

CDL in INFERMIERISTICA & OSTETRICIA

Statistica Medica

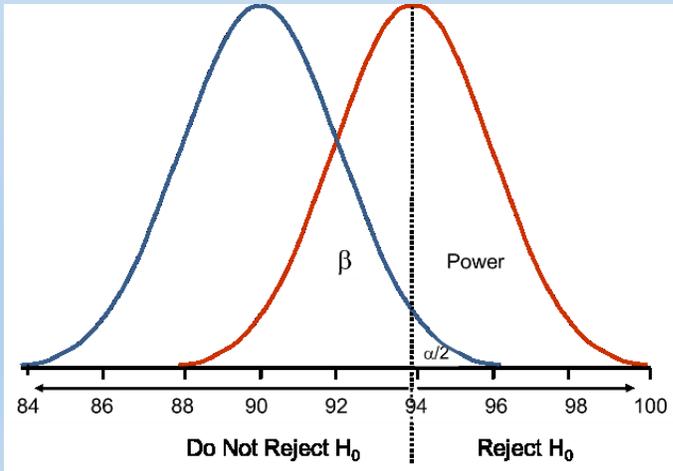
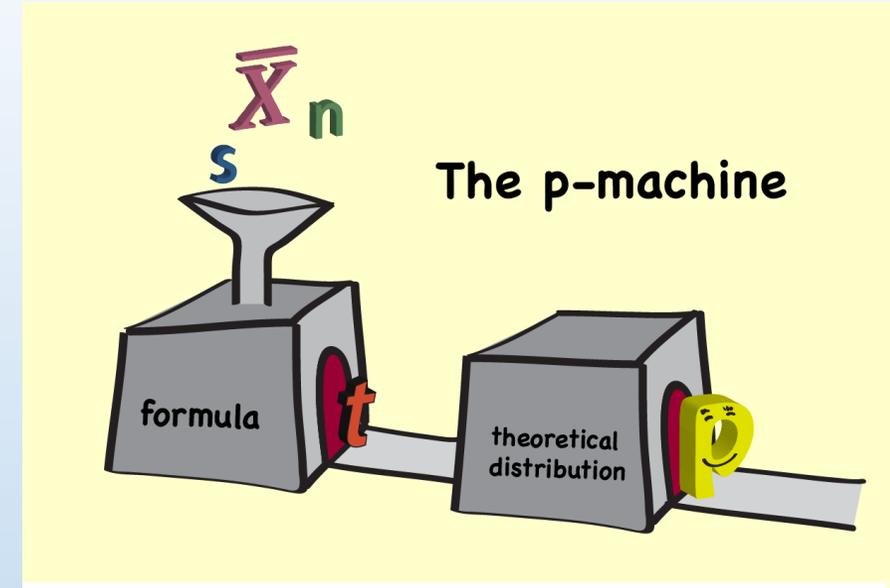
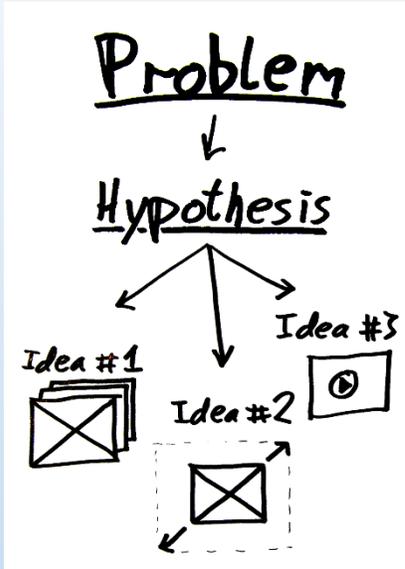
gbarbati@units.it

A.A. 2024-25

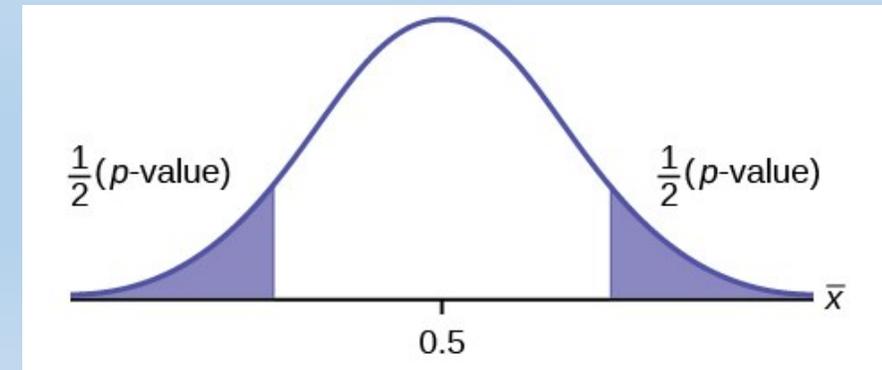


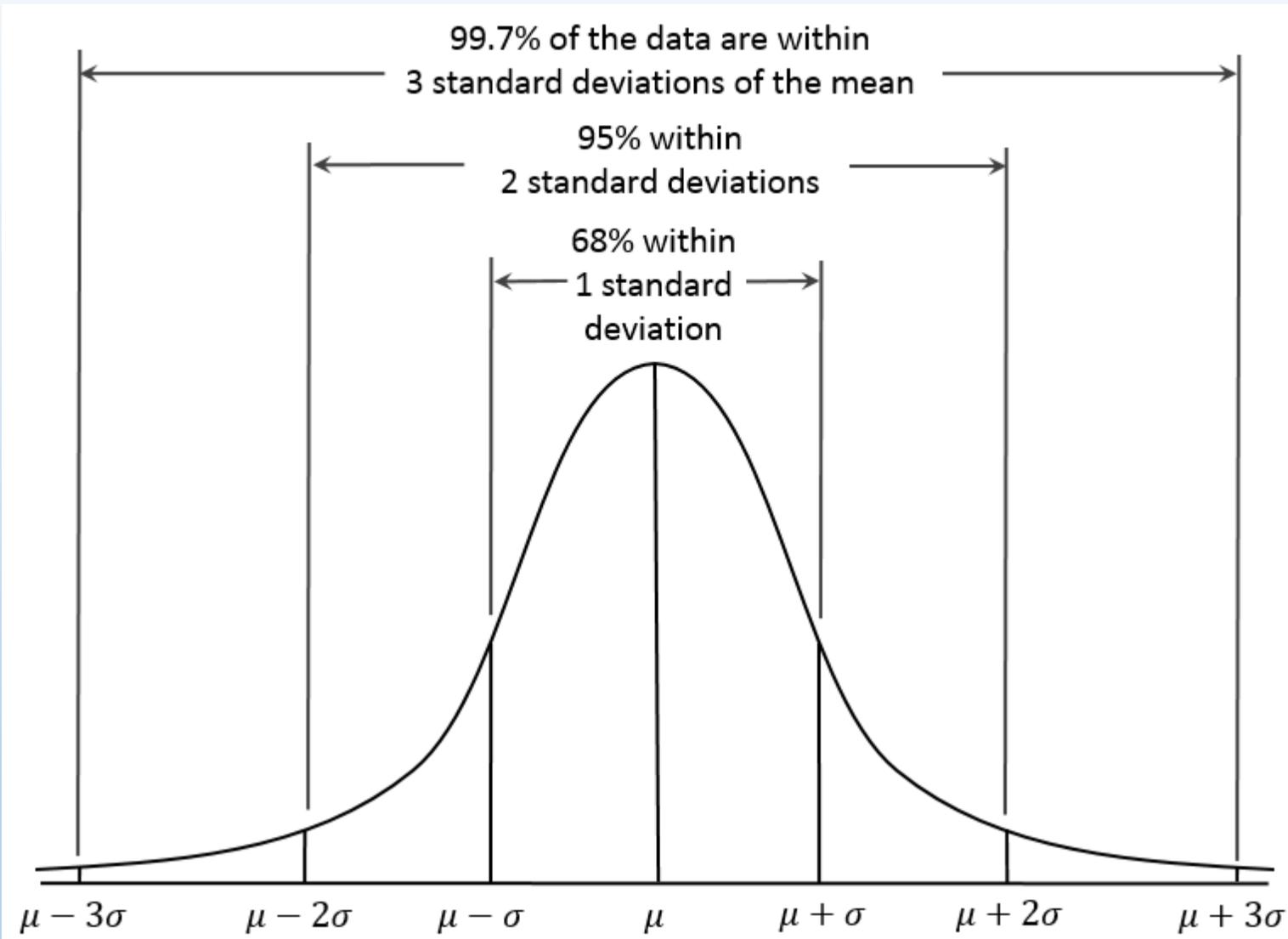
Sommario:

- Stima dei parametri
- Intervalli di confidenza
- Concetti generali del test di ipotesi



STATISTICS
MEAN NEVER
HAVING TO
SAY YOU'RE
CERTAIN



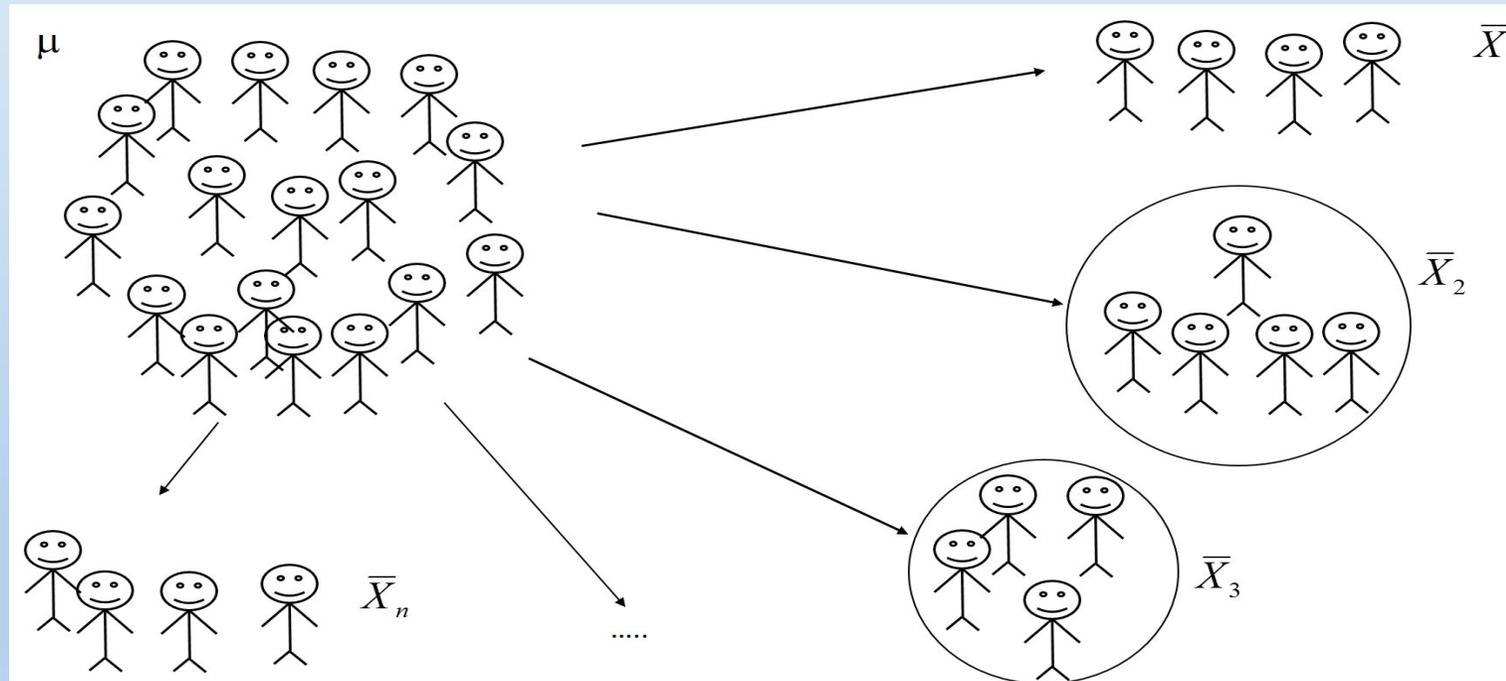


L'importanza della curva normale non risiede solo nella sua capacità di descrivere alcuni fenomeni su scala continua («forma» della distribuzione che andrebbe sempre verificata nei dati.../proprietà in probabilità).

Essa occupa un posto di rilievo nella teoria dell'inferenza...

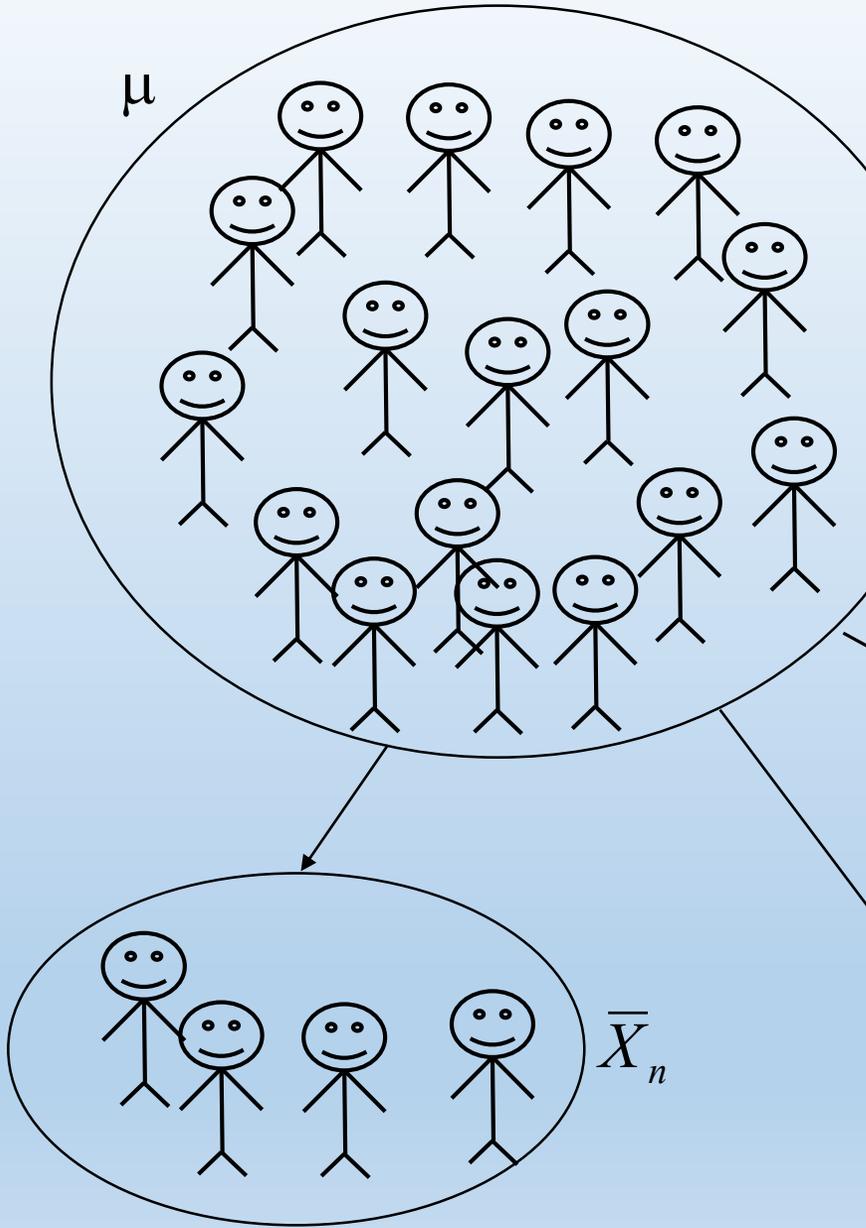
Teorema del limite centrale

La distribuzione di probabilità della somma (media) di un numero *elevato** di variabili aleatorie indipendenti e *identicamente distribuite*** tende distribuirsi normalmente, *indipendentemente* dalla distribuzione delle variabili originali.

