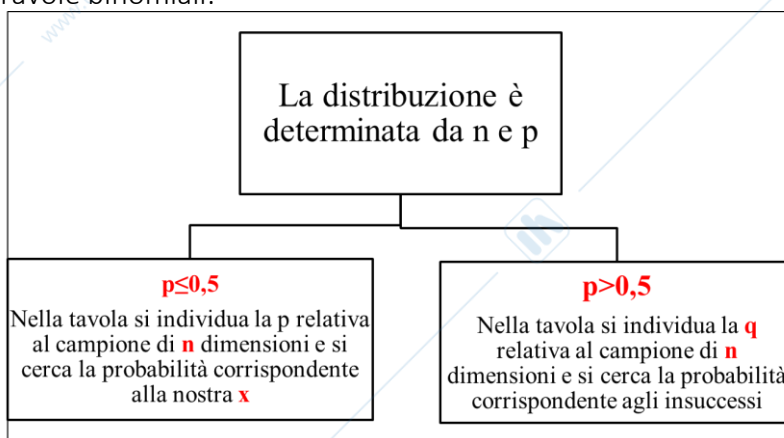
Distribuzione binomiale: $P(x | n) = {}_n C_x p^x q^{(n-x)}$

- L'esperimento ha *un numero fissato di prove*
- Le prove devono essere *indipendenti* (ogni prova non influenza la probabilità delle altre prove)
- L'esito di ogni prova deve appartenere a *due sole categorie (dicotomica)*
- Le probabilità devono rimanere costanti in ogni prova

Si usa quando vogliamo vedere la probabilità di x successi in n prove.

- p successo in una singola prova su n
- $q = (1-p)$ insuccesso in una singola prova su n
- n numero fissato di prove
- x numero di successi in n prove
- $P(x)$ probabilità di avere esattamente x successi in n prove

Tavole binomiali:



Tavole binomiali, unico limite: sono state calcolate con la probabilità massima del 50%, quando è maggiore del 50% si lavora sugli insuccessi.

Ogni valore è cumulato, se prendo solo un valore devo sottrarlo al precedente.

Quando voglio sapere il valore secco faccio il valore richiesto – quello che precede.

Senno cerco il complementare quindi gli insuccessi, quando ne cerco esempio massimo 4.

Variabile aleatoria discreta:

- Se, in base a una data ipotesi, la probabilità di un particolare evento osservato è estremamente bassa concludiamo che l'ipotesi è probabilmente non corretta ($P(13)=0.001$)
- **Risultato Insolitamente alto:** x successi su n prove e un numero di *successi insolitamente alto* se $P(x \text{ o più})$ è molto piccola (0.05 o <0.05)
- **Risultato Insolitamente basso:** x successi su n prove e un numero di *successi insolitamente basso* se $P(x \text{ o meno})$ è molto piccola (0.05 o <0.05)

Due estremi con P minore di 0.5

Valore atteso di una variabile aleatoria è il valore medio (prodotto degli eventi per la probabilità) dei risultati.

$$\mu = np$$

la media deve essere sempre accompagnata dalla deviazione standard.

$$\sigma^2 = npq$$

$$\sigma = \sqrt{npq}$$

Distribuzione di Poisson

- Se x è il numero di volte in cui si è verificato un qualche evento casuale in un intervallo di tempo o spazio la probabilità che x si verifichi è
- $P(x) = e^{-\mu} \mu^x / x!$ $e=2.71828$
- μ è il numero medio di volte in cui l'evento casuale si verifica nell'intervallo
- X il numero di volte in cui l'evento si verifica in un intervallo di tempo prefissato
- Gli eventi devono verificarsi in modo casuale
- Ogni evento deve verificarsi indipendentemente dagli altri
- La media è μ

La deviazione standard è $\sigma = \sqrt{\mu}$

La condizione è di conoscere la media, e la risposta è sempre riferito allo stesso periodo di tempo in cui è stata calcolata la media.

