ASSISTENZA SANITARIA & TECNICHE DELLA PREVENZIONE NELL'AMBIENTE E NEI LUOGHI DI LAVORO

STATISTICA MEDICA

gbarbati@units.it

A.A. 2024-25

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Richiamo: VA Binomiale

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

$$Y = X_1 + X_2 + \ldots + X_n \qquad\qquad Y \approx Bin(n, p)$$

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***

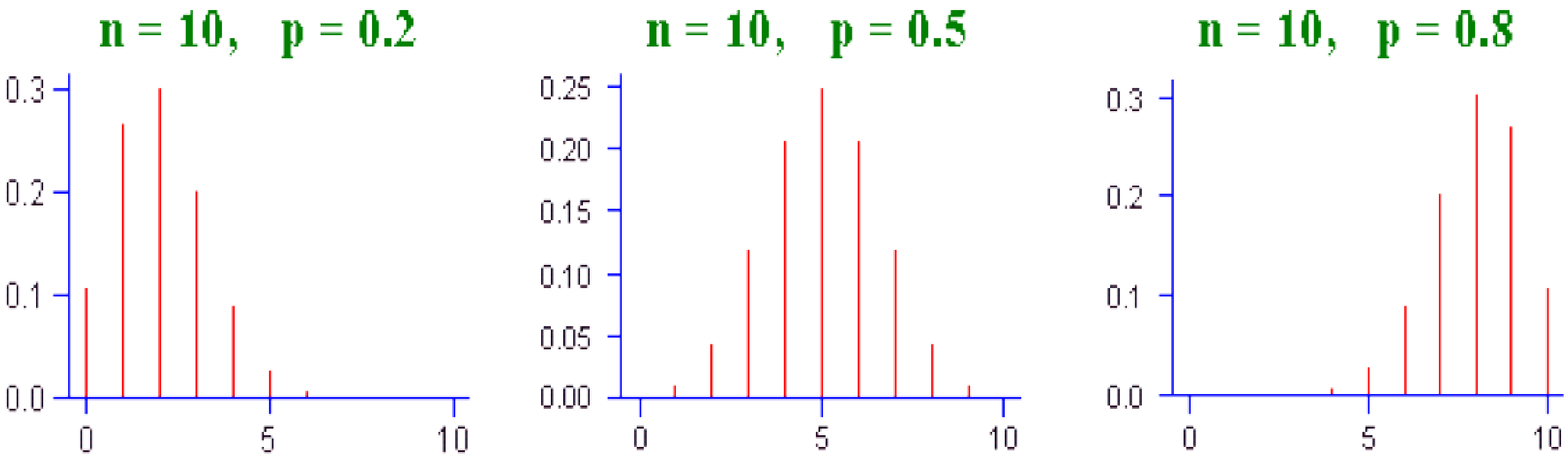
$$P(Y = k) = \frac{n!}{(n-k)!k!} p^k q^{n-k}$$

Coefficiente binomiale:

Conta tutti i possibili modi in cui possiamo estrarre k elementi da n $\longrightarrow$ $\binom{n}{k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{1*2*3\ldots*k} = \frac{n!}{(n-k)!k!}$

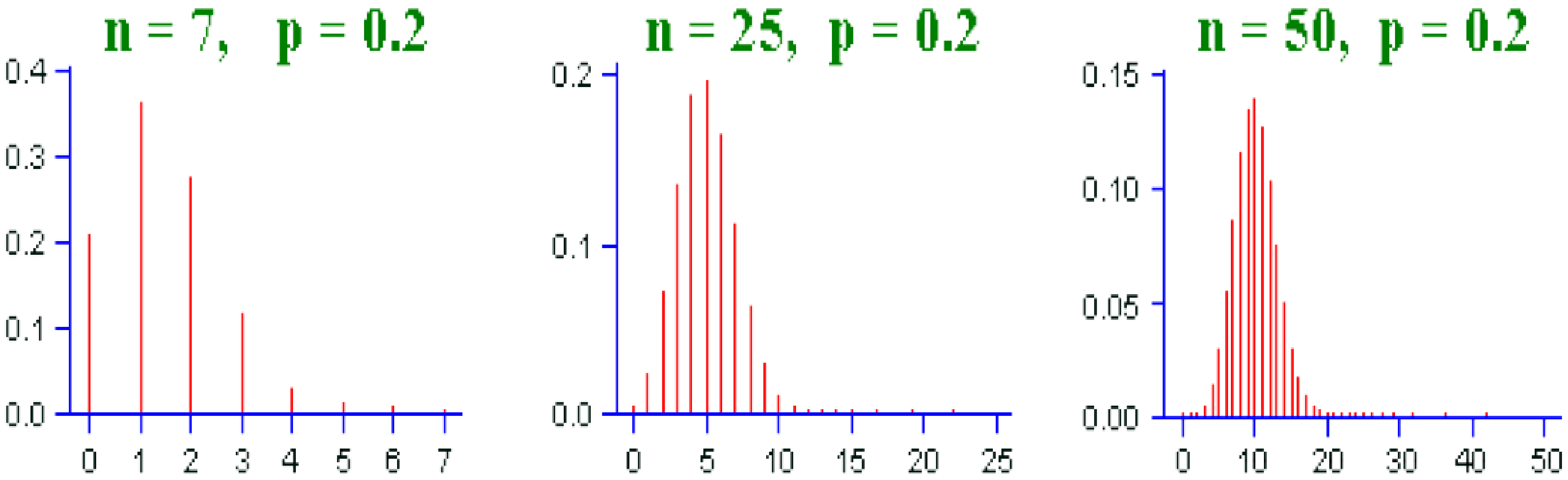
La *forma* della distribuzione della VA Binomiale…

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 $n(1-p) \geq 5$

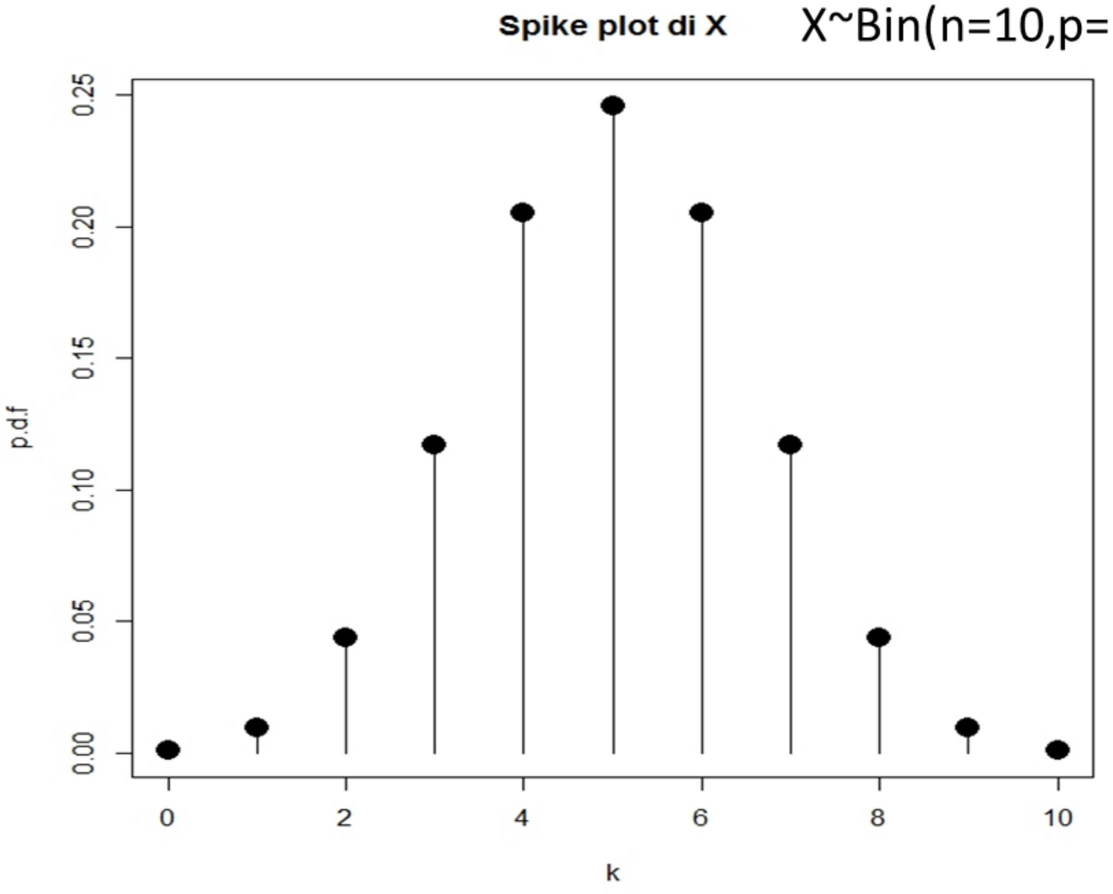
$$Y \approx Bin(n, p)$$

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

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

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

X~Bin(n=10,p=1/2)