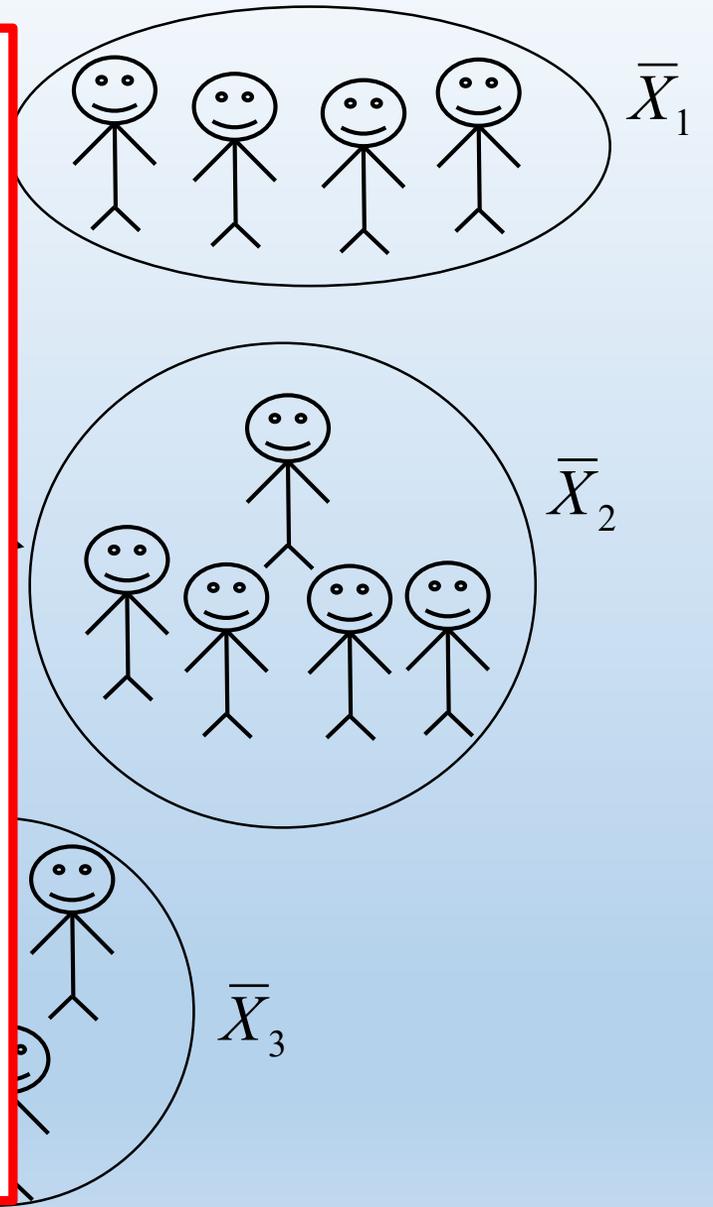


* Un campione di dimensione $n > 30$ è considerato «sufficientemente grande»

**sotto alcune condizioni



La media campionaria è essa stessa una VA perchè dipende dal meccanismo casuale di estrazione del campione...



Stima dei parametri

Obiettivo: Descrivere un fenomeno (su scala continua) oggetto del nostro studio.

Es: Stimare il parametro μ del fenomeno nella popolazione (**ignoto**) tramite un campione di dimensione n

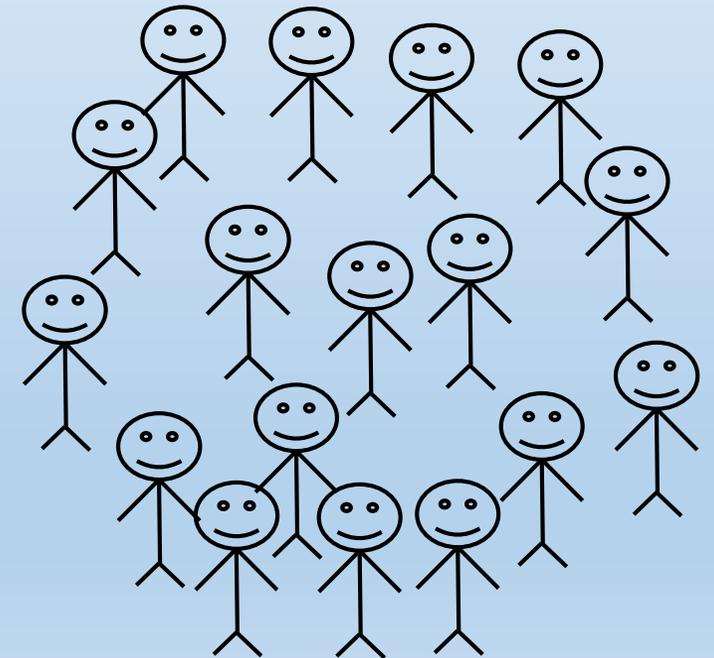
(anche la varianza del fenomeno σ^2 nella popolazione è **ignota**)

Cosa ci assicura la teoria dell'inferenza?

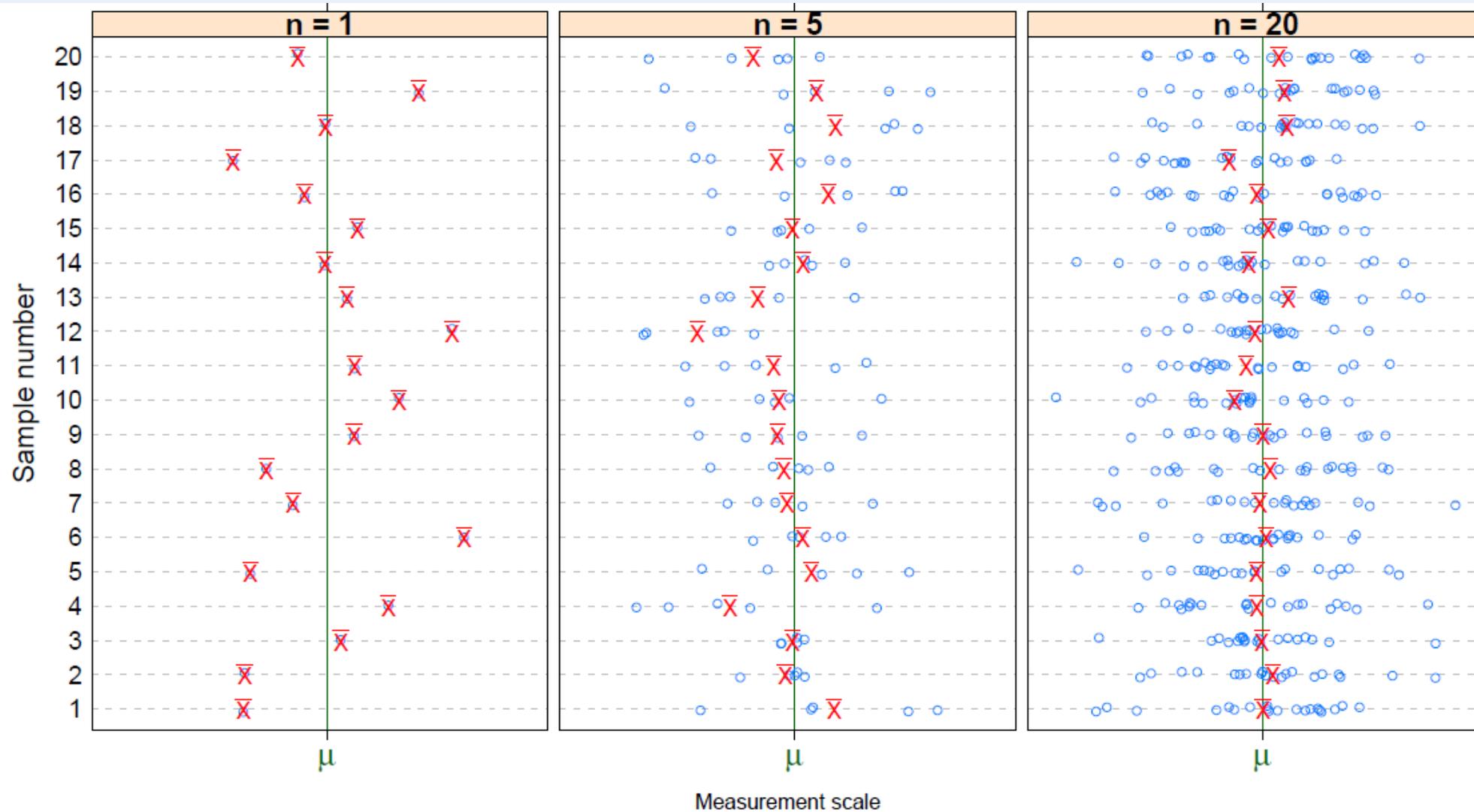
La VA media campionaria \bar{X} è un **buon** stimatore di μ

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\bar{X} \rightarrow \mu \text{ per } n \rightarrow \infty$$



Simulazione della distribuzione della media campionaria



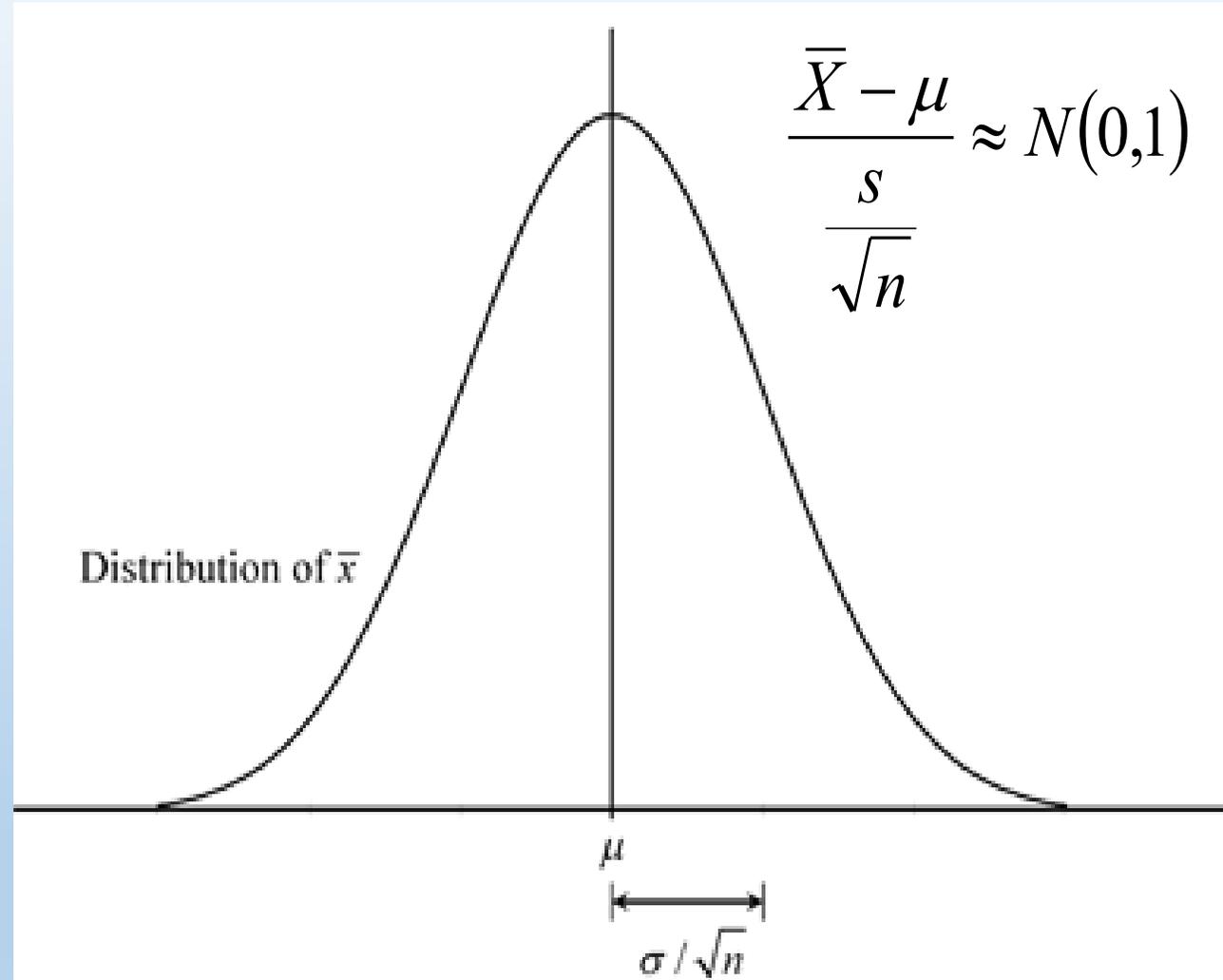
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\bar{X} \rightarrow \mu \text{ per } n \rightarrow \infty$$

L' «**errore**» che commettiamo utilizzando la media campionaria come stima della media ignota è pari a:

$$devst(\bar{X}) = \frac{\sigma}{\sqrt{n}} \rightarrow 0 \text{ per } n \rightarrow \infty$$

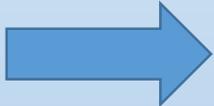
STANDARD ERROR



$\bar{X} \neq \mu$ praticamente sempre... anche se «per caso» $\bar{X} = \mu$...noi non lo sapremmo...

E quindi: quanto è «sbagliata» \bar{X} rispetto a μ ??

La teoria dell'inferenza ci dice che: $devst(\bar{X}) = \frac{\sigma}{\sqrt{n}} \rightarrow 0$ per $n \rightarrow \infty$

Ma la **varianza** della popolazione σ^2 non è nota  come stimare l'errore su \bar{X} ?

La teoria dell'inferenza ci assicura che s^2 è una buona stima di σ^2 :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$P(\text{LOWER LIMIT} < \mu < \text{UPPER LIMIT}) = 0.95 = (1-\alpha)$

$$P\left(-1.96 \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96\right) = 0.95$$

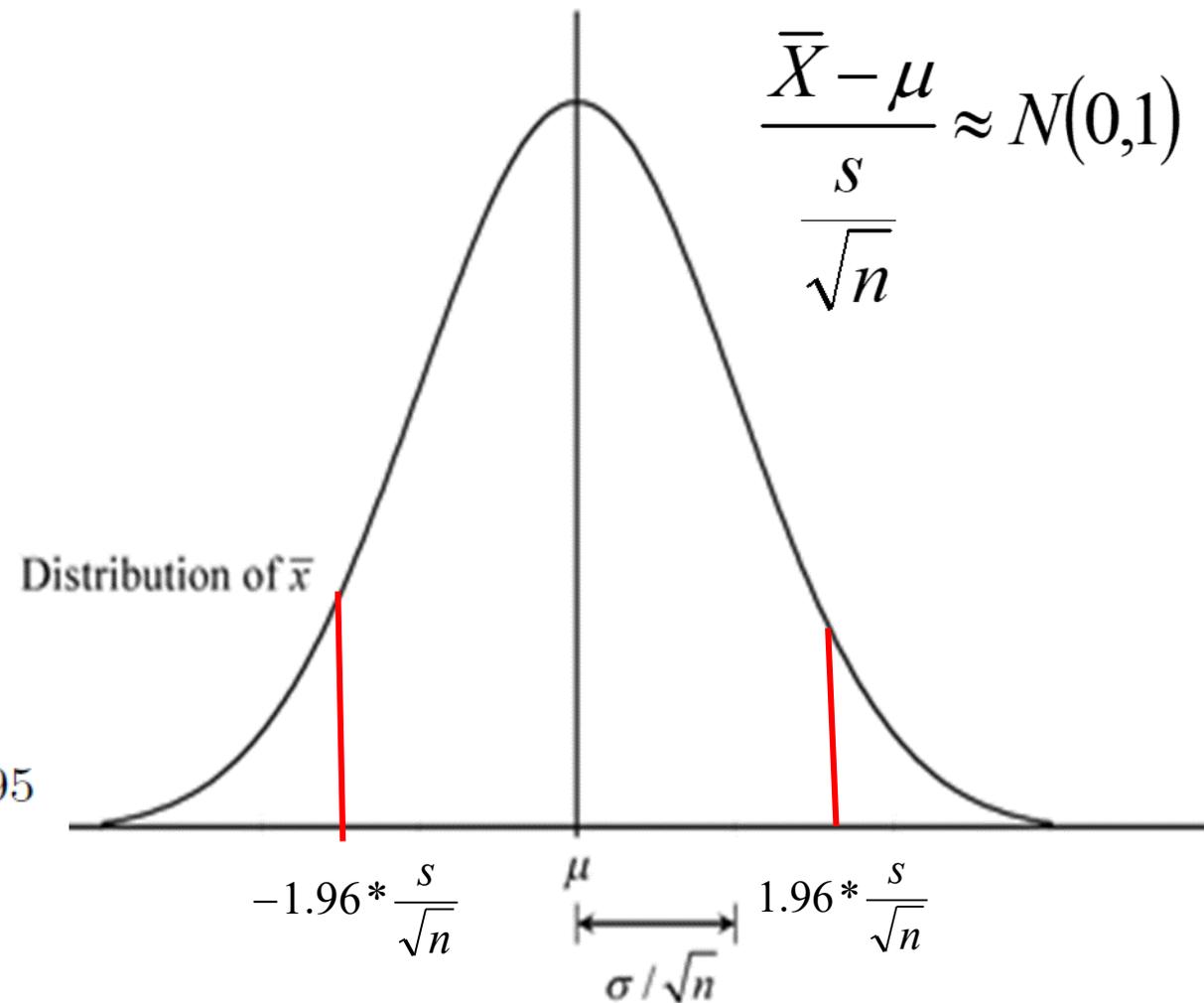
$$P\left(-1.96 \frac{\sigma}{\sqrt{n}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \frac{\sigma}{\sqrt{n}} \leq 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P\left(-1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P\left(-1.96 \frac{\sigma}{\sqrt{n}} - \bar{X} \leq \bar{X} - \mu - \bar{X} \leq 1.96 \frac{\sigma}{\sqrt{n}} - \bar{X}\right) = 0.95$$

$$P\left(1.96 \frac{\sigma}{\sqrt{n}} + \bar{X} \geq \mu \geq -1.96 \frac{\sigma}{\sqrt{n}} + \bar{X}\right) = 0.95$$