DIFFERENZE:

Binomiale

È determinata dalla taglia del campione n e dalla probabilità p

I possibili valori assunti dalla variabile aleatoria x sono $0,1,2,\dots,n$

Poisson

È determinata solo dalla media

I possibili valori assunti dalla variabile aleatoria x sono $0,1,2,\dots$ senza limite

Distribuzione uniforme

Una variabile aleatoria continua (espressa in scala continua) ha una distribuzione uniforme se i suoi n valori si distribuiscono uniformemente nell'intervallo dei valori possibili, il grafico di una distribuzione di P continua è detta

curva di densità:

-L'area sottesa alla curva=1

-Ogni punto della curva deve avere altezza ≥ 0

I valori si approssimano ad una curva. Mediante la Distribuzione Normale o Gaussiana:

- Se una variabile aleatoria continua ha una distribuzione che presenta un grafico a forma di campana si dice che ha una distribuzione normale. L'equazione della distribuzione normale è

$$f(x) = \frac{e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}}{\sigma \sqrt{2\pi}}$$

Caratteristiche della Distribuzione Normale

- π e e sono le costanti matematiche note, μ e σ rispettivamente media e deviazione standard
- È simmetrica intorno alla media
- La media mediana e moda coincidono
- L'area totale sotto la curva vale 1
- $\pm 1\sigma$ 68%, $\pm 2\sigma$ 95%, $\pm 3\sigma$ 99,7% dell'area totale
- La distribuzione normale è completamente specificata dai parametri μ σ
- La media ± 1 deviazione standard deve contenere il 95 % dei valori

Media + 0,67(coefficiente) *deviazione standard è il terzo e primo quartile. Per passare da deviazione standard a quartile)

Perché si possa fare riferimento alla distribuzione di Gauss dobbiamo standardizzare i nostri valori :

- È possibile trasformare una distribuzione normale non standard in una distribuzione normale standard usando la seguente formula

$$z = \frac{x - \mu}{\sigma}$$

(a) Distribuzione normale non standard

(b) Distribuzione normale standard

x = mio valore

le tavole gaussiane cumulano a sinistra.

Formule su excel:

Funzione :

=DISTRIB.NORM.N(x ; m ; ds ; cumulativo)

Con VERO \rightarrow probabilità, unico valore non cumulativo

=INV.NORM.S(probabilità) restituisce il valore z

=INV.NORM.N(prob; m ; ds) restituisce il valore x

Distribuzione campionaria della media:

- **La distribuzione campionaria della media** è la distribuzione di probabilità delle medie campionarie, quando tutti i campioni hanno la stessa taglia n

Campionamento con reimmissione:

Quando il campionamento è fatto con reimmissione il numero di tutti i possibili campioni è N^n

Campionamento senza reimmissione:

Quando il campionamento è fatto senza reimmissione il numero di possibili campioni è

$${}_N C_n = \frac{N!}{n!(N-n)!}$$

La media di tutti i campioni corrisponde con la media della popolazione.

Più è piccolo l'errore standard meno è la differenza tra tutti i campioni.

Errore standard: stima del valore del campione

- **Variabile distribuita normalmente allora la distribuzione avrà queste caratteristiche**

La distribuzione della media nei vari campioni sarà ancora normale

La media delle medie sarà uguale alla media della popolazione N

La varianza della distribuzione sarà uguale all'errore standard

- **Variabile non distribuita normalmente allora la distribuzione avrà queste caratteristiche**

- Sarà data da una trasformazione con la formula :

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- se $n > 30$ la distribuzione delle MC può essere approssimata ad una D.N.

APPLICAZIONE:

- La distribuzione della media nei vari campioni sarà ancora normale
- La media delle medie sarà uguale alla media della popolazione N
- La varianza della distribuzione sarà uguale all'errore standard
- Supponiamo sia noto che, in una certa popolazione umana, la lunghezza del cranio è approssimativamente distribuita come una normale con una media 185.6 mm ed una deviazione standard di 12.7 mm. Qual è la probabilità che un campione casuale semplice di dimensione 10, estratto da questa popolazione abbia una media più grande di 190 mm?

Interpretazione:

- Per stimare un parametro della popolazione alcune statistiche delle distribuzioni campionarie risultano adeguate altre non sono così precise queste statistiche sono:
- **Stimatori non distorti**: media, varianza, proporzione
- **Stimatori distorti** : mediana, range, deviazione standard

Calcola a media: media * frequenza / numero totale di n

Non si può accettare valori scala numerica continua senza intervalli contigui es: 5-8.3, 8.4-9.6

Studi epidemiologici:

Random

Epidemiologia: studio della frequenza delle malattie.

Lo studio epidemiologico è uno studio osservazionale, non ne intervengo, sono free living, non c'è nessuna manipolazione.