*Teorema di De Moivre-Laplace

Intervallo di confidenza per una proporzione

Stima di una proporzione nella popolazione: da un campione casuale di **VA X_1, X_2, …, 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: **$p=P(X_i=1)$.**

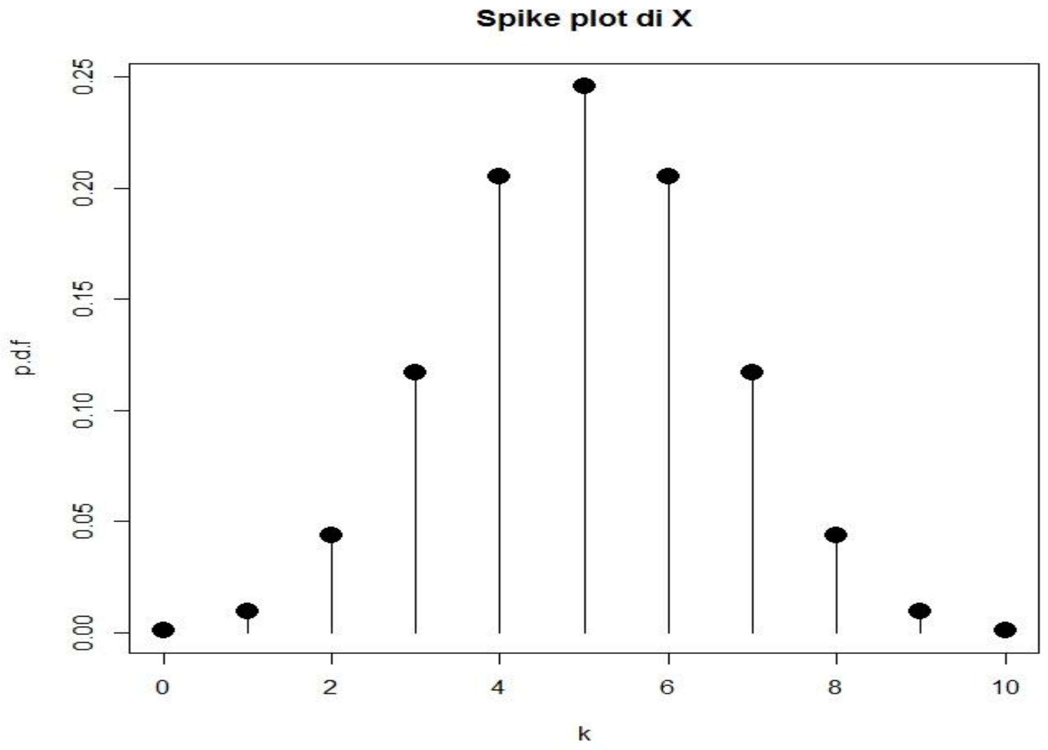
$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{stimatore di p}$$

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

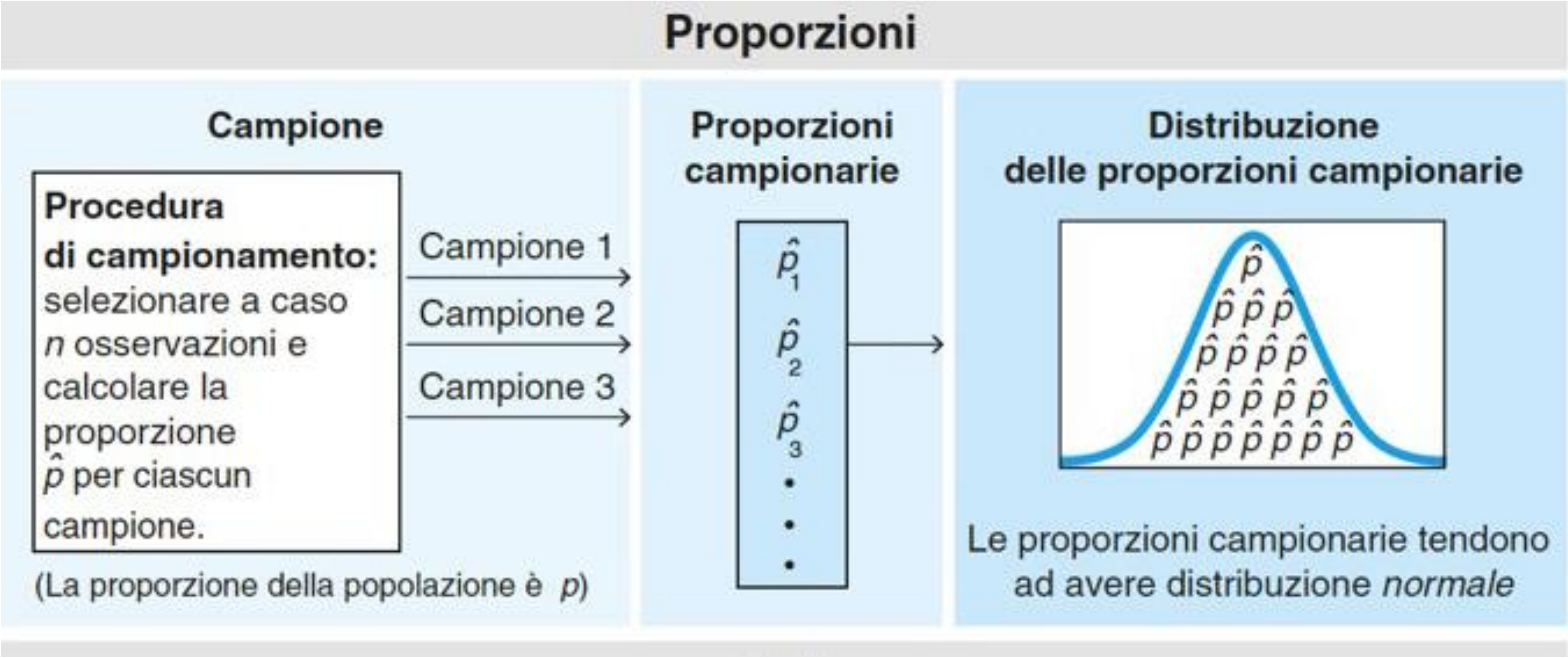
$$\boxed{\bar{p} \pm 1.96 * SE(\bar{p})}$$

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

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



Distribuzione di probabilità della proporzione campionaria



Intervallo di confidenza per una proporzione: esempio sui dati

Si definisce '*attività fisica moderata*' una attività che viene praticata almeno 5 volte a settimana per circa 30 minuti.
Si estrae un campione casuale di **125** studenti, dei quali **47** affermano di fare esercizio meno di 5/settimana.
Quale è l'intervallo di confidenza al **95%** per la «reale» proporzione nella popolazione di studenti «pigri» dell'ateneo?

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

p= 47/125=0.376

$$stderrp = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.376*(1-0.376)}{125}} = 0.0433$$

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

p-1.96*stderrp=0.29

p+1.96*stderrp=0.46

95% IC: [0.29 ; 0.46]

1.96 = costante della gaussiana

Introduzione al Test di Ipotesi: concetti di base

- Si ha una specifica ipotesi su un certo fenomeno nella popolazione che si vuole verificare (ipotesi **nulla** *vs* ipotesi **alternativa**)

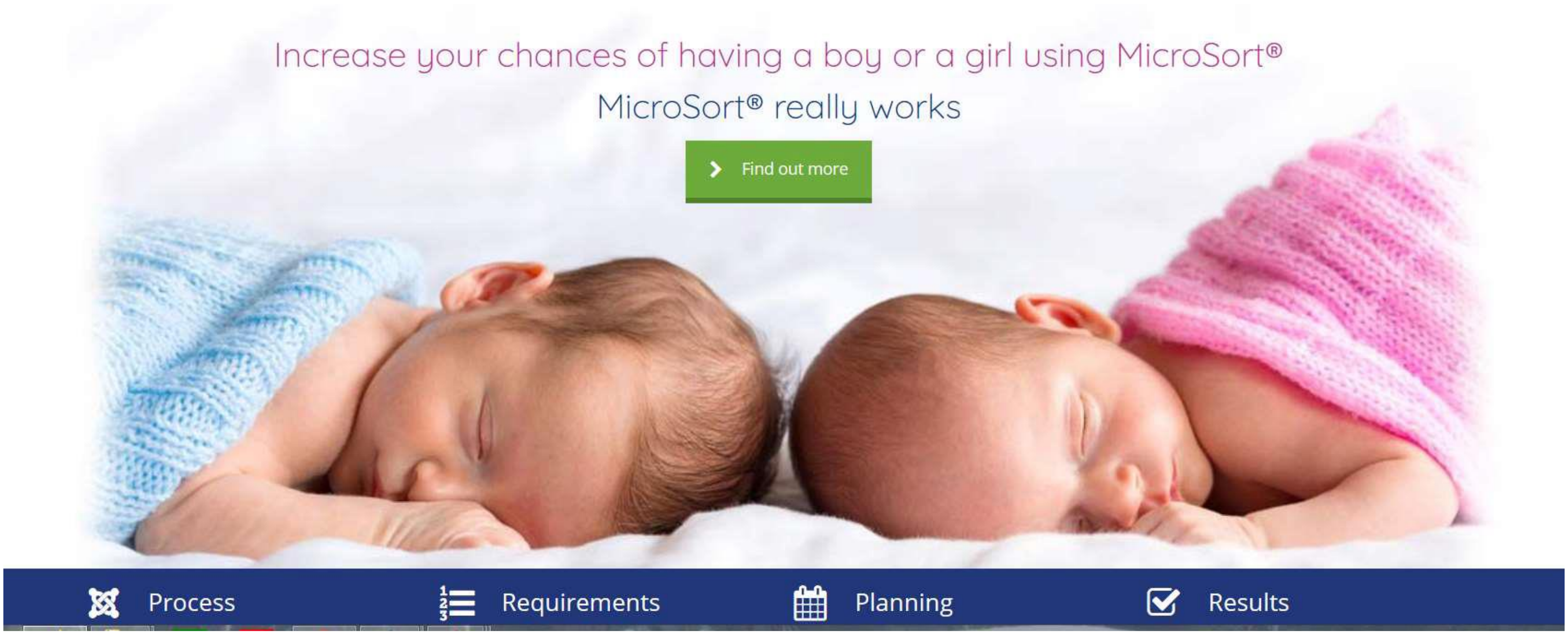
- Si raccolgono dei dati attinenti al problema (dati campionari) [di numerosità ***sufficiente…***]