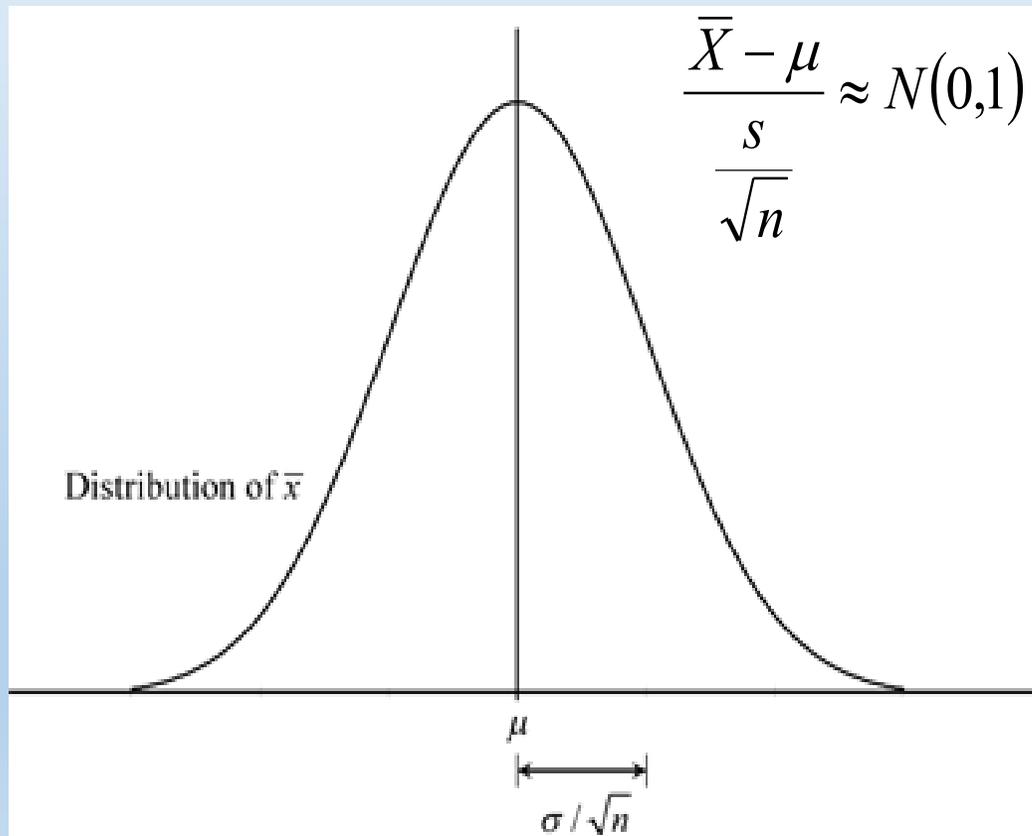
$$P\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$



INTERVALLO DI CONFIDENZA (IC)

$$\left[\bar{X} - \text{cost} * \frac{s}{\sqrt{n}}; \bar{X} + \text{cost} * \frac{s}{\sqrt{n}} \right]$$

Questo **intervallo** «*contiene*» il parametro μ con una probabilità determinata dalla costante utilizzata (nel lungo termine)...



Se vogliamo un IC al 95%:

$$\left[\bar{X} - 1.96 * \frac{s}{\sqrt{n}}; \bar{X} + 1.96 * \frac{s}{\sqrt{n}} \right]$$

Standard Error : SE

Confrontiamo queste due espressioni:

1. L'intervallo di confidenza contiene il parametro ignoto μ con una probabilità del X%
2. Il parametro ignoto μ ha una probabilità del X% di cadere in quell'intervallo...

Sembrano simili, ma nell'impostazione frequentista* la seconda è sbagliata.

Il parametro μ **NON E'** una variabile aleatoria; è un valore unico e non noto.

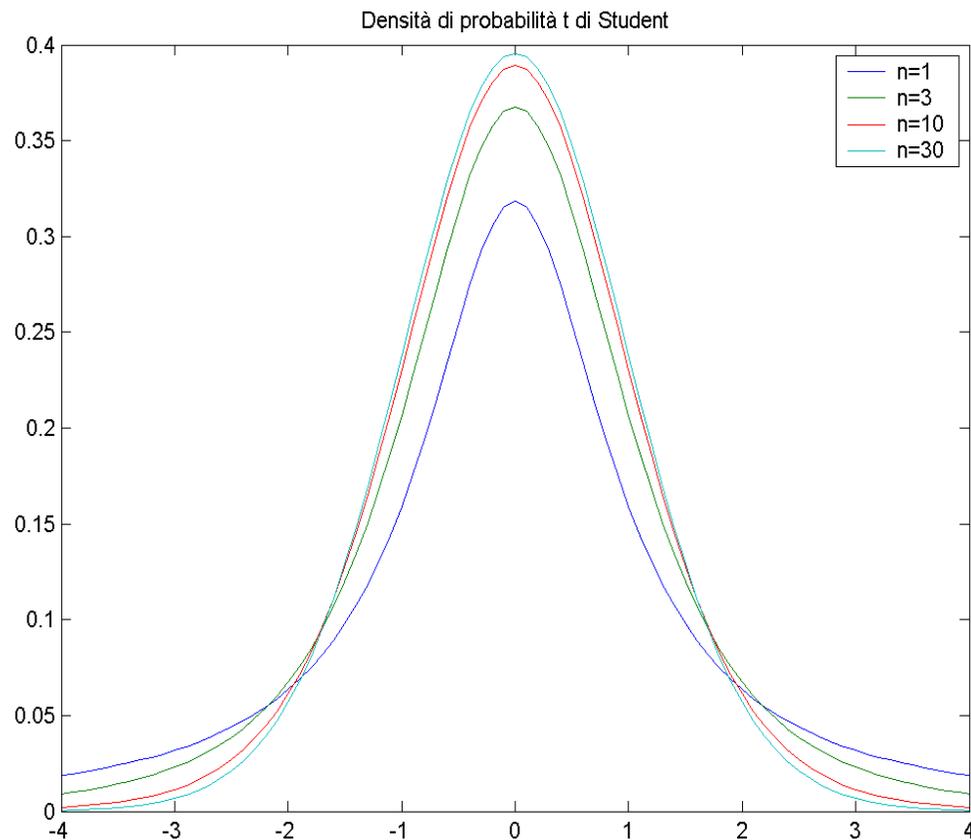
Per ogni specifico campione estratto abbiamo solo 2 possibilità: o μ è dentro l'intervallo o è fuori dall'intervallo.

Stiamo quindi facendo riferimento *ad una serie ipotetica di campionamenti ripetuti*, e stiamo dicendo che nel X% dei casi gli intervalli di confidenza calcolati conterranno μ . Nel restante 100-X% non lo conterranno.

*l'impostazione «**bayesiana**» che deriva dalla definizione soggettiva di probabilità cambia un po' le carte in tavola...

Correzione per n «piccolo»: distribuzione t di Student

Per una migliore stima dell'IC in piccoli campioni si definisce la VA di Student (t di Student):

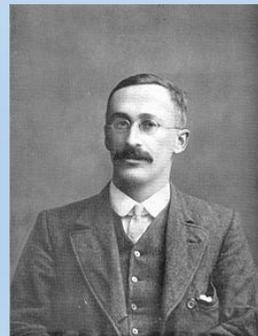
$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$


La distribuzione **t di Student** è molto simile alla gaussiana. Ma cambia di *forma* in relazione alla numerosità n del campione: tende ad avvicinarsi alla **distribuzione normale standard** $N(0,1)$, al crescere di n .

Per $n > 30$ le due distribuzioni sono indistinguibili.

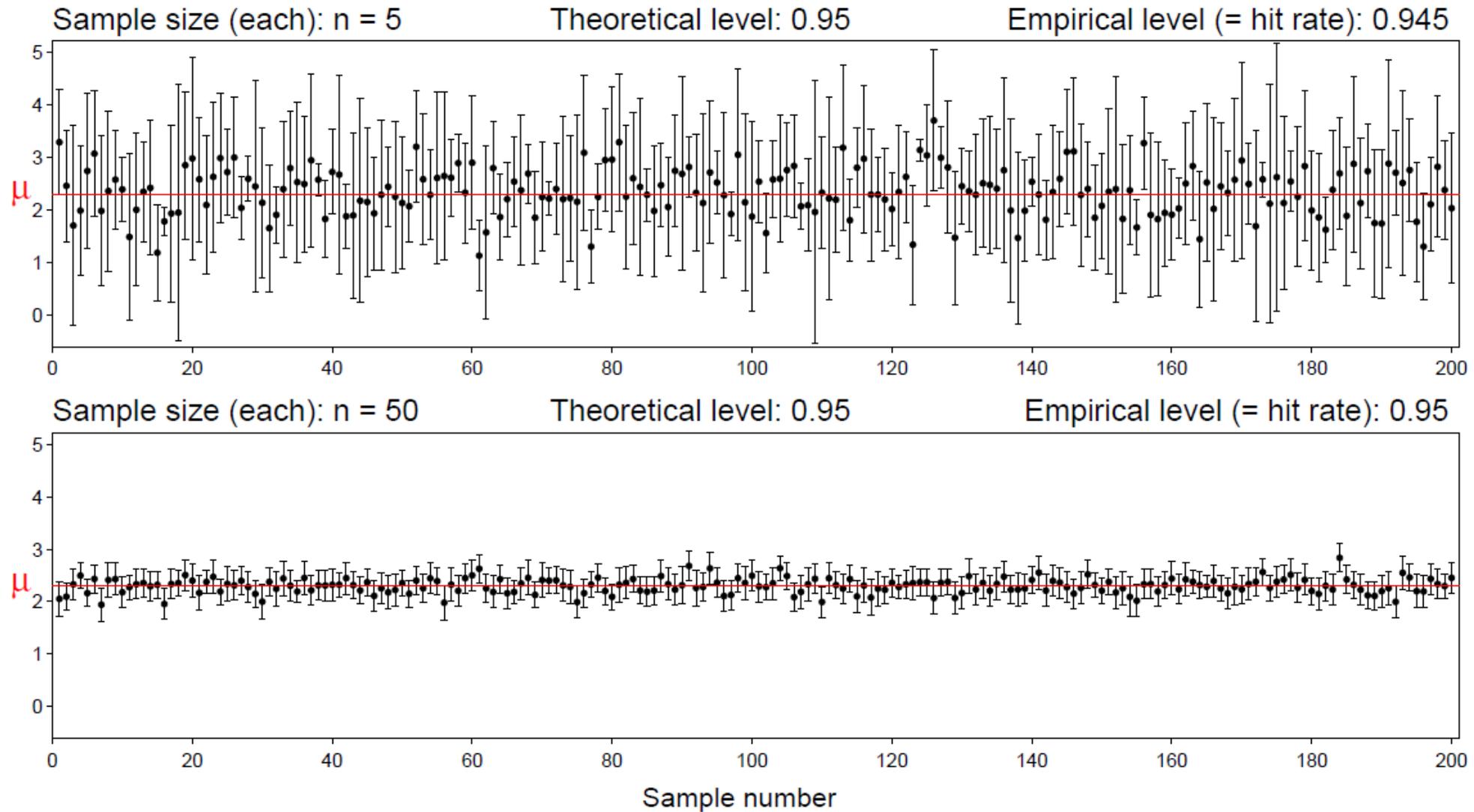
IC al 95%:

$$\left[\bar{X} - t_n * \frac{s}{\sqrt{n}} ; \bar{X} + t_n * \frac{s}{\sqrt{n}} \right]$$



William Gosset, detto "Student" (1876-1937)

«Hit rate» dell'IC di Student per 2 x 200 campioni simulati



VA Binomiale (RECAP)

Una VA X Binomiale «conta» il numero k dei successi di n VA di Bernoulli.

$$Y = X_1 + X_2 + \dots + X_n \qquad Y \approx \text{Bin}(n, p)$$

Es: Quale è la probabilità che in n lanci di una moneta esca k volte testa?

Due parametri descrivono la distribuzione di X :

- il numero delle prove n
- la probabilità di successo p



Conta tutti i possibili modi in cui possiamo estrarre k elementi da n

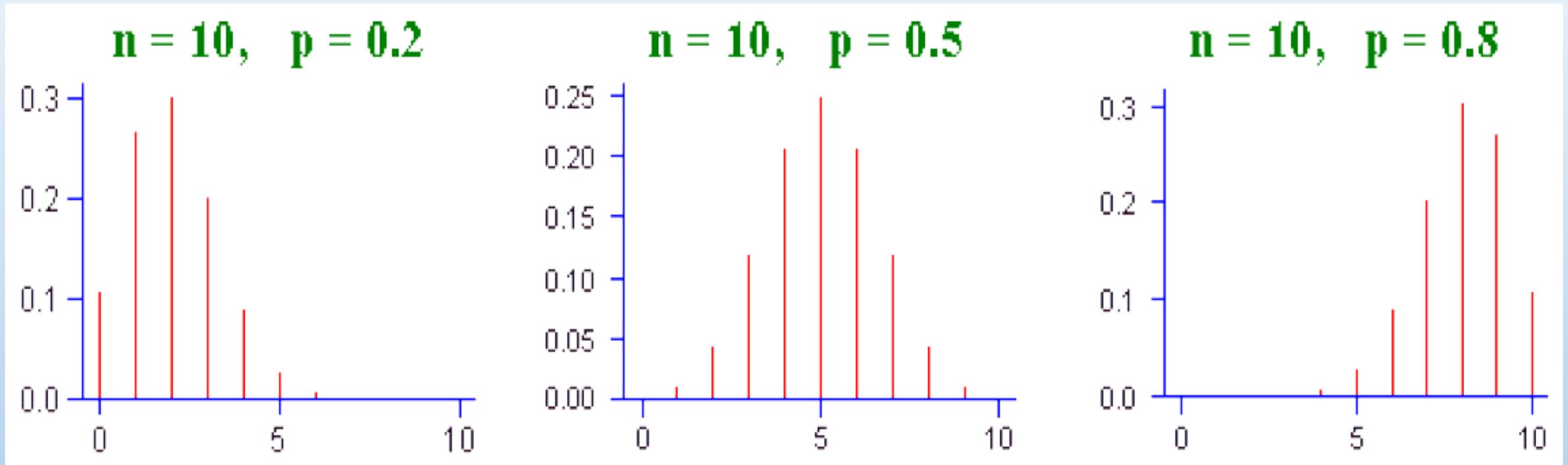
$$P(Y = k) = \frac{n!}{(n-k)!k!} p^k q^{n-k}$$