Ci sono due tipi di popolazione:

coorti fisse: ne escono solo quando diventano casi

popolazione dinamica: vista per un periodo più o meno lunga poi non viene più vista.

STUDIO TRASVERSALE: studio fotografia del periodo, si individuano le malattie presenti, le caratteristiche dei soggetti. È uno studio molto veloce, non costa come studio e si osservano particolari. Ha un obiettivo ma non lo mantiene sullo stesso campione. La popolazione è dinamica. Svantaggio è che risente molto del periodo in cui viene svolta. Non ci consentono di vedere le relazioni temporali, non si sa cosa è successo prima e cosa succederà quindi non si possono fare ipotesi.

STUDIO PROSPETTICO LONGITUDINALE: popolazione free living, in un periodo detto follow up vengono visti i soggetti. Non finisce mai, si può continuare all'infinito. Studio di coorte fissa, sempre stessa popolazione, escono solo i casi. Si possono anche vedere gli effetti sul tempo. Sono molto costosi, esistono problemi etici. È osservazionale, si possono vedere situazioni allarmanti e vedere cosa succede. No casi rari.

STUDIO PROSPETTICO CASO CONTROLLO: si prendono i soggetti esposti al caso di rischio verso chi non è esposto al parametro e vedere chi tra loro manifesta la patologia. Valuta se c'è fattore di rischio in base ai criteri epidemiologici. Sono molto veloci, poichè i soggetti sono già esposti. Possono essere usati per le patologie rare. Controllo costante della popolazione. Problemi etici, molto costoso e si va in drop out, i soggetti possono anche rifiutarsi di continuare. Se si sceglie male il gruppo di controllo lo studio può essere fallace. Problemi di bias di selezione. Non si può parlare di prevalenza e incidenza.

STUDIO RETROSPETTIVO: più usato. Si parte da un gruppo caso, soggetti già ammalati verso chi non ha la patologia e si verificano i parametri prima di ammalarsi. Non ci sono popolazioni particolari, solo quelli con patologia. Guardo al passato, non al futuro. Si può calcolare l'ODD RATIO

Atraverso indicatori possiamo avere info molto importanti. Si possono dividere in

Misure di insorgenza:

PREVALENZA numero di casi con patologia presente

INCIDENZA: n casi prima del controllo ritenuti sani. Si deve avere studio prospettico di coorte.

Il loro legame tra prevalenza e incidenza è strettissimo. I nuovi casi si aggiungono alla prevalenza.

Aumenta più velocemente in base all'incidenza. Rimane costante se incidenza è =0.

È in base alla prevalenza che si faranno poi nuovi studi per vedere le cause della prevalenza della malattia. Nuovi casi che erano sani prima del controllo.

RISCHIO: probabilità di sviluppare la malattia in un certo periodo di tempo. Numero di nuovi casi tra i soggetti esposti a quel fattore. Soggetti che già avevano un parametro.

Modelli di rischio

RISCHIO ASSOLUTO: ce l'hanno tutti sia che ha parametro alterato chi no.

RISCHIO RELATIVO: rischio assoluto tra gli esposti/ rischio assoluto non esposti. Determina se il rischio è un fattore di rischio poichè se ho più casi tra esposti= fattore di rischio.

RISCHIO ATTRIBUIBILE: rischio assoluto esposti-rischio assoluto non esposti (RA in %)

RISCHIO RELATIVO: studio prospettico caso controllo se è > di 1 è fattore di rischio

ODDS RATIO: è valido se presente intervallo di confidenza (calcolato sulla popolazione, legame tra studio di campione e risultato sulla popolazione, possibili valore che assume lo stesso indicatore) di 95%

MORBOSITÀ: frequenza di malattia in un campione di popolazione

MORBILITÀ: quanti reazioni avverse, quanti rigetti. Non sono nuovi casi, tasso e rapporti legati perché scritte in esponente di 10

blinding: paziente e medico non conoscono la molecola

studi cross over: si studiano due trattamenti incrociati, lo studio mi impone che ci fosse una sospensione prima dell'incrocio. Ma è rischioso per chi ha la malattia, vengono usati per studi di medicina sportiva.

Errori:

validità interna è influenzata da scelta errata del gruppo di confronto, da stratificazione non corretta e numerosità troppo diversa tra i gruppi
validità esterna: come può essere compromesso il risultato