- 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 ***rigettare*** l'ipotesi nulla

[approccio **frequentista**]

Example: Does the **MicroSort** Method of Gender Selection Increase the Likelihood That a Baby Will Be a Girl?



The Genetics & IVF Institute claims that its XSORT method allows couples **to increase** the probability of having a baby girl.

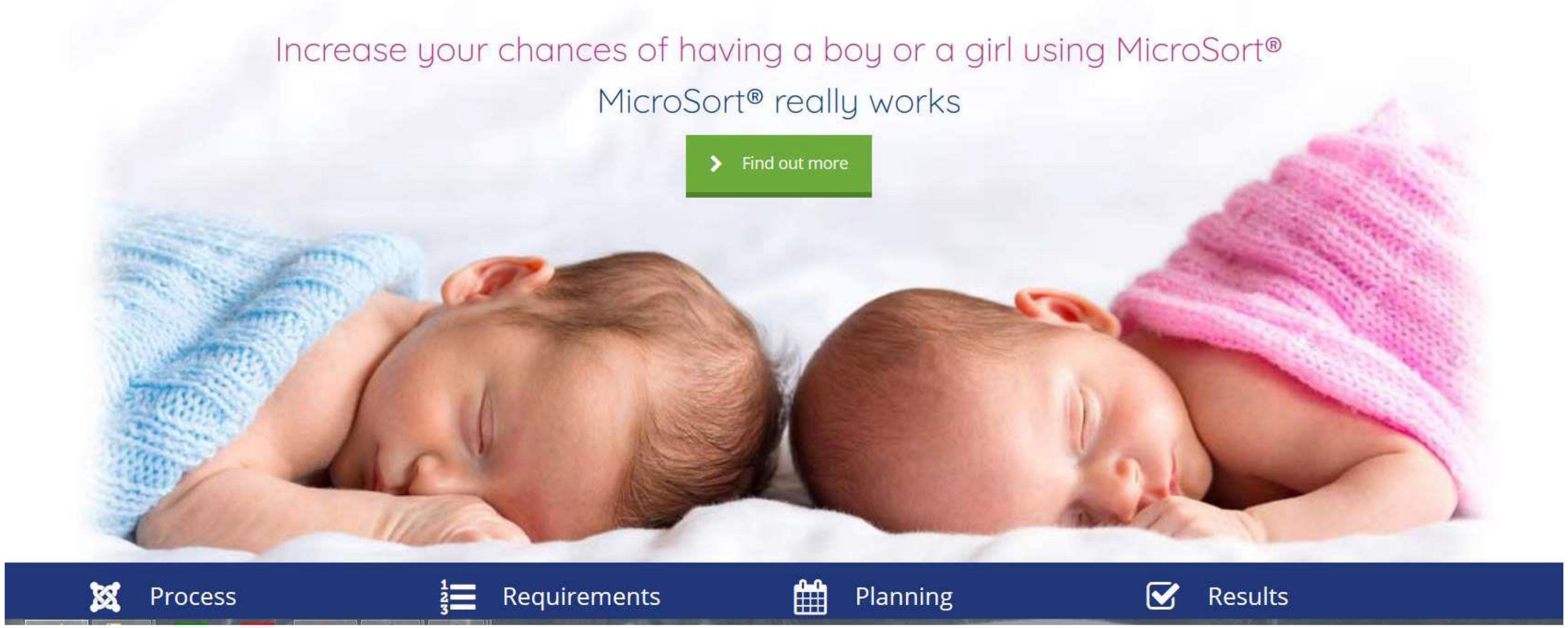
Preliminary results:

- **14** babies born to couples using the XSORT method of gender selection
- **13** of the babies were girls.

Under normal circumstances with no special treatment, girls occur in about **50%** of births.

(Actually, the current birth rate of girls is 48.8%, but we will use 50% to keep things simple.)

Can we *actually support* the claim that the XSORT technique is effective in increasing the probability of a girl?



Test di Ipotesi: tentativo di analogia

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

(a) Ipotesi nulla (H_0): presunzione di innocenza
(b) Ipotesi alternativa (H_1): colpevolezza dell'indagato
(c) Si raccolgono informazioni (**evidenze=dati**) sulla questione
(d) Il giudice **valuta** gli indizi raccolti
(e) 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…

A result of **8 girls out of 14** (or 57.1%) could easily occur *by chance* under normal circumstances with no treatment, so 8 is not significantly high.

The actual result of **13 girls** (or 92.9%) appears to be *significantly* high…

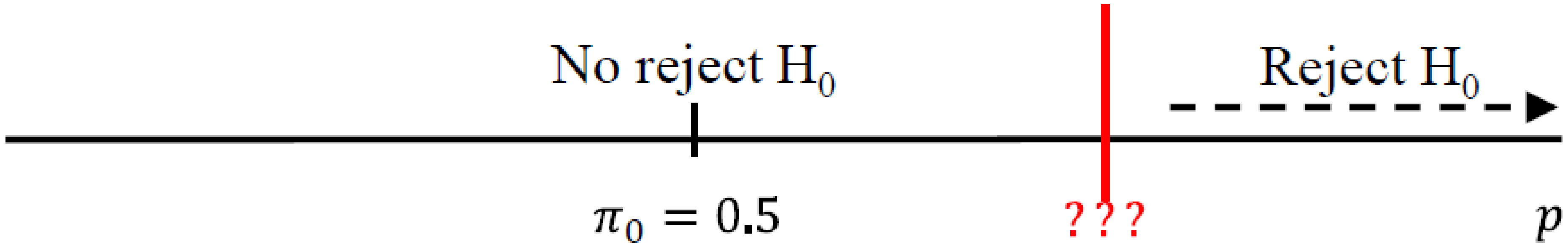
H_0: $\pi = \pi_0 = 0.5$ (null hypothesis: e.g. *the probabilty to get a girl is 50%*)

H_1: $\pi > 0.5$ (alternative hypothesis e.g. *the probabilty to get a girl is higher then 50%*)

Even if the *true probability* of getting a girl is 50% it is possible that *by chance* we observe a **sample probability** which is higher than 50%.

Even if the *true probability* of getting a girl is higher than 50% it is however possible that a **sample probability** is observed that is very close to 50% (or even lower).

In defining the *reject region* we need to control **randomness** or the probability of making mistakes and this can be done by using statistical inference.