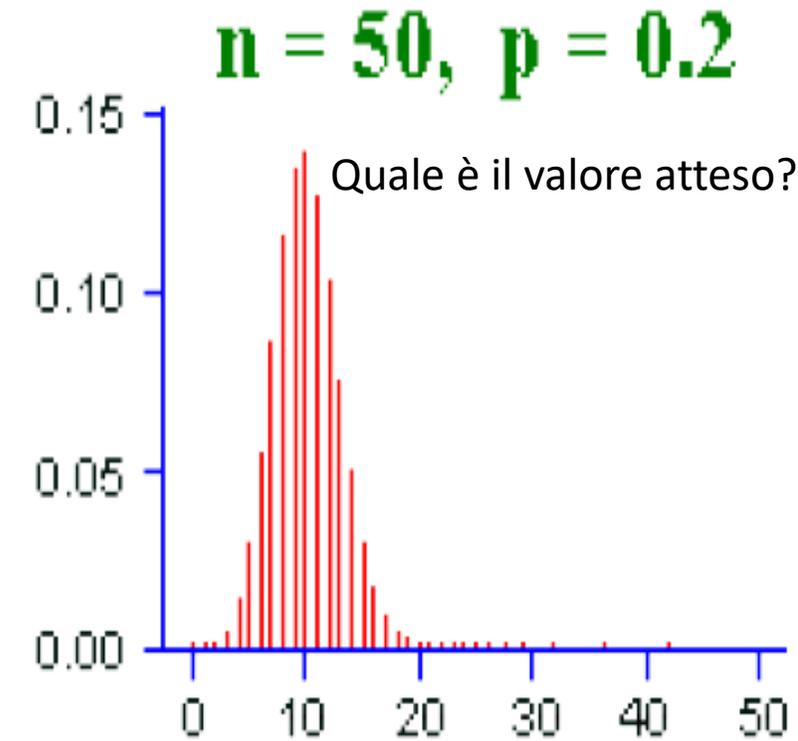
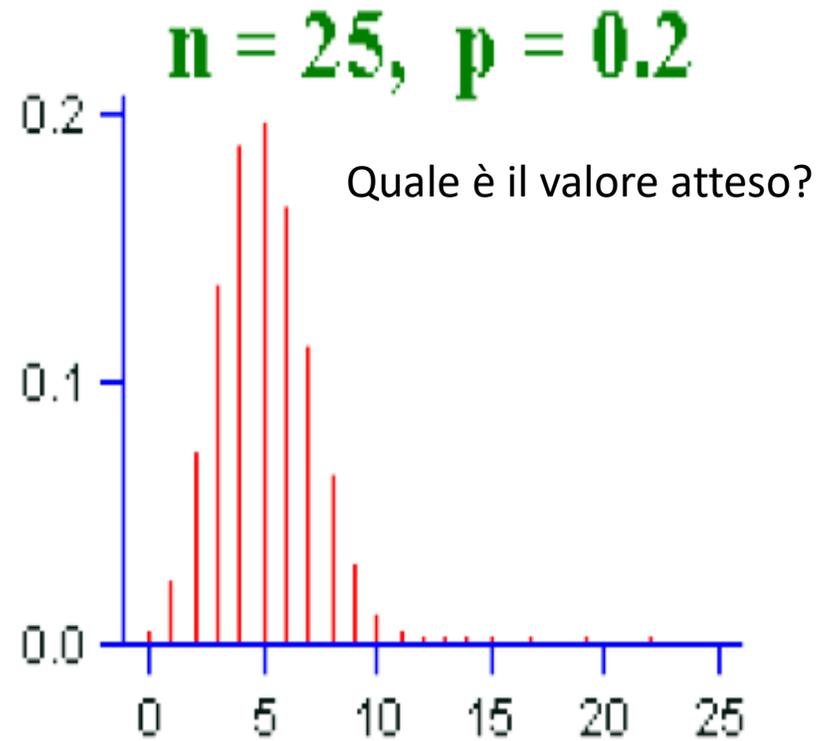
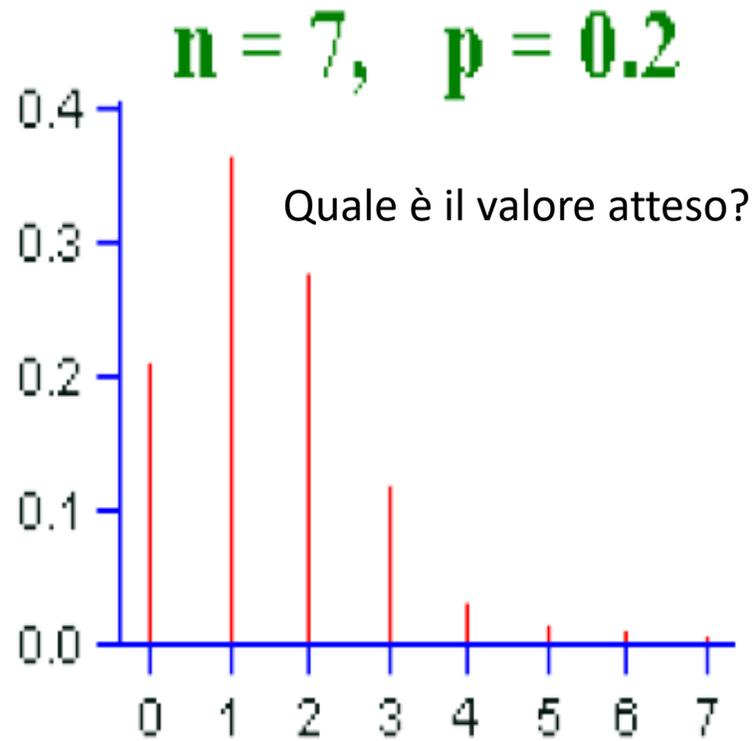
Coefficiente binomiale:

$$\rightarrow \binom{n}{k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{1*2*3\dots*k} = \frac{n!}{(n-k)!k!}$$

La “**forma**” della distribuzione di probabilità una VA binomiale dipende dai valori di n e p :



- Quando p è piccolo (0.2), la distribuzione binomiale è asimmetrica a destra
- Quando $p = 0.5$, la distribuzione binomiale è simmetrica
- Quando p è > 0.5 , la distribuzione binomiale è asimmetrica a sinistra...

Distribuzioni binomiali per diversi valori di n , dato $p=0.2$:

Al crescere di n , la distribuzione binomiale diventa sempre più simmetrica..

La distribuzione binomiale **può essere approssimata da una distribuzione normale** (per qualsiasi valore di p) al crescere di n^* .

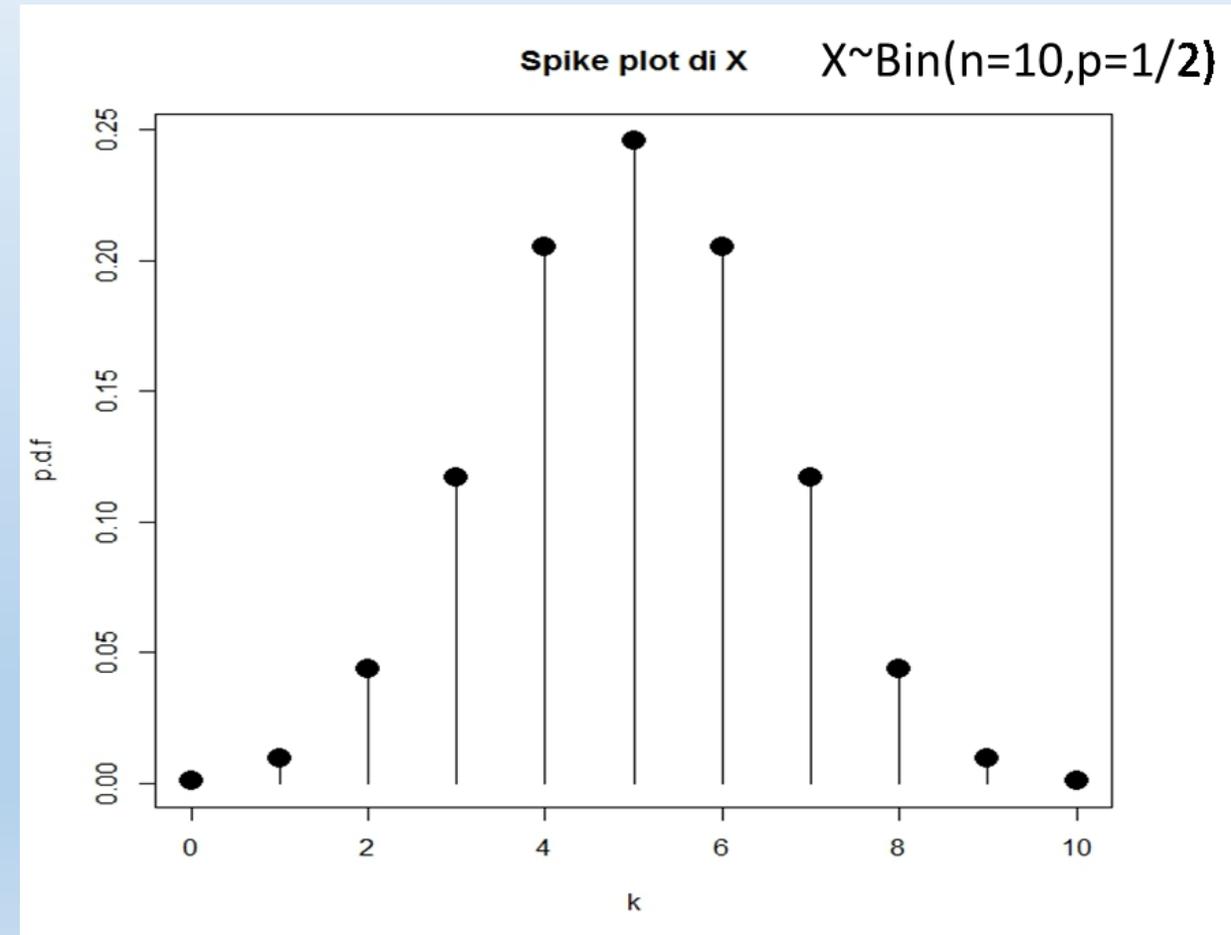
Questa approssimazione è accettabile data la condizione: $np \geq 5$ e $n(1 - p) \geq 5$

$$Y \approx \text{Bin}(n, p)$$

$$Z = \frac{Y - np}{\sqrt{np(1-p)}} \rightarrow N(0,1)$$

La media di una VA Binomiale è: $\mu = np$

La deviazione standard è: $\sigma = \sqrt{np(1-p)}$



Intervallo di confidenza per una proporzione

Stima di una proporzione nella popolazione: da un campione casuale di **VA** X_1, X_2, \dots, X_n che possono assumere solo due possibili valori 0 o 1 (sano o malato...) siamo interessati a stimare la proporzione di eventi (=malati) nella popolazione: $p = P(X_i=1)$.

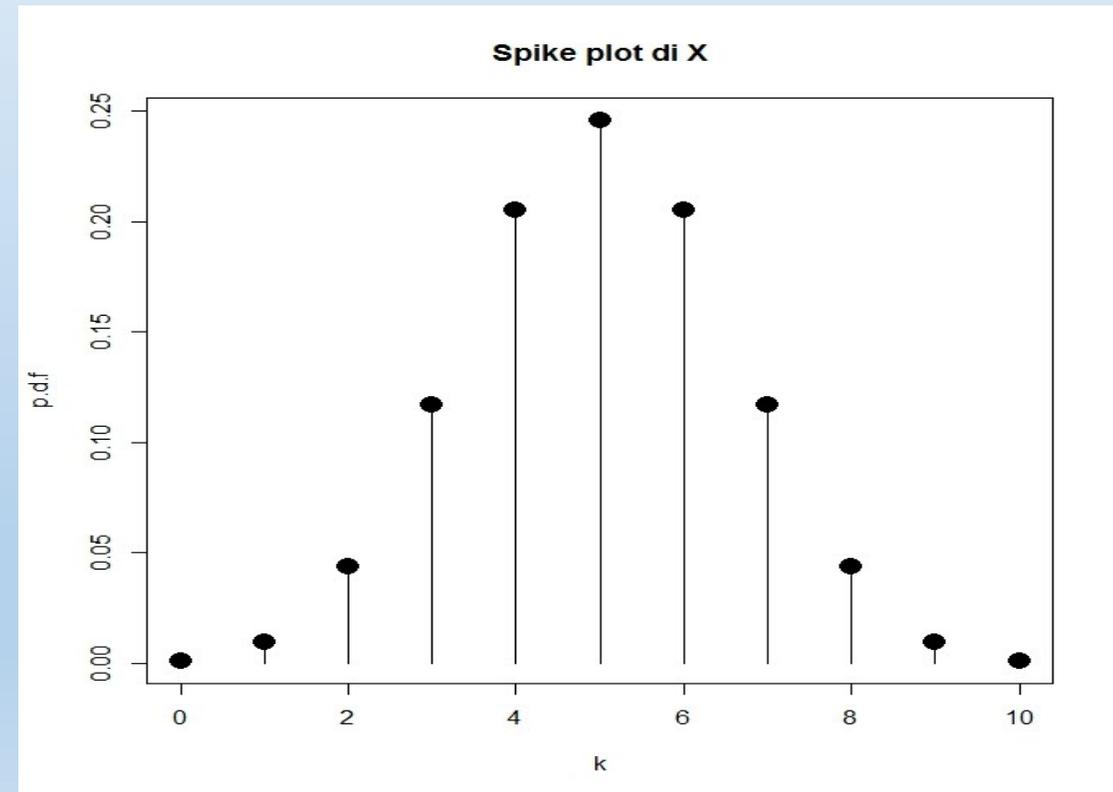
$$\bar{p} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{stimatore di } p$$

(ben approssimata dalla gaussiana per $np \geq 5$ e $n(1-p) \geq 5$)

$$\frac{\bar{p} - p}{SE(\bar{p})} \approx N(0,1)$$

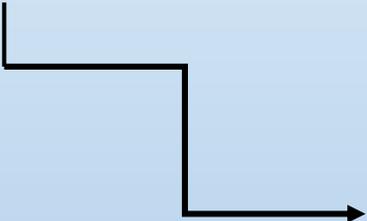
$$\bar{p} \pm 1.96 * SE(\bar{p})$$

$$SE(\bar{p}) = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad \text{errore standard di } \bar{p}$$



Relazione tra la numerosità campionaria e l'ampiezza dell'IC

L'ampiezza totale dell'intervallo di confidenza per la media di una popolazione è data da:

$$d = 2 * z_{\frac{\alpha}{2}} * \sqrt{\frac{\sigma^2}{n}}$$


Costante della gaussiana (es: 1,96)

E' facile osservare che, a parità del livello $1-\alpha$ scelto per l'intervallo di confidenza e della varianza nella popolazione, l'ampiezza dell'intervallo dipende dalla dimensione campionaria n , al crescere della quale l'ampiezza si riduce.

In molti casi applicativi, la dimensione campionaria n è fissata in partenza e dipende dal budget a disposizione per l'estrazione del campione.

In altri casi (ad esempio negli studi clinico-epidemiologici) è molto importante fissare l'ampiezza d^* che l'intervallo **non può superare** e determinare la dimensione campionaria minima n^* che garantisce tale requisito, cioè tale per cui quando $n < n^*$ si ottiene un intervallo con ampiezza $d > d^*$ (ovviamente per tutti gli $n > n^*$ si ottiene un intervallo con ampiezza $d < d^*$).

Per effettuare il calcolo di n^* è sufficiente osservare che se deve essere:

$$2 * z_{\frac{\alpha}{2}} * \sqrt{\frac{\sigma^2}{n}} \leq d^*$$

$$\sqrt{\frac{\sigma^2}{n}} \leq \frac{d^*}{2 * z_{\frac{\alpha}{2}}} \quad \Rightarrow \quad \frac{\sigma^2}{n} \leq \left(\frac{d^*}{2 * z_{\frac{\alpha}{2}}} \right)^2 \quad \Rightarrow \quad n \geq \left(\frac{2 * z_{\frac{\alpha}{2}} * \sigma}{d^*} \right)^2$$

Esempio: Da informazioni derivanti da una precedente analisi, si sa che la durata delle telefonate che arrivano al CUP di un ospedale si distribuisce in modo approssimativamente normale con media μ incognita e deviazione standard pari a 4 minuti.

Si desidera calcolare la dimensione campionaria minima necessaria per costruire un intervallo di confidenza per la stima della durata media delle chiamate al livello 95% che abbia un'ampiezza massima di 5 minuti.

La dimensione richiesta è data da:

$$n^* = \left(\frac{2 * z_{\frac{\alpha}{2}} * \sigma}{d^*} \right)^2 = \left(\frac{2 * z_{\frac{\alpha}{2}} * 4}{5} \right)^2 = 9,8 \quad \text{Bisogna avere un campione di almeno 10 telefonate}$$

Si osservi che la conoscenza di σ^2 è cruciale per la determinazione della dimensione campionaria ottimale. Tale conoscenza può derivare da studi precedenti o da studi pilota.

E' comunque consigliabile sovrastimare la varianza (piuttosto che sottostimarla) in quanto è meglio utilizzare una dimensione campionaria *troppo elevata* che una *troppo bassa*.

Il calcolo della dimensione campionaria ottimale può essere fatto anche quando l'intervallo di confidenza è calcolato per una proporzione incognita p .

Naturalmente, in questo caso la precisione dello stimatore (e quindi l'ampiezza dell'intervallo) dipende dal valore assunto da p , che è incognito.

In particolare, ci si accorge che la funzione $p*(1 - p)$ è una funzione concava che vale 0 quando $p= 0$ oppure 1 e raggiunge il suo massimo quando $p = 0.5$. Se ne deduce che a parità di dimensione campionaria e di livello di confidenza otterremo intervalli di confidenza generalmente più stretti quando p si trova vicino agli estremi 0 e 1, e più larghi quando p si trova in un intorno di 0.5.

E' dunque appropriato usare come misura cautelativa proprio $p=0.5$: $p*(1-p)=0.25$

Se dunque desideriamo calcolare la dimensione minima richiesta per avere un intervallo per p che non superi l'ampiezza globale massima d^* dobbiamo cercare il minimo valore di n tale che:

$$2 * z_{\frac{\alpha}{2}} * \sqrt{\frac{0.25}{n}} \leq d^*$$

Dopo qualche passaggio si ottiene:

$$n^* = 0.25 * \left(\frac{2 * z_{\frac{\alpha}{2}}}{d^*} \right)^2$$

Secondo tale formula, se ad esempio programmiamo uno studio epidemiologico per stimare la proporzione di soggetti con una certa malattia e desideriamo che l'intervallo di confidenza al livello $1 - \alpha = 0.95$ non superi l'ampiezza di 2 punti percentuali, avremo bisogno di un minimo di:

$$n^* = 0.25 * \left(\frac{2 * 1.96}{0.02} \right)^2 = 9604$$

Dovremo estrarre dalla popolazione target un campione di almeno 9604 persone.

Test di Ipotesi: concetti di base

- Si ha una ipotesi su un certo fenomeno nella popolazione che si vuole verificare (ipotesi nulla vs ipotesi alternativa)
- Si raccolgono dei dati attinenti al problema (dati campionari)
- Si combinano i pezzi di informazione per ottenere una «misura di evidenza» a favore o contro l'ipotesi
- Si decide se c'è abbastanza evidenza dai dati per accettare o rigettare l'ipotesi nulla

Test di Ipotesi: tentativo di analogia

Una persona viene accusata di un reato: viene arrestata e portata davanti ad un tribunale

- (a) Ipotesi nulla: Presunzione di innocenza; ipotesi alternativa: la colpevolezza dell'indagato
- (b) Si raccolgono informazioni (evidenze=dati) sulla questione
- (c) Il giudice valuta gli indizi raccolti
- (d) Il giudice decide se incolpare o meno l'indagato

Il principio fondamentale:

Evidenza non sufficiente -> Verdetto di non colpevolezza (in dubio pro reo)

Purtroppo: può succedere che un innocente vada in galera,
così come un colpevole sia lasciato libero...



Errori di I e di II tipo

Studio campionario  Possibilità di decisioni sbagliate

		VERITA' (Ignota)	
		H ₀ è vera	H ₀ è falsa
Decisione presa sui dati campionari	Rigetto H ₀	Errore di I tipo *	ok
	Non rigetto H ₀	ok	Errore di II tipo **

*Errore di I tipo:

- Rigettare H₀ quando in realtà è vera; (falso positivo = innocente in galera);
- Si associa una probabilità α di commettere questo errore: livello di significatività
- α è *sotto controllo*, perché il test è disegnato in modo tale che α non sia più grande di una soglia pre-specificata

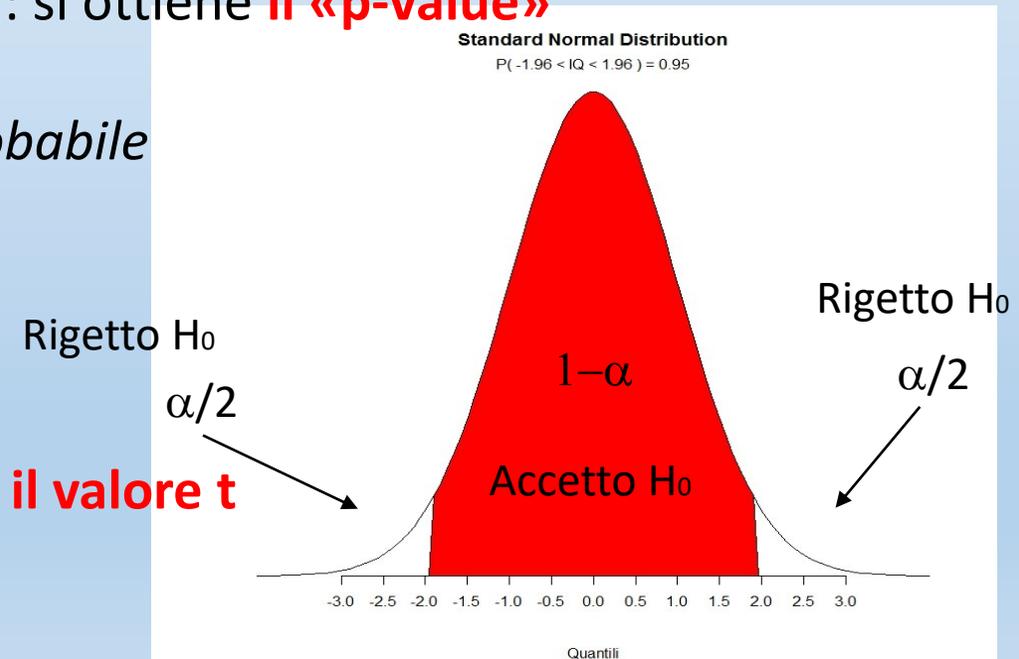
**Errore di II tipo:

- Non rigettare H₀ quando in realtà è falsa; (falso negativo = colpevole in libertà);
- Si associa una probabilità β di commettere questo errore: $1-\beta$ =Potenza del test
- β non è solitamente sotto controllo, perché la distribuzione della statistica di test è nota solo sotto l'ipotesi nulla...

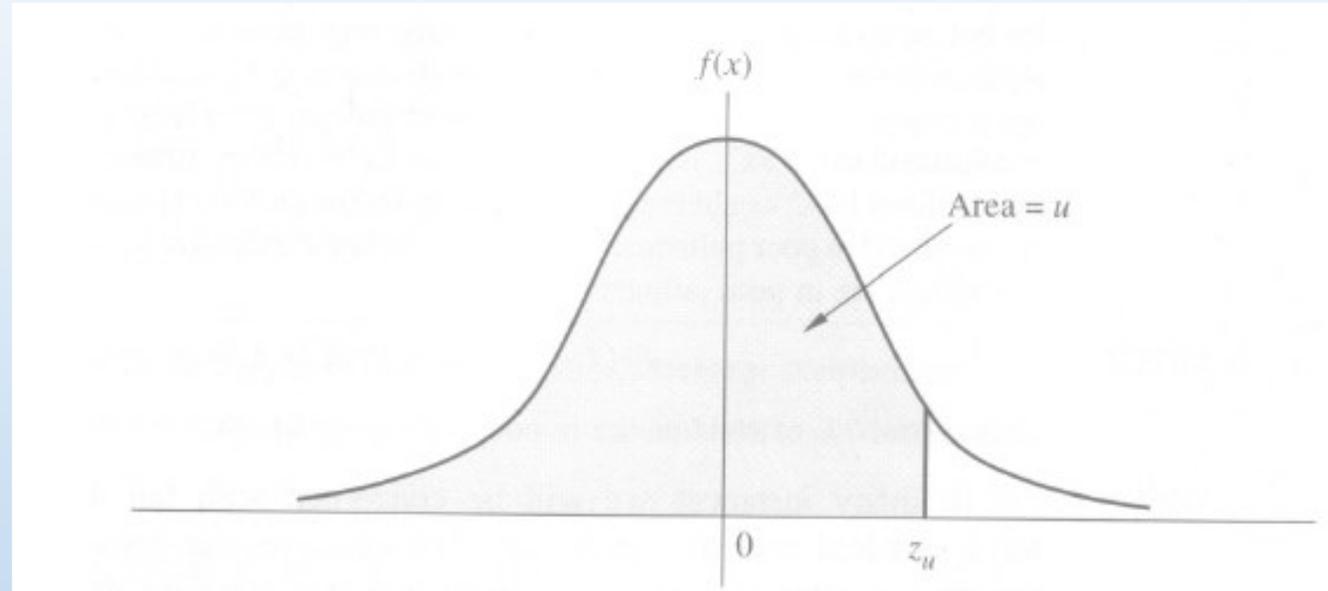
Effettuare un test statistico (strategia generale)

- (a) Ipotesi *nulla* **H₀** versus Ipotesi *alternativa* **H₁** (mutualmente esclusive)
- (b) Si disegna lo studio con le VA attinenti al problema: X₁, X₂, ... (es: glicemia/peso...)
- (c) Si calcola una *statistica di test* **T(x₁, x₂, ...)=t** sul campione, la cui distribuzione di probabilità è nota se vale H₀ (e differisce rispetto a quello che avrebbe sotto H₁) : si ottiene il **«p-value»**
- (d) Si «rigetta» H₀ se il valore osservato di **t** è troppo *poco probabile* (se H₀ fosse vera): **p ≤ α**

p-value = probabilità sotto H₀ che la variabile casuale T abbia il valore t osservato sui dati campionari o un valore più «estremo»



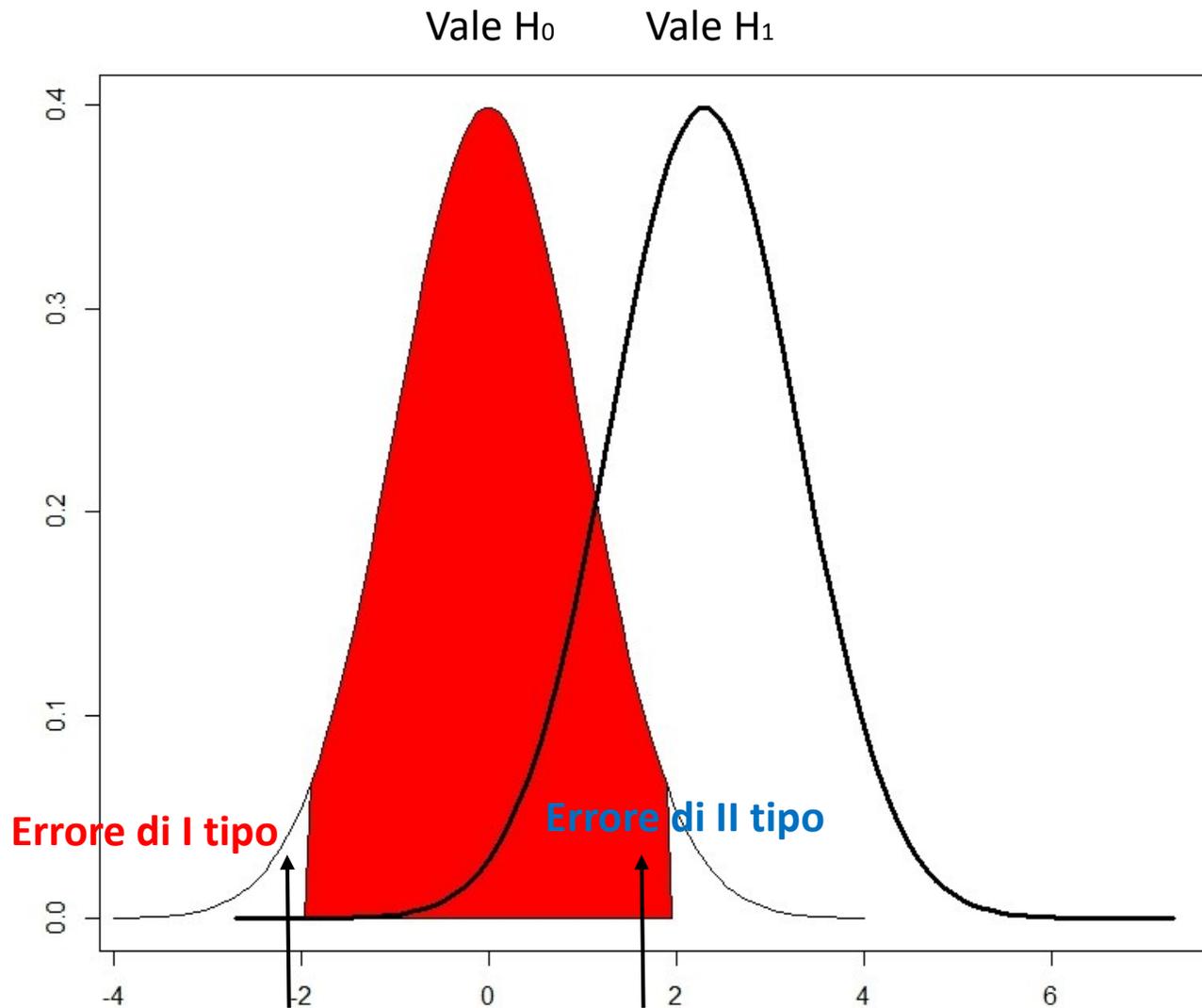
Quantili della normale standard



Il $(100-u)$ percentile di una normale standard è indicato con z_u ed è definito dalla relazione:

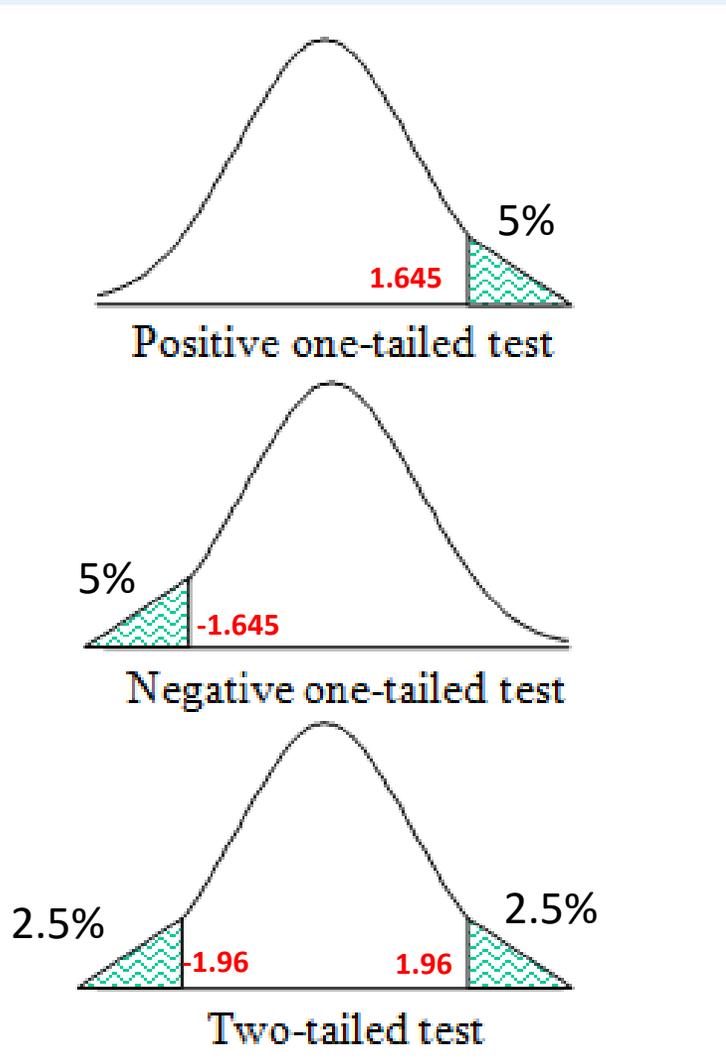
$$P(X < z_u) = u$$

$$z_{.975} = 1.96 \quad z_{.95} = 1.645 \quad z_{.5} = 0 \quad z_{.025} = -1.96$$



- Il calcolo della statistica di test e del *p-value* viene effettuato dal software;
- Nel riportare i risultati di un test vanno specificate le ipotesi (H₀, H₁) il nome del test utilizzato, il *p-value* e l'IC;
- Il valore di *p* quantifica la **plausibilità della ipotesi nulla** : quanto più è piccolo, tanto meno plausibile (verosimile) appare H₀

Test di ipotesi ad una coda oppure a due code?*



$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

Supponiamo di dover confrontare due farmaci «A» e «B».

Se si crede che il farmaco «A» sia meglio del farmaco «B», si farà il test ad una coda. In questo caso però si rischia di accettare l'ipotesi nulla di uguaglianza anche nel caso che «A» sia inferiore a «B»

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

Solo nel caso si consideri questo problema *trascurabile*, si potrà usare il test ad una coda...