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 H_0 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 H_0 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…

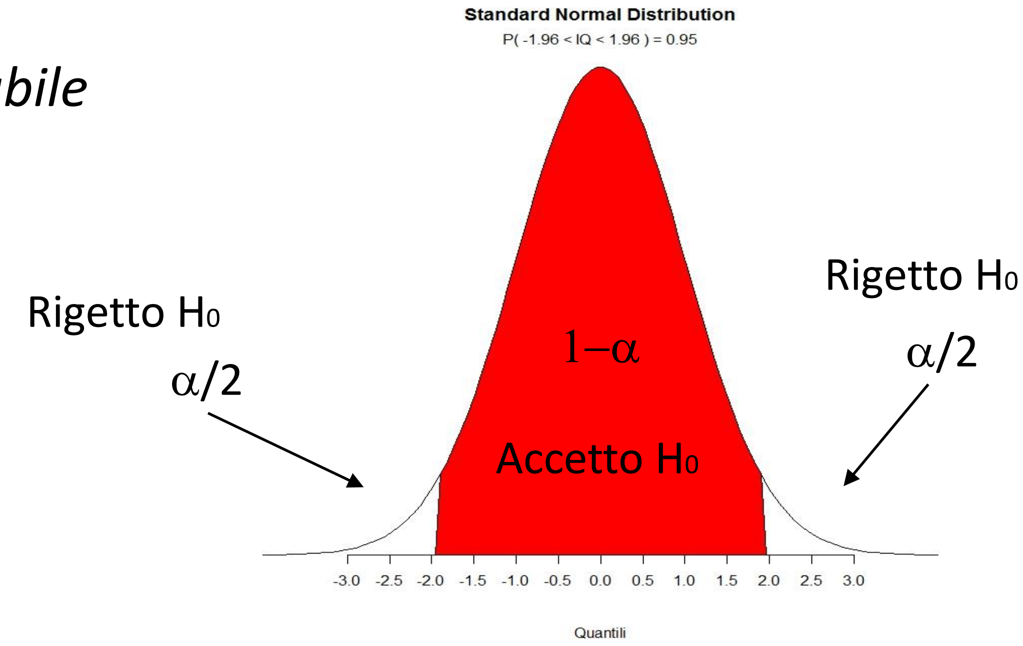
Effettuare un test statistico (strategia generale)

(a) Ipotesi *nulla* **H_0** *versus* Ipotesi *alternativa* **H_1** (mutualmente esclusive)

(b) Si raccoglie un campione di dati **x** (es: glicemia/bimbi nati da coppie in trattamento…)

(c) Si calcola sui dati una *statistica di test* **T(x1, x2,…)=t** la cui distribuzione di probabilità è nota se vale H_0 (e differisce rispetto a quello che avrebbe sotto H_1)

(d) Si rigetta H_0 se il valore osservato di **t** è troppo *poco probabile* (se H_0 fosse vera): ***p* <= α**



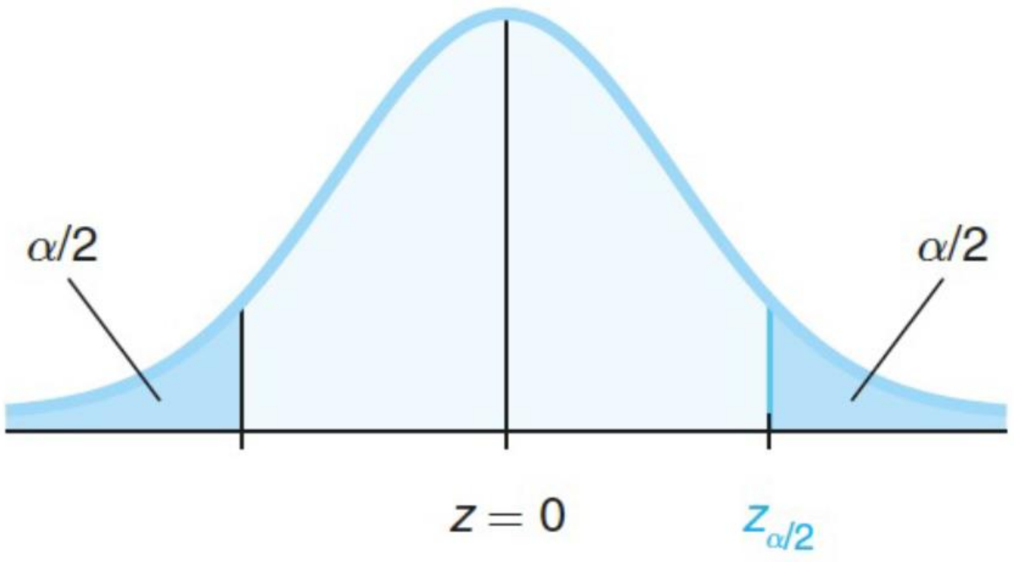
p-value*= probabilità sotto H_0 che la variabile casuale T abbia il valore t osservato sui dati campionari o un valore più *«estremo»

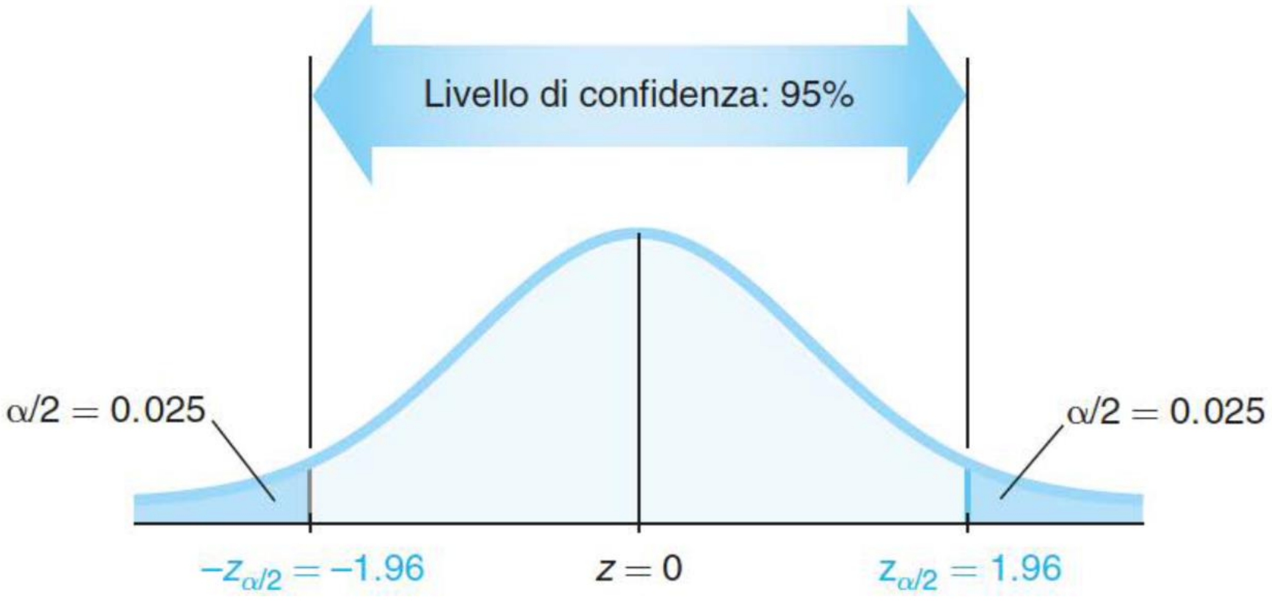
-> P(Data | H_0)

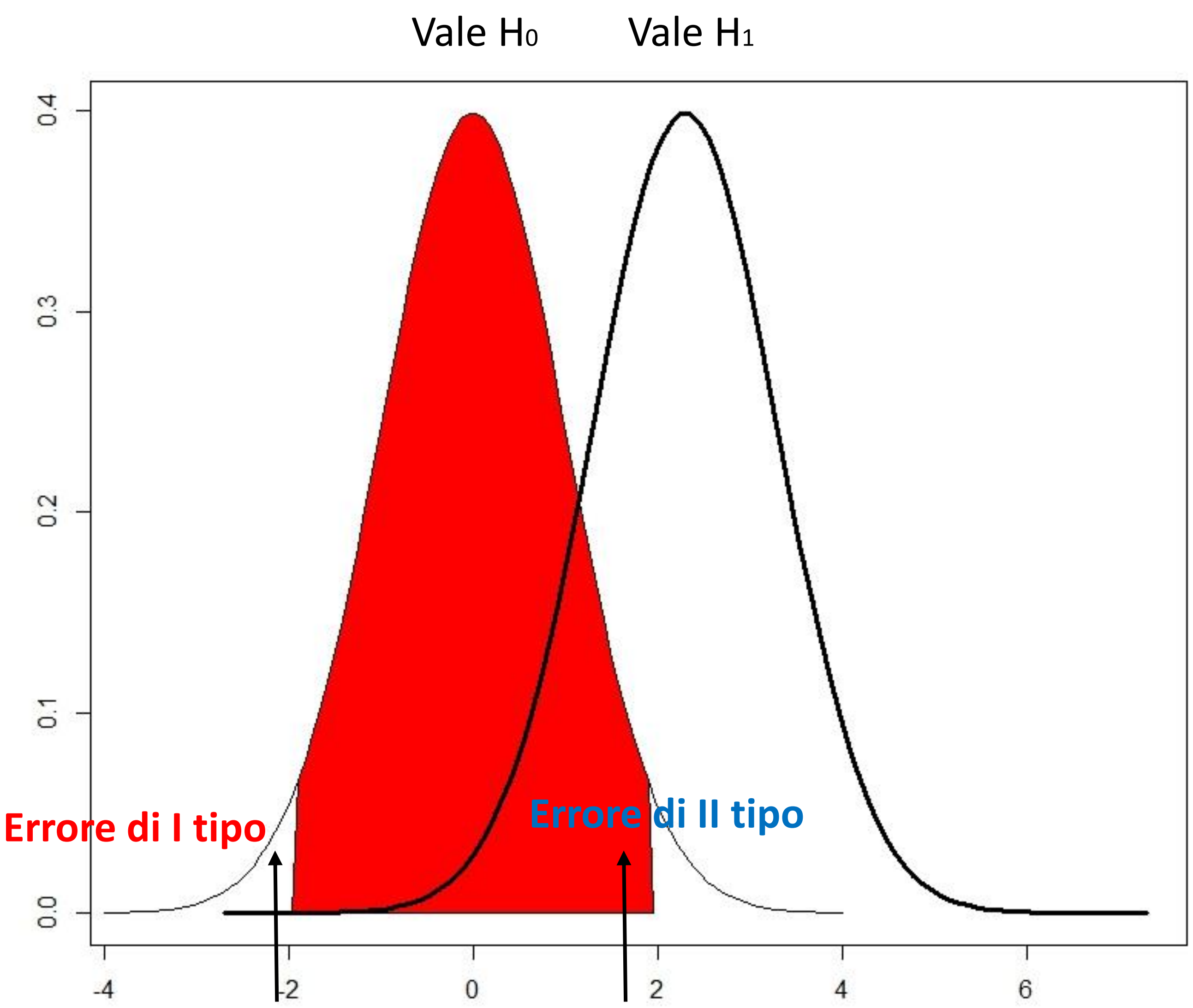
Un **valore critico** è il valore corrispondente al quantile della distribuzione che delimita l'area con maggiore possibilità di verificarsi.