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

*(se la statistica di test ha una distribuzione simmetrica)

Conclusioni & Conseguenze

Supponiamo che α sia piccolo, tipicamente $\alpha \leq 0.05$

- Se H_0 viene «rigettata» questa decisione **è affidabile**:

la probabilità di errore sempre nota a priori α è piccola («*il risultato del test è statisticamente significativo*»);

- Se H_0 non è rigettata, si conclude che i dati non offrono sufficiente evidenza per *rigettare H_0* ; questa decisione **può non essere così affidabile**:
la probabilità di errore β non è (generalmente) fissata a priori e potrebbe essere grande («*il risultato del test non è statisticamente significativo*»);

“absence of evidence is not evidence of absence”



*** una ipotesi H_0 può non essere rigettata (quando invece dovrebbe) perché la dimensione campionaria è troppo piccola!!!**

Trasformazioni delle scale di misura (in relazione ai test)

La trasformazione di una variabile (su scala quantitativa) può cambiare la sua distribuzione da asimmetrica a una distribuzione più simmetrica, più *simile* alla distribuzione normale.

Questo passaggio è utile per vari scopi: molti metodi statistici (inferenziali) assumono la **normalità** delle variabili in esame.

Ad esempio, se siamo interessati a fare un test statistico sulla differenza tra due medie, la media deve essere un valore rappresentativo della distribuzione del fenomeno che stiamo studiando.

Inoltre, se stiamo studiando l'**associazione** tra una coppia di variabili, opportune trasformazioni della loro scala migliorano la *linearità* della associazione.

Bisogna però poi fare attenzione a come vengono riportati i risultati delle analisi con le variabili trasformate.

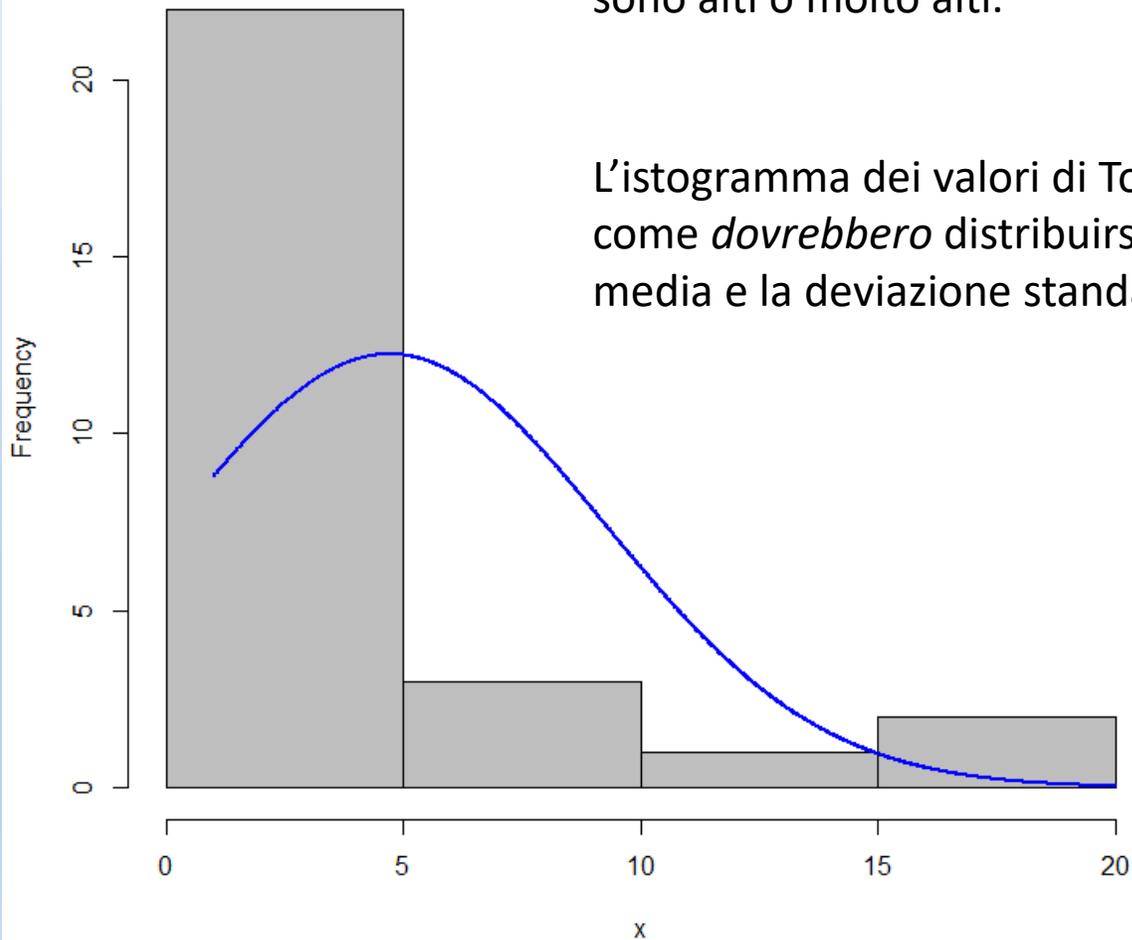
Per presentare i valori degli indici di posizione o dispersione, è a volte consigliabile *ri-trasformare* sulla scala originale per evitare ambiguità nell'interpretazione.

Questo esempio utilizza dati ipotetici della torbidità (x) dell'acqua del fiume.

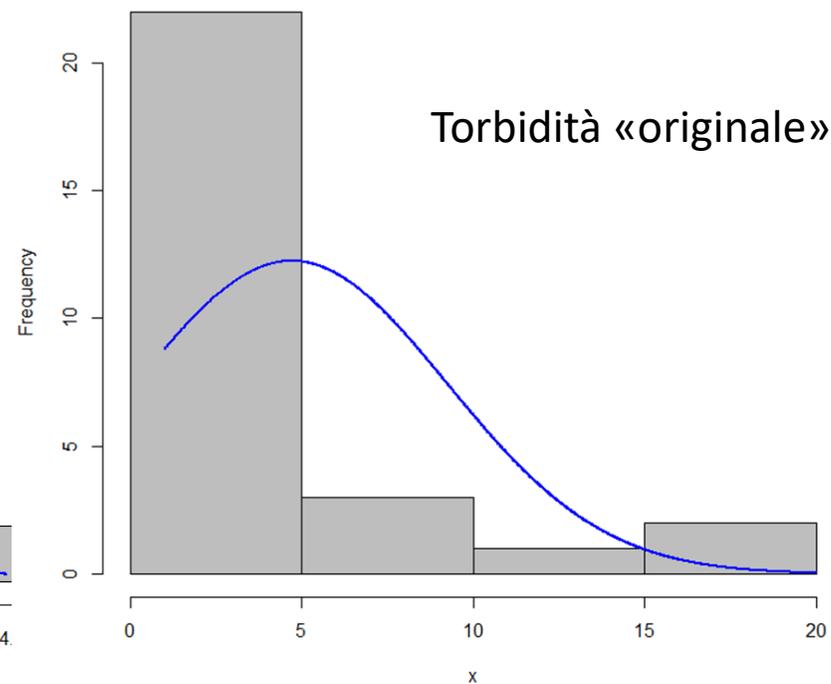
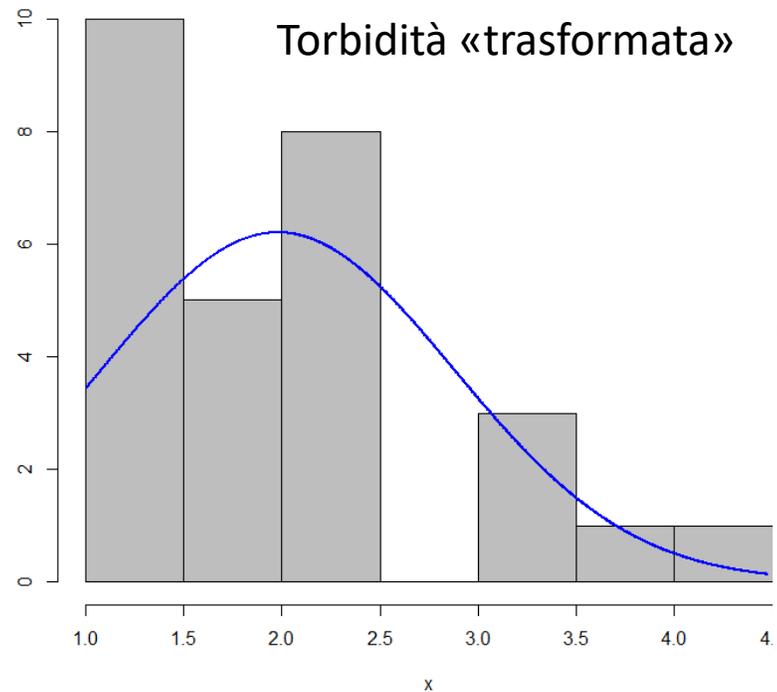
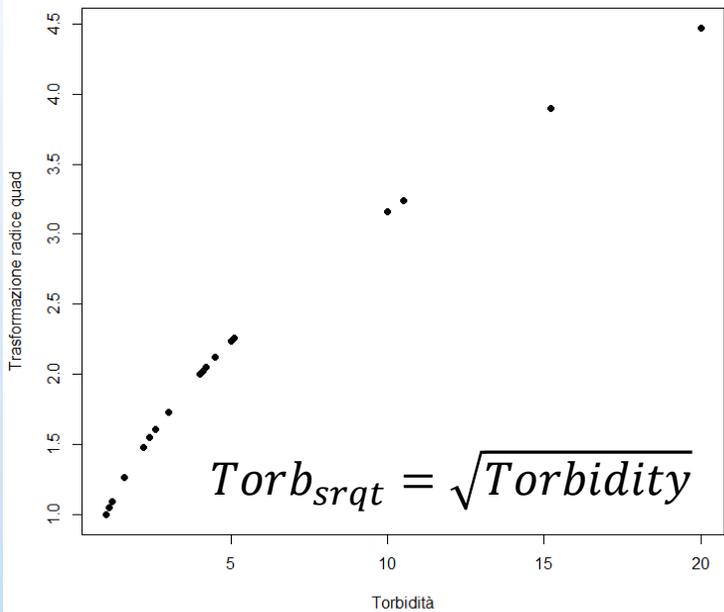
La torbidità è una misura di quanto l'acqua possa diventare torbida a causa del materiale sospeso (inquinanti per lo più).

La caratteristica di questa distribuzione è che i valori sono per lo più bassi, ma occasionalmente sono alti o molto alti.

L'istogramma dei valori di Torbidità, con una **curva normale (blu)** sovrapposta [ovvero come *dovrebbero* distribuirsi i dati, se provenissero da una distribuzione normale, data la media e la deviazione standard osservate] ci mostra la forte asimmetria a destra.



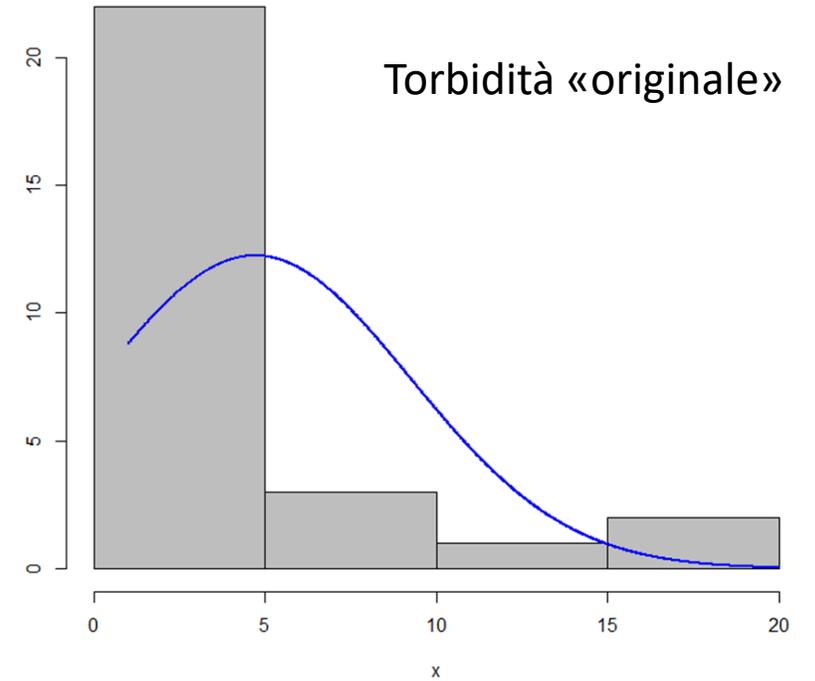
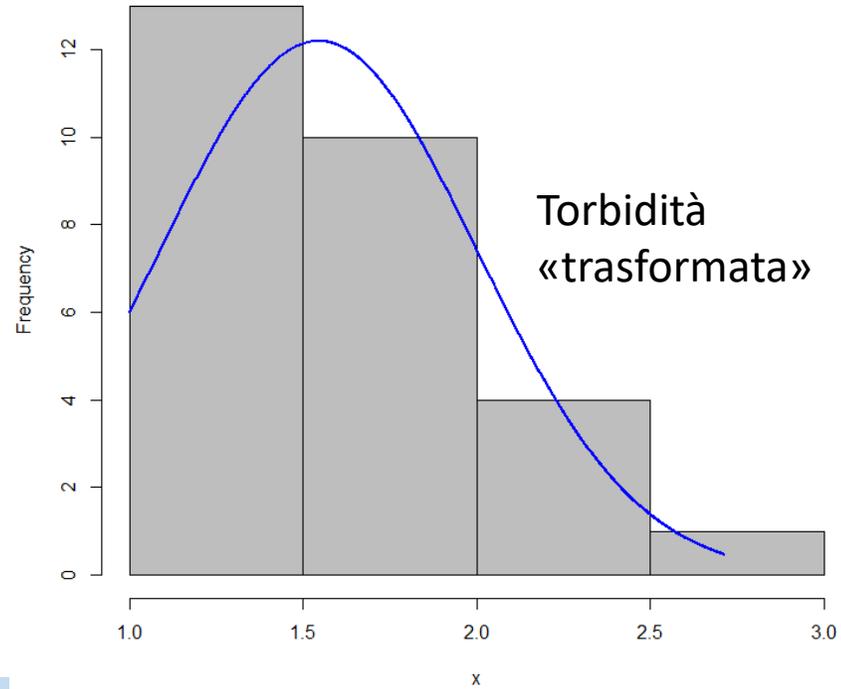
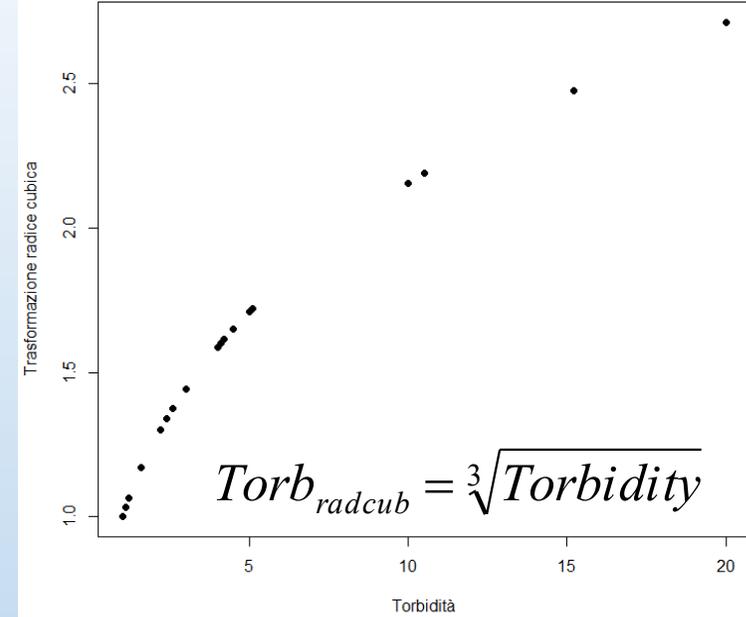
Proviamo adesso ad applicare alcune delle trasformazioni più comuni per i dati asimmetrici: la radice quadrata, la radice cubica ed il logaritmo.