Il quantile $z_{\alpha/2}$ è un valore critico, cioè un valore z che ha la proprietà di separare un'area di probabilità $\alpha/2$ nella coda destra della distribuzione normale standard:

Spesso nei test di ipotesi il valore critico delimita una probabilità complessiva nelle code del **5%**.







		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 **

Conclusioni & Conseguenze

Supponiamo che α sia piccolo, tipicamente $\alpha <= 0.05$

- 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!!!**

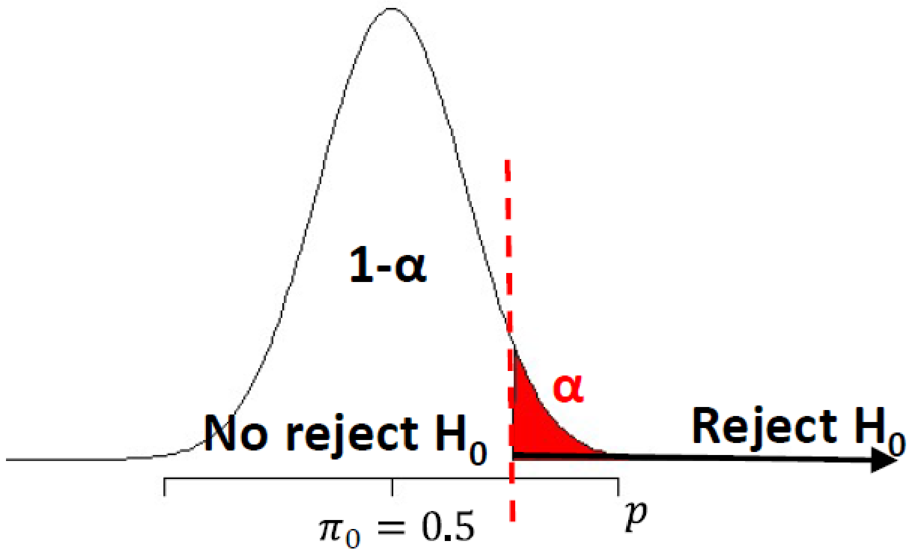
We know the **theoretical distribution** of the sample probability:

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{stimatore di p} \qquad \frac{\bar{p}-p}{SE(\bar{p})} \approx N(0,1)$$

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

Under the null (true probability of getting a girl is 50%,): $\boldsymbol{H_0: \pi = \pi_0 = 0.5}$

the theoretical distribution of the sample probability: $\boldsymbol{p \sim N(0.5, \sqrt{0.5 * 0.5/14})}$



We can then define the **critical rejection region** in order to establish a priori the probability of making mistakes when we reject H_0.

This probability is called significance level α.

Under the null (H_0): XSORT does not work

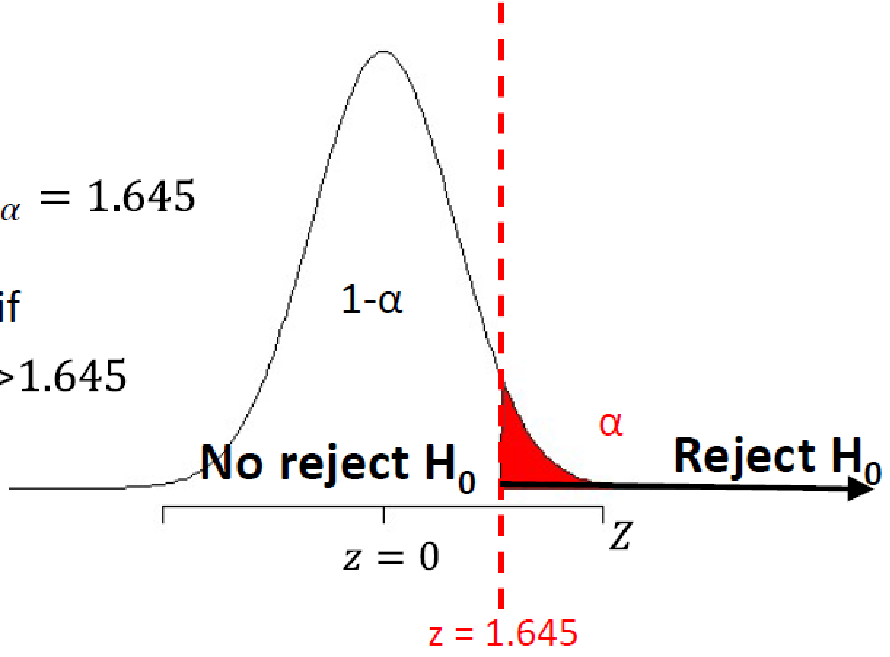
$n=14$
$p=13/14=0.929$
$\alpha = 0.05 \quad z_{0.95} = 1.645$
$$z = \frac{p-\pi_0}{se(p)}$$
$$z = \frac{0.929 - 0.5}{\sqrt{0.5 * 0.5/14}} = 3.21$$

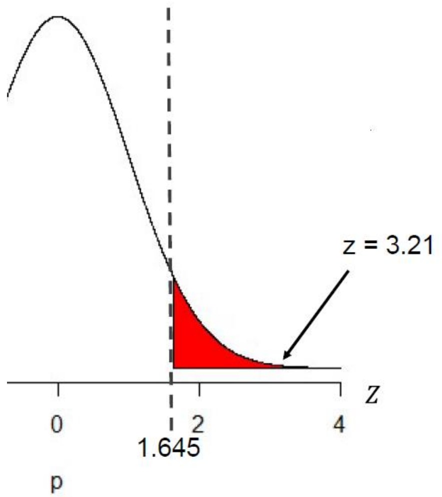
Better standardise:
$$z = \frac{p-\pi_0}{se(p|H_0)} \qquad Z \sim N(0,1)$$
$$z = \frac{p - 0.5}{\sqrt{0.5 * 0.5/14}}$$

When the level of significance α is set, the threshold is the (1-α)th percentile $z_{1-\alpha}$

Ex. $\alpha = 0.05$ -> $z_{1-\alpha} = 1.645$

Thus we will reject if
$$z = \frac{p-0.5}{\sqrt{(0.5*0.5)/14}} > 1.645$$





Reject the null hypothesis! There is sufficient sample evidence to support the claim that for couples using the «XSORT gender selection method», most (more than half) of their babies are girls.

Esempio di test di ipotesi per la differenza tra due medie

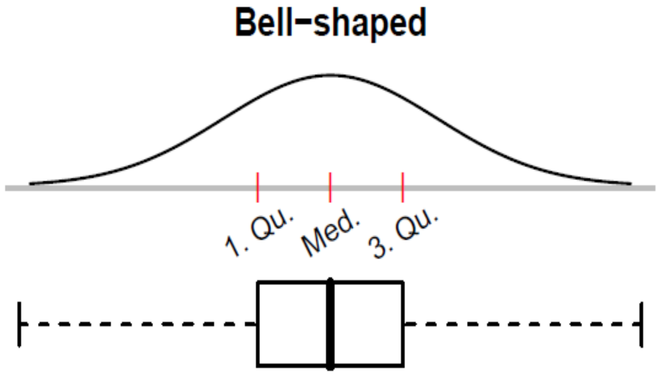
Dati due campioni del carattere x* misurato in due gruppi:

$$H_0:\ \varepsilon = 0$$

$$H_1:\ \varepsilon \neq 0$$

$$\varepsilon = \bar{x}_1 - \bar{x}_2$$

* per i quali abbia senso utilizzare la media come indice di posizione…



- Se le VA sono **indipendenti** → t-test
  Es: due gruppi diversi di pazienti

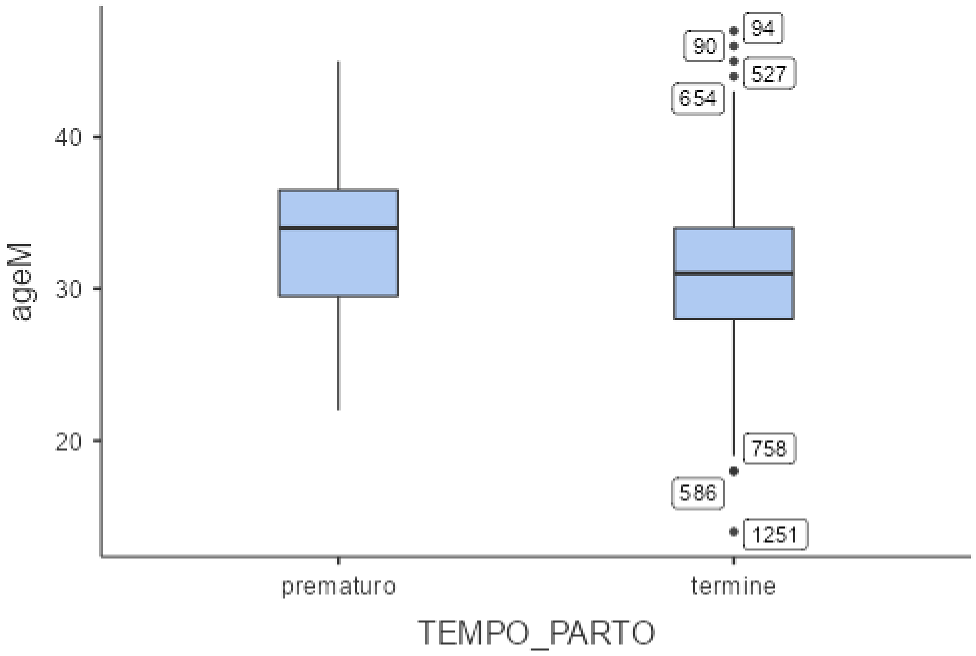
- Se le VA *non sono indipendenti* → t-test per dati ***accoppiati***
  Es: stesso gruppo pre-post

NB: nell'utilizzo del t-test occorre specificare se le varianze dei due campioni sono ***simili*** oppure no (F-test)

Test di ipotesi per la differenza tra due medie: esempio

C’è una differenza nella età media delle madri inglesi che hanno avuto un parto prematuro rispetto alle madri che hanno avuto un parto a termine?

1. Verifichiamo la simmetria rispetto alla media delle distribuzioni:



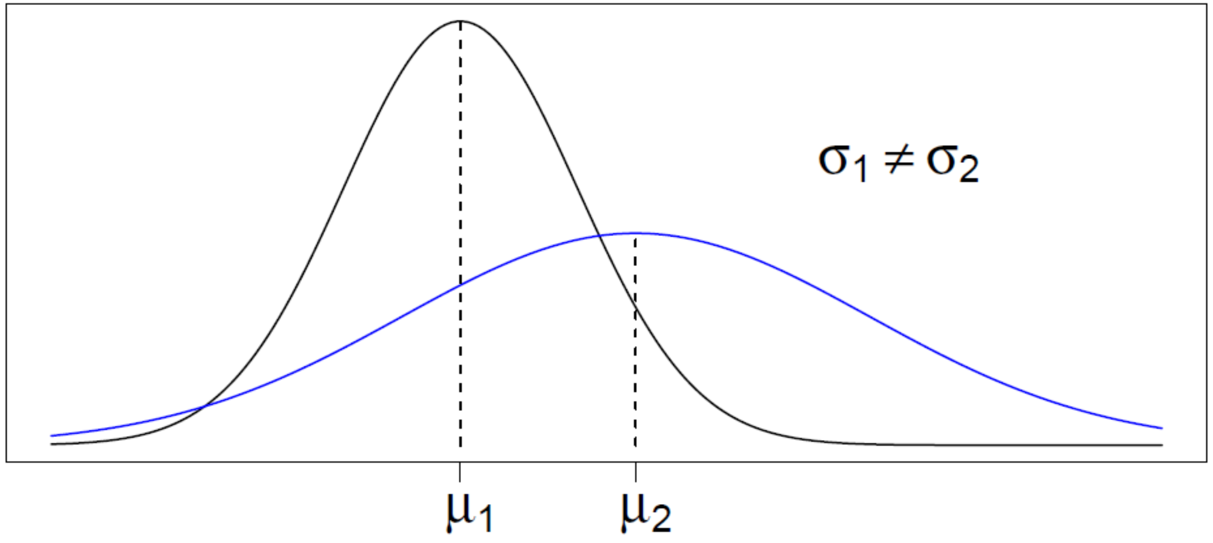
Descrittive

	TEMPO_PARTO	ageM
N	prematuro	99
	termine	1344
Media	prematuro	33.4
	termine	30.8
Mediana	prematuro	34
	termine	31.0
Deviazione standard	prematuro	4.89
	termine	4.14
Minimo	prematuro	22
	termine	14
Massimo	prematuro	45
	termine	47
Asimmetria	**prematuro**	**0.0417**
	termine	**0.0840**
Curtosi	**prematuro**	**-0.425**
	termine	**0.398**

Test di ipotesi per la differenza tra due medie: esempio

2. Prima di effettuare il test sulla media, dobbiamo verificare se le varianze* delle distribuzioni siano omogenee (si utilizza la statistica di test F):

$$H_0 : \sigma_X = \sigma_Y$$
$$H_1 : \sigma_X \neq \sigma_Y$$



Ipotesi

Test di Omogeneità delle Varianze (Levene)

	F	gdl	gdl2	p
ageM	4.47	1	1441	0.035

Nota. Un piccolo valore di p suggerisce una violazione dell'assunzione di varianze uguali

Rigettiamo l'ipotesi nulla di omogeneità delle varianze

*Il rapporto fra due varianze segue la distribuzione F

Test di ipotesi per la differenza tra due medie: esempio

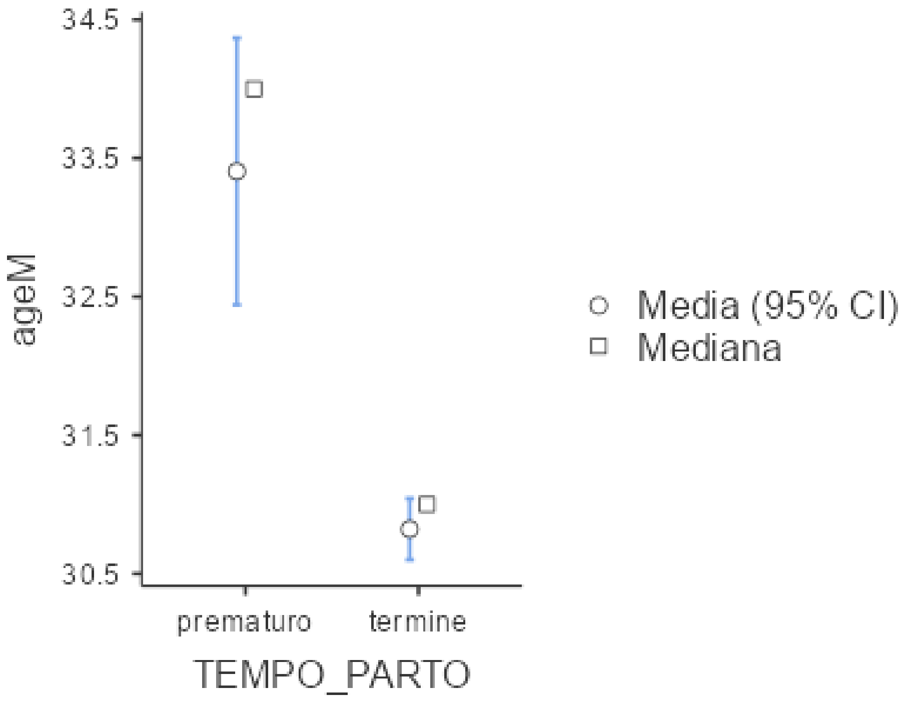
3. Effettuiamo adesso il t-test (varianze non omogenee):