C'è un lieve miglioramento,
ma non soddisfacente

	Media	Dev Std	Mediana	Q1	Q3	Min	Max	Asimmetria	Curtosi
Turbidity	4.7	4.56	4	1.5	5	1	20	1.81	2.87
Sqrt(Turbidity)	1.98	0.9	2	1.22	2.24	1	4.47	1.06	0.49

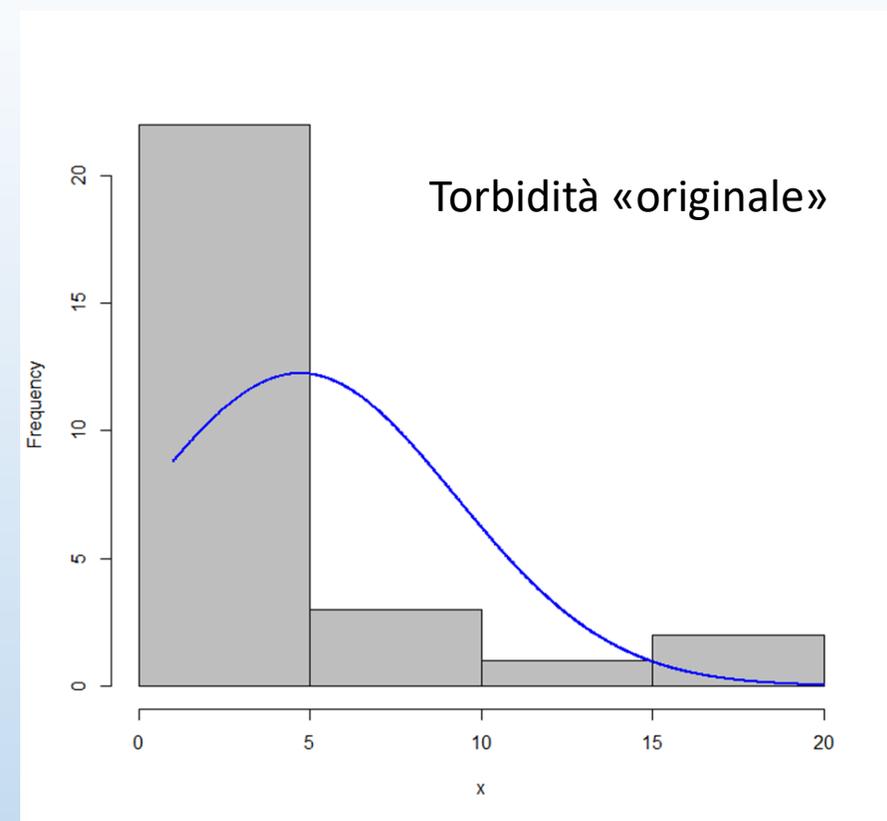
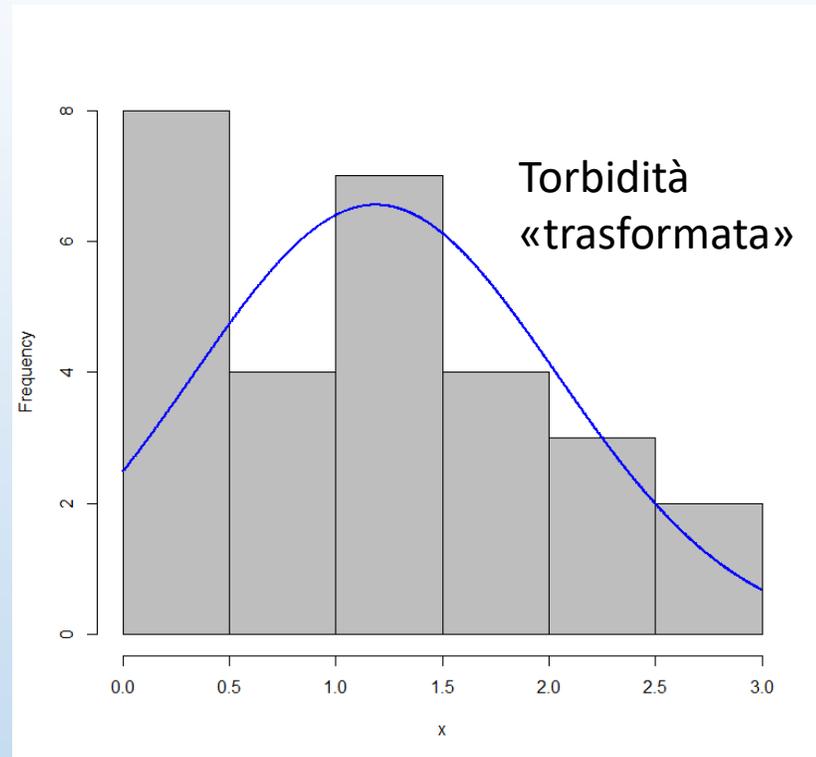
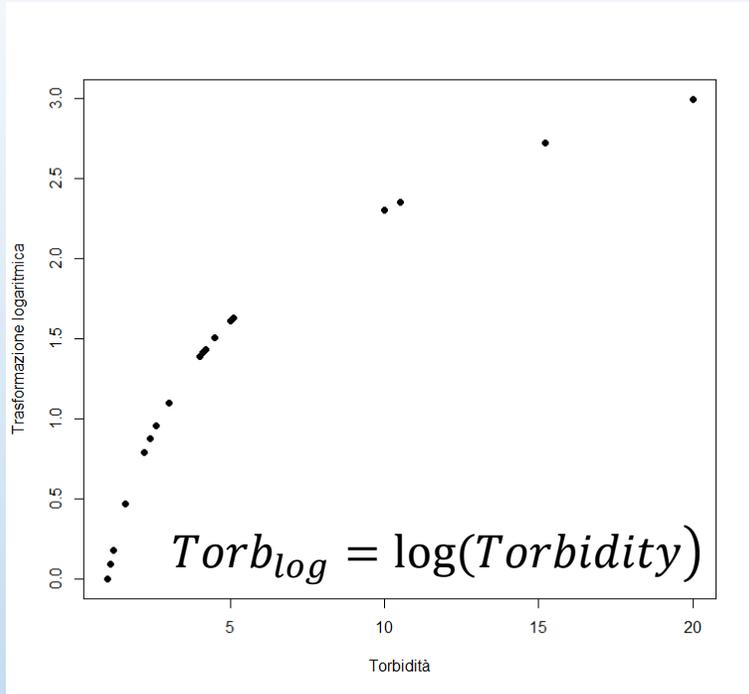
Può essere calcolata su dati ≥ 0 .



C'è un discreto miglioramento,
ma ancora non del tutto soddisfacente

	Media	Dev Std	Mediana	Q1	Q3	Min	Max	Asimmetria	Curtosi
Turbidity	4.7	4.56	4	1.5	5	1	20	1.81	2.87
Radcub(Turbidity)	1.54	0.46	1.59	1.14	1.71	1	2.71	0.79	-0.10

Può essere calcolata anche su dati negativi e che contengono il valore zero.

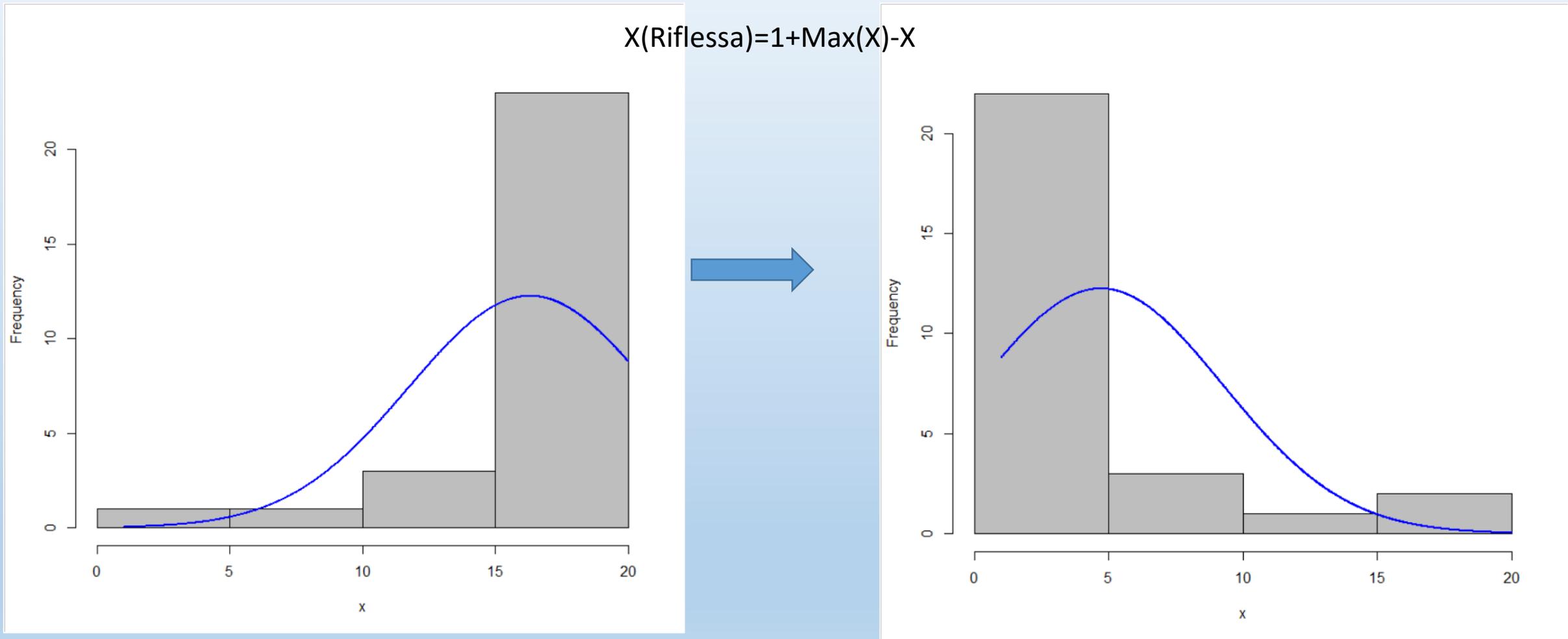


Decisamente meglio: visto il tipo di dati è probabilmente il risultato migliore che possiamo ottenere.

	Media	Dev Std	Mediana	Q1	Q3	Min	Max	Asimmetria	Curtosi
Turbidity	4.7	4.56	4	1.5	5	1	20	1.81	2.87
log(Turbidity)	1.19	0.85	1.39	0.4	1.61	0	3	0.28	-0.87

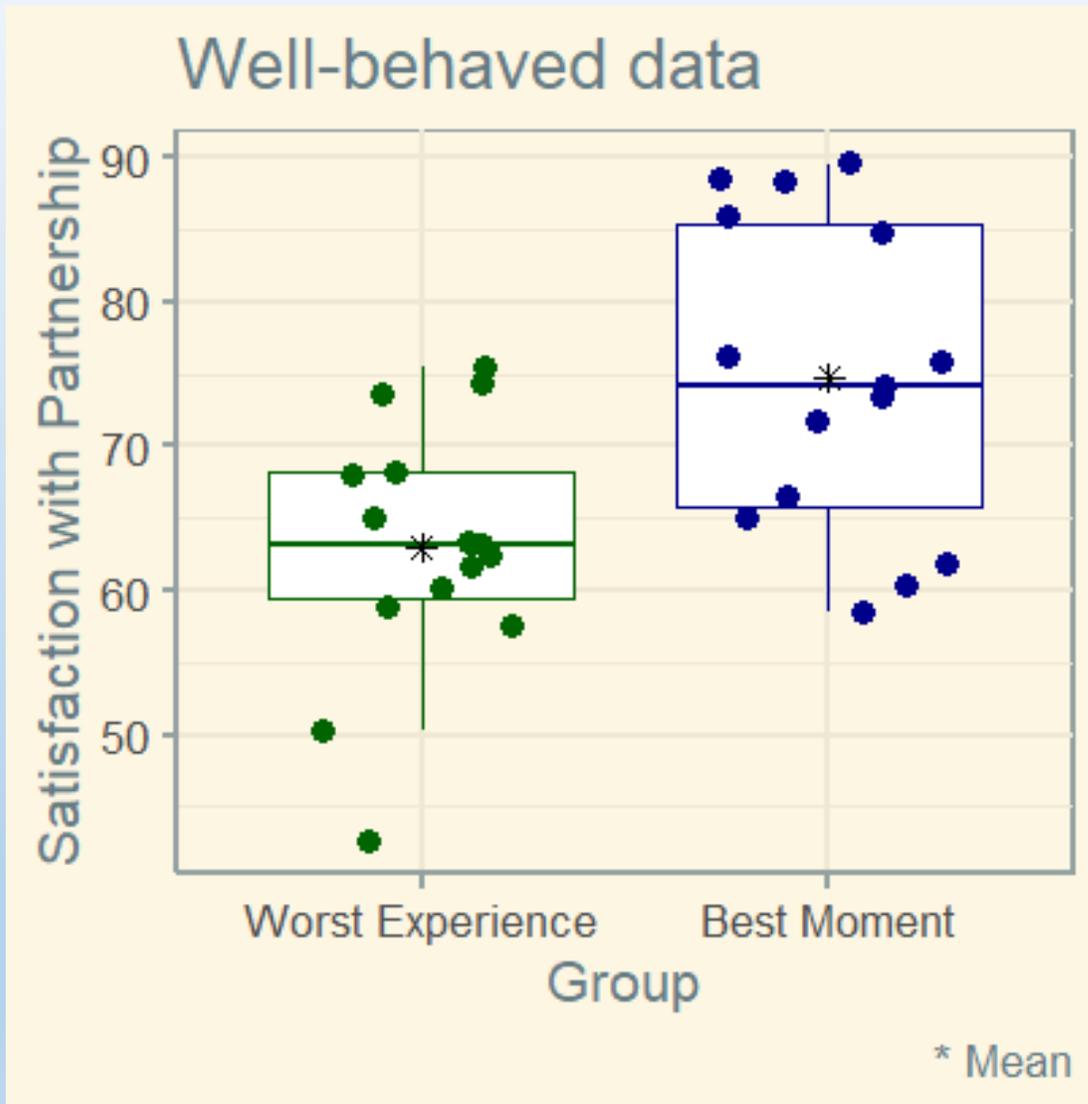
Logaritmo naturale: base è il numero di Nepero (circa 2.7); solo per numeri >0, se la distribuzione contiene 0 va aggiunta una costante

Teniamo presente che per le distribuzioni asimmetriche a sinistra, è opportuno prima di tutto creare la variabile «riflessa» e poi si possono applicare le trasformazioni viste (radice quadrata, cubica o logaritmo).



Ci poi sono alcune funzioni che possiamo applicare sui dati per ottenere «*la migliore trasformazione possibile*», ma non entriamo in questo dettaglio tecnico.

Impatto della asimmetria/outliers sugli esiti del test



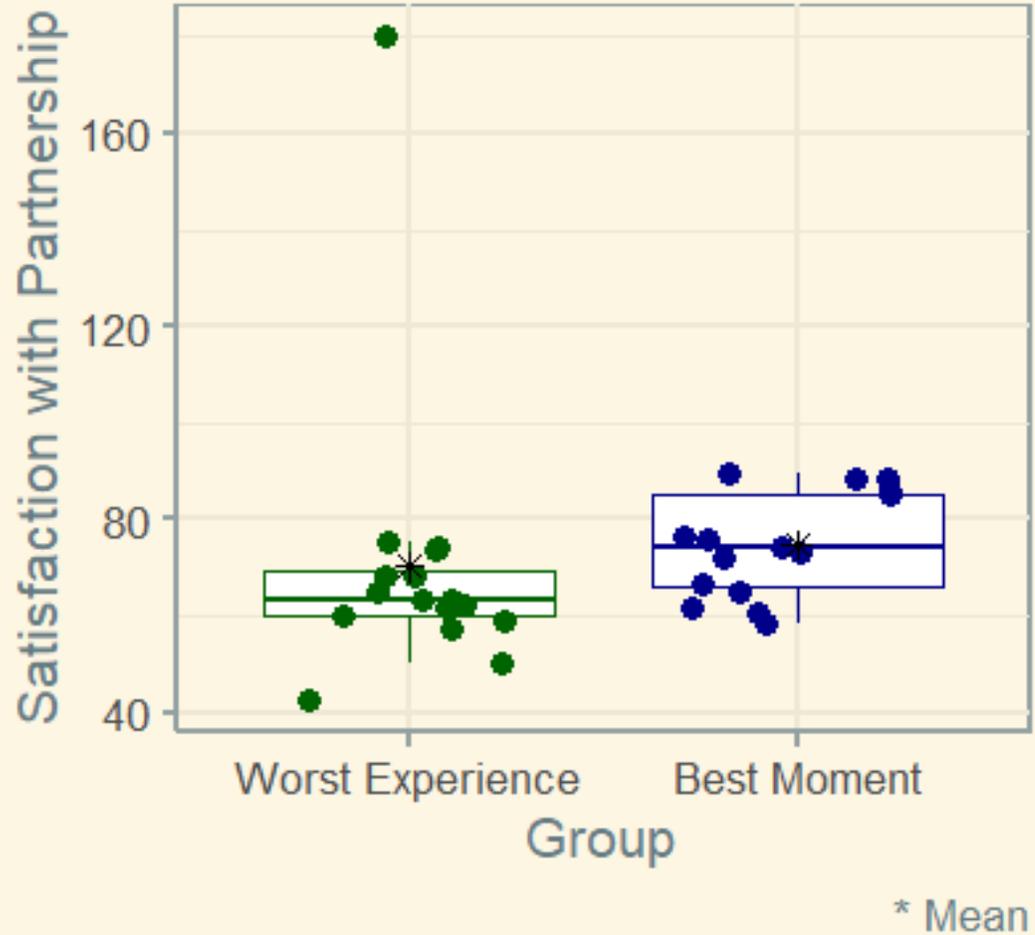
Characteristic	Worst Experience, N = 15 ¹	Best Moment, N = 15 ¹	p-value ²
satisfaction	63 (9)	75 (11)	0.003

¹ Mean (SD)

² Welch Two Sample t-test

Il t-test (opzione varianze non omogenee) trova una differenza significativa tra le medie di questi due gruppi.