Test t a campioni indipendenti

Test t a campioni indipendenti

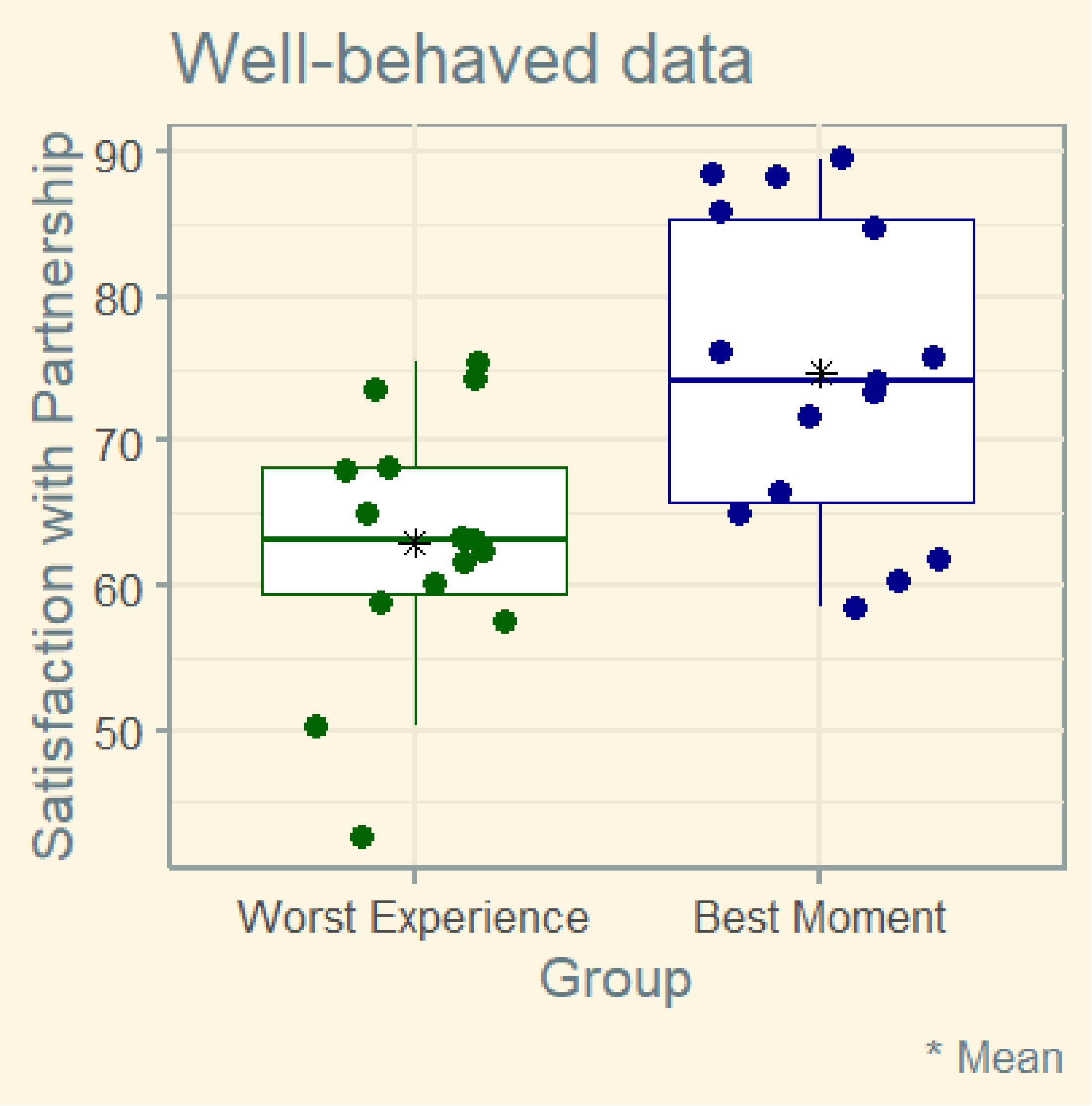
		Statistiche	gdl	p	Differenza media	Differenza SE	95% Intervallo di Fiducia	
							Inferiore	Superiore
ageM	t di Welch	5.12	109	< .001	2.58	0.504	1.58	3.58

Nota. H_a $\mu_{prematuro} \neq \mu_{termine}$



Suggerisce una differenza fra le medie significativa ($p < 0.001$)

Impatto della asimmetria/outliers sugli esiti del test

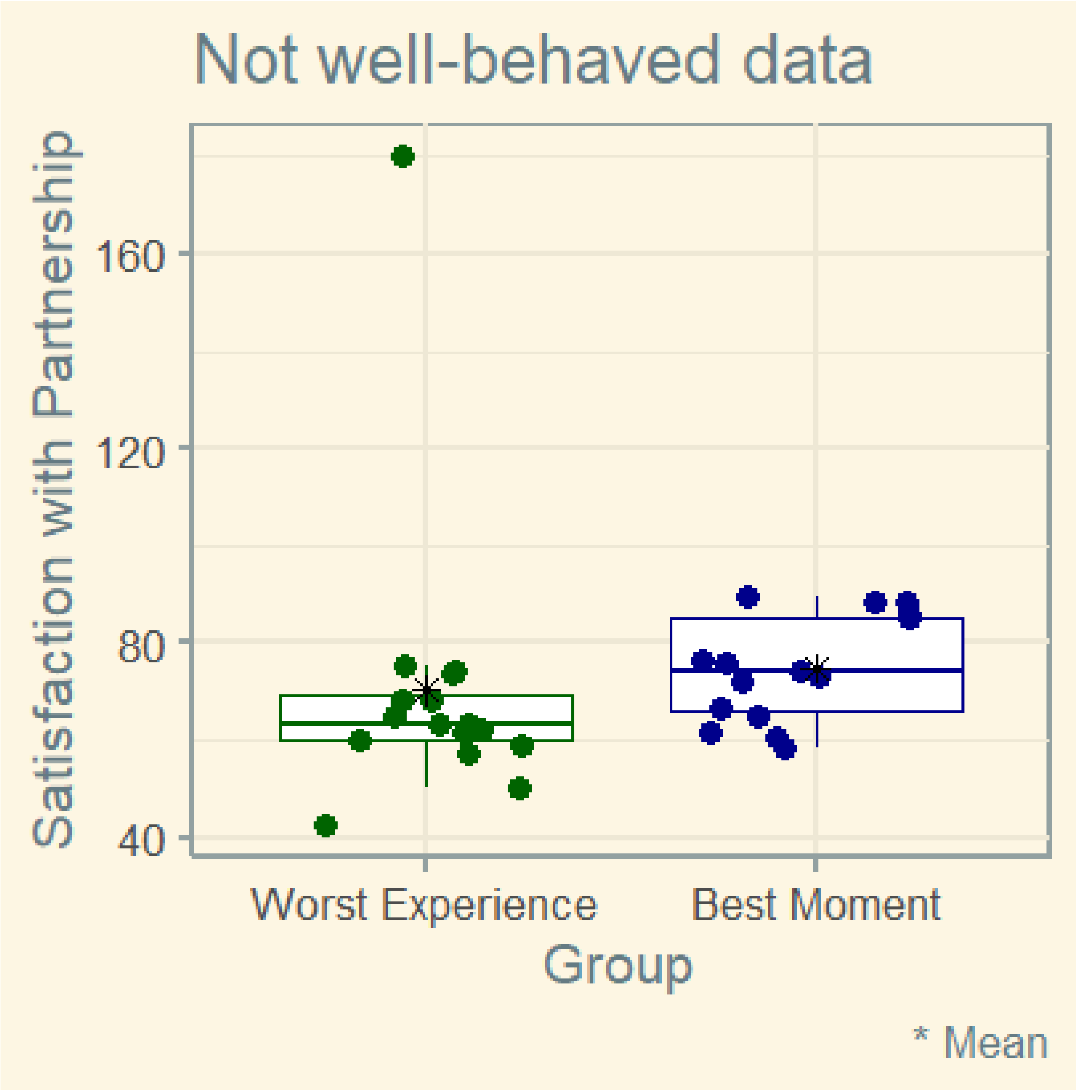


Characteristic	Worst Experience, N = 15[1]	Best Moment, N = 15[1]	p-value[2]
satisfaction	63 (9)	75 (11)	0.003

[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.



Characteristic	Worst Experience, N = 16[1]	Best Moment, N = 15[1]	p-value[2]
satisfaction	70 (30)	75 (11)	0.6

[1] Mean (SD)

[2] Welch Two Sample t-test

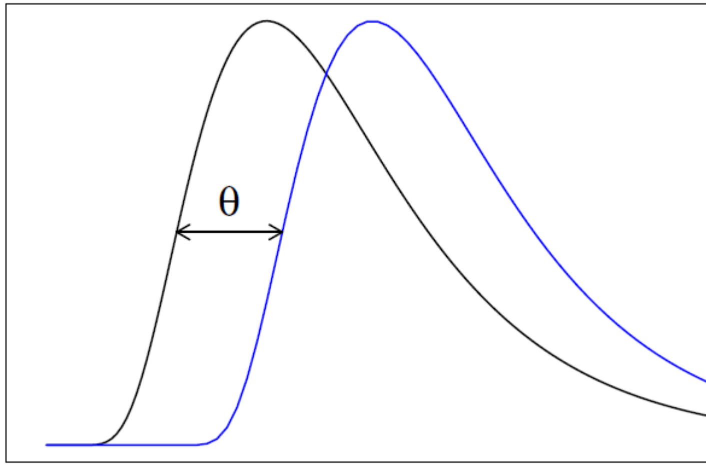
Il t-test (opzione varianze non omogenee) **non trova** una differenza significativa tra le medie di questi due gruppi ….

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:



$$H_0 : \theta = 0$$

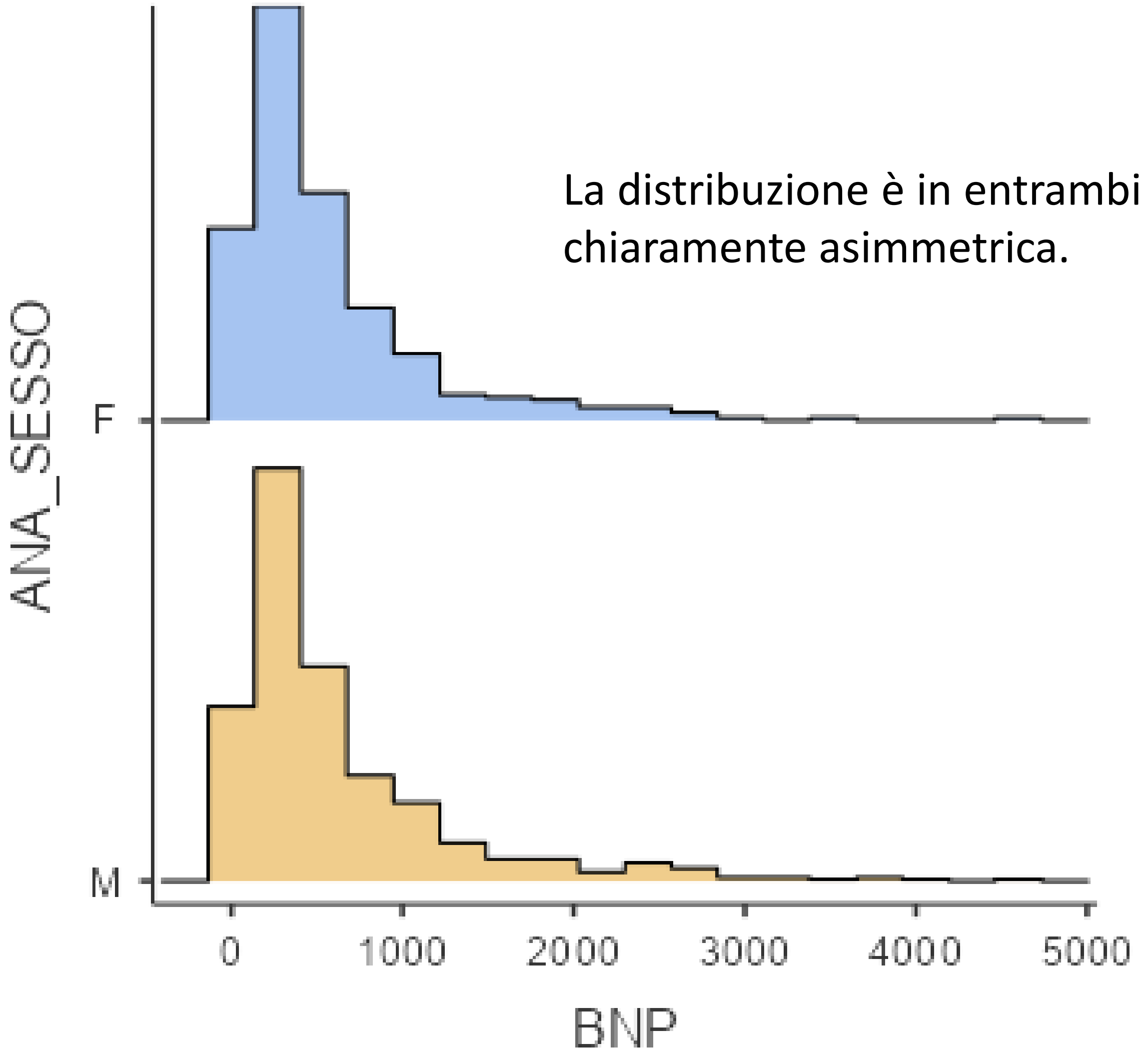
$$H_1 : \theta \neq 0$$

- 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

Wilcoxon-Mann-Whitney test: esempio sui dati

Istogramma della distribuzione dei valori del BNP in un campione di pazienti cardiovascolari, distinti per sesso.



La distribuzione è in entrambi i gruppi chiaramente asimmetrica.

Descrittive

	ANA_SESSO	BNP
N	F	439
	M	618
Mancanti	F	619
	M	852
Media	**F**	**548**
	M	**613**
Mediana	**F**	**361**
	M	**395**
Deviazione standard	F	585
	M	662
Minimo	F	6.10
	M	8.20
Massimo	F	4606
	M	4559
Asimmetria	**F**	**2.54**
	M	**2.36**
Curtosi	**F**	**8.96**
	M	**6.85**

Wilcoxon-Mann-Whitney test: esempio sui dati

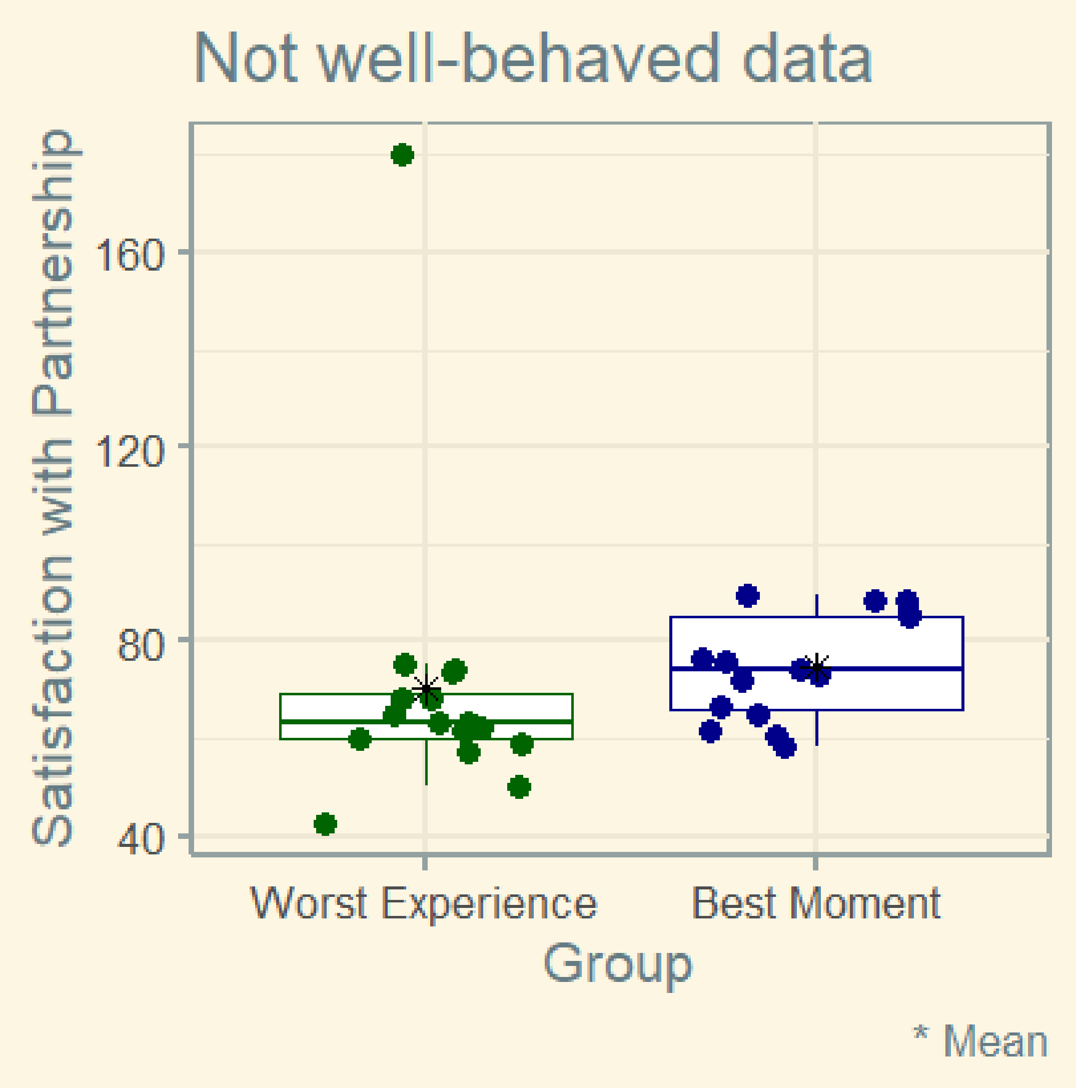
Test non parametrico

		Statistiche	p
BNP	U di Mann-Whitney	128633	0.151

Nota. H_a $\mu_F \neq \mu_M$

	SESSO	N	Mancanti	Media	Mediana	SD	Minimo	Massimo	Percentili		
									25th	50th	75th
BNP	F	439	619	548	361	585	6.10	4606	174	361	686
	M	618	852	613	395	662	8.20	4559	192	395	773

Non si rileva una differenza *statisticamente significativa* tra i percentili delle distribuzioni di BNP per sesso (p=0.151).



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.

Test di ipotesi per dati di tipo categorico/espressi in classi

- 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					
	1	2	3	. . .	C	Total
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 x 54 = 22.5	50/120 x 66 = 27.5	50
No	70/120 x 54 = 31.5	70/120 x 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 o in uno studio trasversale (cross-sectional study);

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

Facciamo il punto:

- Per (quasi) ogni tipo di ipotesi da sottoporre a verifica si può trovare il test opportuno…;
- 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!