Not well-behaved data



	Worst	Best	
	Experience, N	Moment, N	p-value²
Characteristic	= 16 ¹	= 15 ¹	
satisfaction	70 (30)	75 (11)	0.6

¹ Mean (SD)
² Welch Two Sample t-test

Il t-test (opzione varianze non omogenee) **non trova** una differenza significativa tra le medie di questi due gruppi

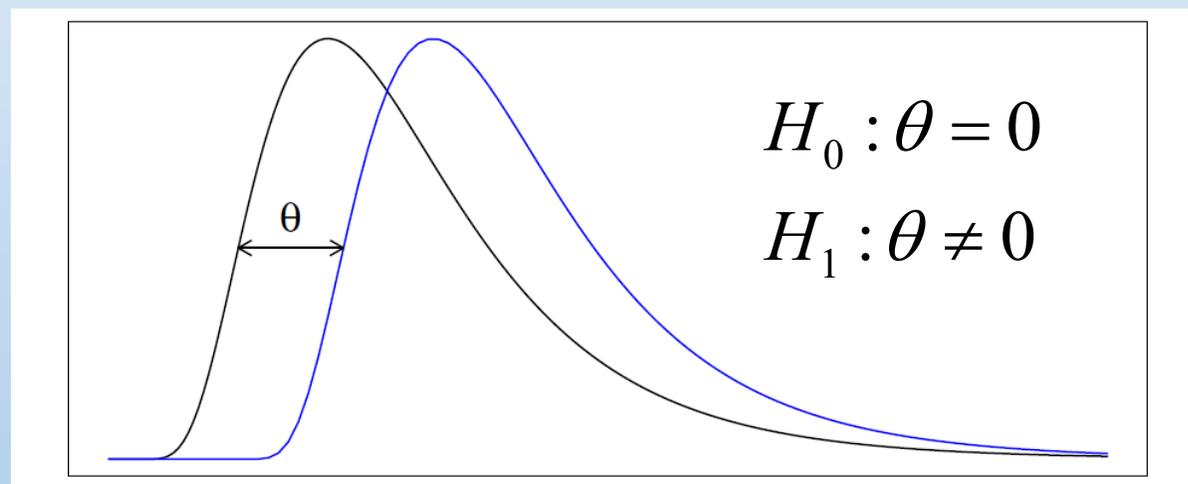
Come risolvere [se non è possibile trasformare]?

Test di ipotesi per la differenza tra due distribuzioni non gaussiane

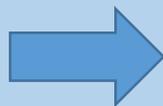
Quando le distribuzioni non sono «gaussiane» la media può non essere più rappresentativa...



Il test di significatività può essere fatto sui «ranghi»* della distribuzione:

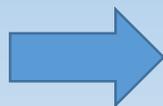


• Se le VA sono indipendenti



Wilcoxon-Mann-Whitney test (Wilcoxon rank sum test)

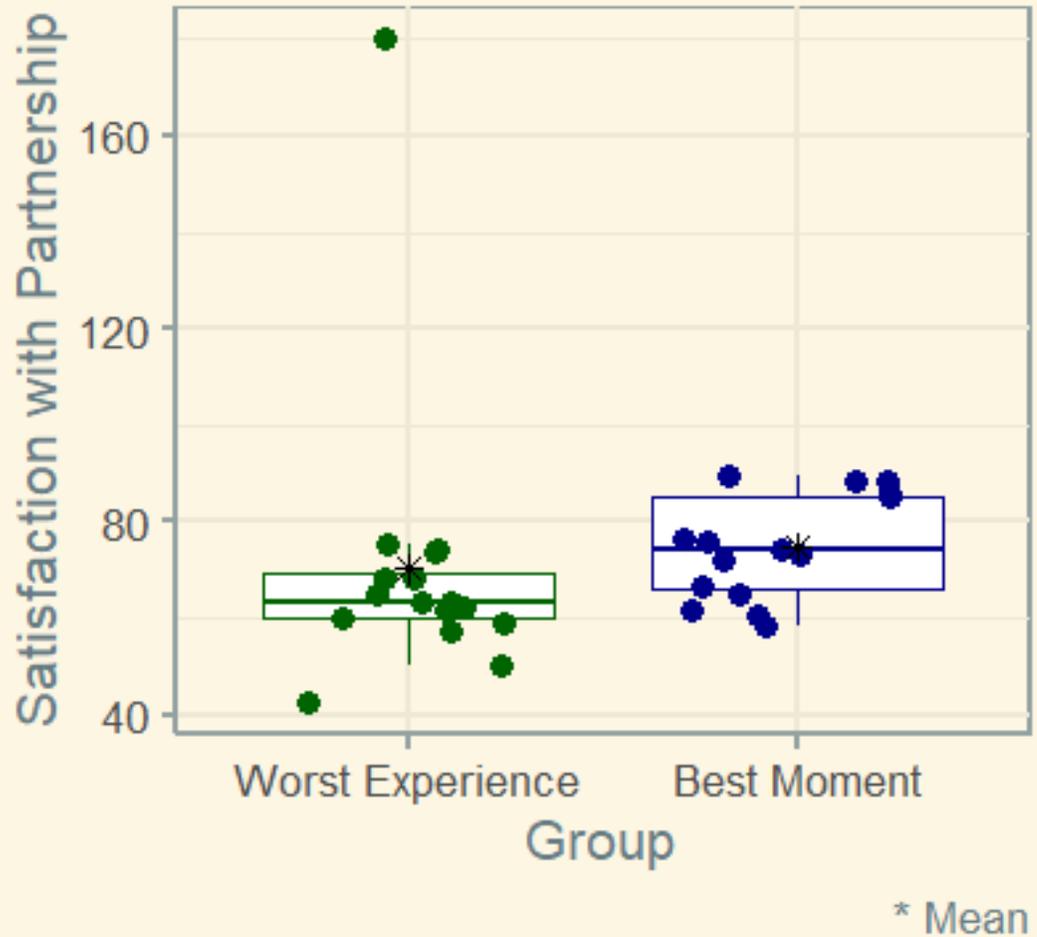
• Se le VA non sono indipendenti



Wilcoxon test per dati accoppiati (Wilcoxon signed rank test)

*posizioni delle osservazioni nell'ordinamento

Not well-behaved data



Characteristic	Worst Experience, N = 16 ¹	Best Moment, N = 15 ¹	p-value ²
satisfaction	63 (60, 70)	74 (66, 85)	0.027

¹ Median (IQR)
² Wilcoxon rank sum exact test

Il test non parametrico **trova correttamente** una differenza significativa nelle distribuzioni di questi due gruppi.

Number Needed to Treat (NNT)

In uno studio clinico con una risposta binaria, come vivo o morto, ci sono molti modi per quantificare la differenza tra due trattamenti.

Ad esempio, è possibile utilizzare la differenza tra due proporzioni: $p_{new} - p_{old}$

p_{new} = proporzione di successi con il nuovo trattamento ; p_{old} = proporzione di successi con il vecchio trattamento

$$NNT = \frac{1}{p_{new} - p_{old}}$$

NNT è il numero di pazienti da sottoporre al nuovo trattamento per ottenere un ulteriore successo rispetto al vecchio trattamento.

$$1 \leq NNT < \infty \left\{ \begin{array}{l} \frac{1}{1-0} \text{ Il nuovo trattamento è sempre efficace mentre il vecchio non lo è mai} \\ \frac{1}{0} \text{ Il nuovo trattamento non è in alcun modo efficace (la differenza è pari a zero)} \end{array} \right.$$

Potremmo addirittura avere la situazione opposta, cioè che il nuovo trattamento sia dannoso, e quindi ottenere un valore negativo:

$NNT < 0 \rightarrow NNH =$ number needed to harm (numero necessario per avere un danno)

Esempio:

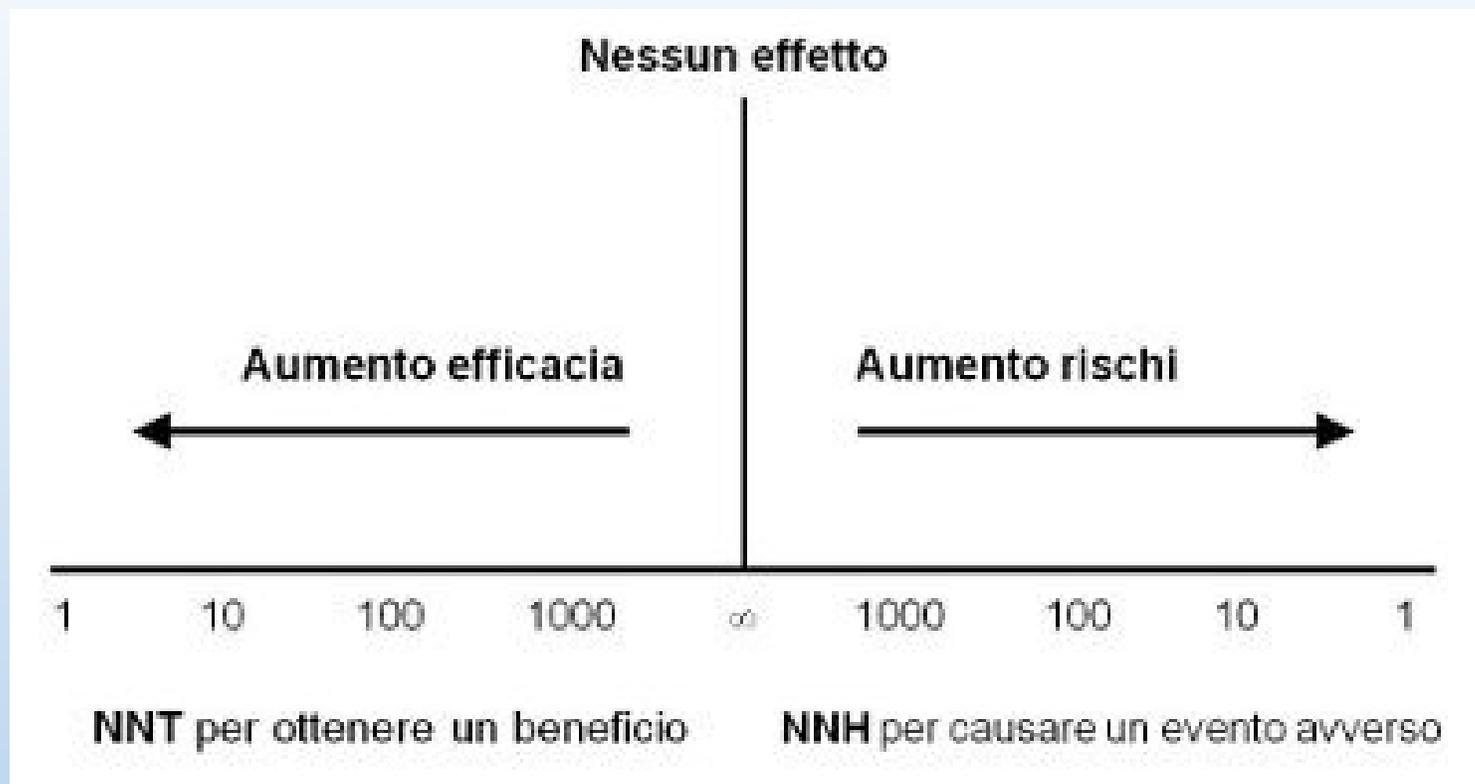
In uno studio sulla streptomina, nel quale i soggetti erano selezionati tra i pazienti affetti da tubercolosi polmonare, la proporzione di sopravvissuti a 6 mesi era 93% nel gruppo trattato vs 73% nei controlli.

$$NNT = \frac{1}{p_{new} - p_{old}} = \frac{1}{0.93 - 0.73} = \frac{1}{0.20} = 5$$

Quindi il numero necessario di pazienti da trattare per prevenire un decesso a 6 mesi era pari a 5.

E' possibile calcolare anche gli intervalli di confidenza attorno a NNT, ma non entriamo in questo dettaglio*.

*Sorgono dei problemi qualora la differenza tra le proporzioni non sia statisticamente significativa, perché l'intervallo di confidenza in quel caso contiene lo zero, in tal caso infinito è un possibile valore per NNT...così come sono ammissibili valori negativi.



Al **diminuire** del NNT aumenta l'efficacia del trattamento, per cui 1 è il NNT ideale: un successo terapeutico per ciascun paziente trattato (misura di **efficacia**).

All'**umentare** del NNH si riduce la probabilità di eventi avversi e aumenta la sicurezza del trattamento, per cui il NNH ideale tende all'infinito, documentando l'assenza di eventi avversi (misura di **sicurezza**).

Test di ipotesi per dati di tipo categorico

- I dati di tipo “categorico” possono essere descritti tramite una **tabella di contingenza**
- Il test chi-quadrato confronta i valori osservati in ogni cella della tabella rispetto a quelli che ci saremmo attesi **se non ci fosse associazione** tra la variabile sulle righe e la variabile sulle colonne
- L'ipotesi nulla è quindi: **non c'è associazione** tra i due fenomeni

Criterion 2	Criterion 1					Total
	1	2	3	...	C	
1	n_{11}	n_{12}	n_{13}	...	n_{1c}	r_1
2	n_{21}	n_{22}	n_{23}	...	n_{2c}	r_2
3	n_{31}				⋮	
⋮	⋮				⋮	
⋮	⋮				⋮	
r	n_{r1}	n_{rc}	r_r
Total	C_1	C_2			C_c	n

La relazione tra una malattia e un fattore di esposizione può essere descritta tramite una tabella di contingenza (valori osservati= O_i):

Disease			
Exposure	Yes	No	Total
Yes	37	13	50
No	17	53	70
Total	54	66	120

$37/54 = 68\%$ di **individui malati** era esposta

$13/66 = 20\%$ di **individui non malati** era esposta

Questi dati suggeriscono una associazione tra malattia ed esposizione?

Sotto l'ipotesi nulla di assenza di associazione, i valori attesi (E_i) nelle celle sarebbero:

Disease			
Exposure	Yes	No	Total
Yes	$50/120 \times 54 = 22.5$	$50/120 \times 66 = 27.5$	50
No	$70/120 \times 54 = 31.5$	$70/120 \times 66 = 38.5$	70
Total	54	66	120

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Differenze valori osservati vs attesi

Chi quadro= 29.1

La probabilità di ottenere questo valore sotto l'ipotesi nulla è < 0.001

- Il chi-quadro è un test di associazione tra due variabili di tipo categorico [o continue ma raggruppate in classi];
- Si può applicare per studiare la associazione tra una malattia ed un fattore di esposizione in uno studio di coorte o in uno studio caso-controllo non appaiato o in uno studio trasversale (cross-sectional study);
- **Il test di McNemar è una variante del Chi-quadro per dati appaiati** e può essere utilizzato negli studi caso-controllo appaiati

Disease			
Factor	Present (D)	Absent (\bar{D})	Total
Present (F)	a	b	a+b
Absent (\bar{F})	c	d	c+d
Total	a+c	b+d	N

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Rischio Relativo

Odds ratio $OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}$

Facciamo il punto:

- Per (quasi) ogni tipo di ipotesi da sottoporre a verifica si può trovare il test opportuno...(o lo si può creare);
- Il test statistico è soggetto ad errore, essendo generato da un meccanismo basato sulla probabilità;
- L' errore può essere «minimizzato» **ma non eliminato**;
- La significatività statistica **non sempre coincide** con la rilevanza clinica;
- E' opportuno consultare lo statistico **nella fase di disegno dello studio** per pianificare le analisi da condurre...ex-post è sempre tardi!