

# TEST D'IPOTESI

## **Il confronto di proporzioni**

# Test su una proporzione

## Esempio

Supponiamo di voler saggiare se la *vera* sensibilità ( $\pi$ ) di un nuovo test diagnostico è pari a  $\pi_0=0.95$ , come dichiarato dal produttore del test, oppure ha un valore inferiore

**H<sub>0</sub>** :  $\pi = \pi_0$  (la *vera* sensibilità del test è pari a 0.95)

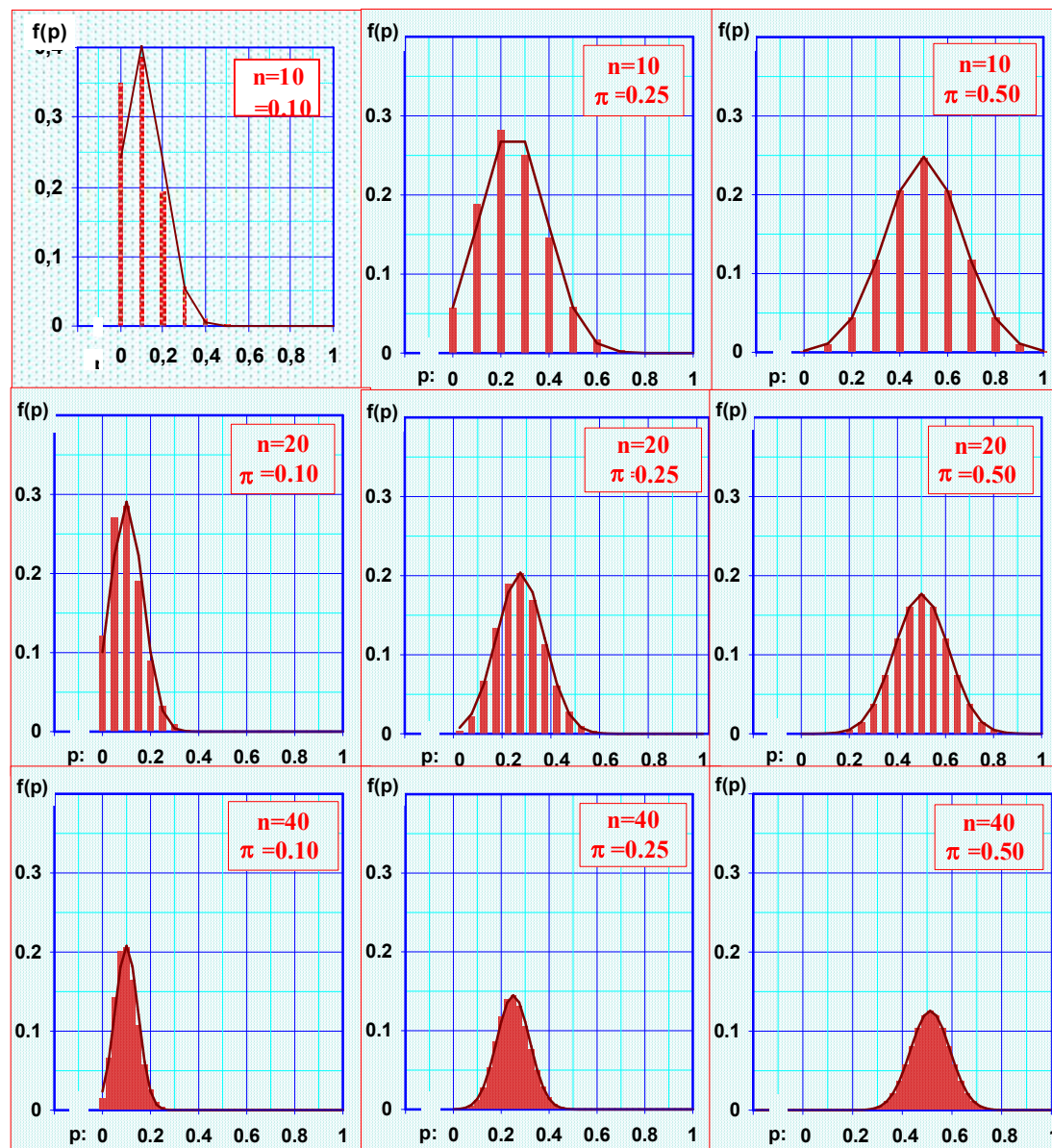
**H<sub>1</sub>** :  $\pi < \pi_0$  (la *vera* sensibilità del test è minore di 0.95)

L'ipotesi alternativa è unidirezionale e quindi il test sarà ad una coda.

Stabiliamo il livello di significatività  $\alpha = 0.05$

## Distribuzione delle frequenze relative: Binomiale e Gaussiana

La frequenza relativa ( $p$ ) di un carattere, calcolata su campioni tratti da un universo in cui il carattere ha prevalenza pari a  $\pi$ , ha distribuzione **binomiale**.  
All'aumentare della dimensione ( $n$ ) del campione, i valori di  $p$  mostrano tendenza a crescere ed accentrarsi attorno al parametro  $\pi$ , **approssimando la distribuzione gaussiana**.



# Test d'ipotesi su una proporzione

Sappiamo che, se  $n$  è sufficientemente grande (e il valore *più piccolo* tra  $n\pi$  e  $n(1-\pi)$  è maggiore di 5) la distribuzione di una proporzione (variabile binomiale) è approssimata dalla gaussiana e, sotto  $H_0$

$$p \sim N(\pi_0, \pi_0(1-\pi_0)/n)$$

Consideriamo allora la corrispondente devziata gaussiana standardizzata

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \sim N(0,1)$$

Dopo aver ottenuto la nostra stima campionaria di  $p$ , calcoleremo il valore  $z$  della statistica test. **Se la stima di  $p$  è molto lontana da  $\pi_0$ ,  $z$  avrà un valore estremo che ci porterà a rifiutare  $H_0$ .** In caso contrario, non rifiuteremo  $H_0$ .

# Esempio

In un campione di 290 soggetti ammalati, 272 risultano positivi al test

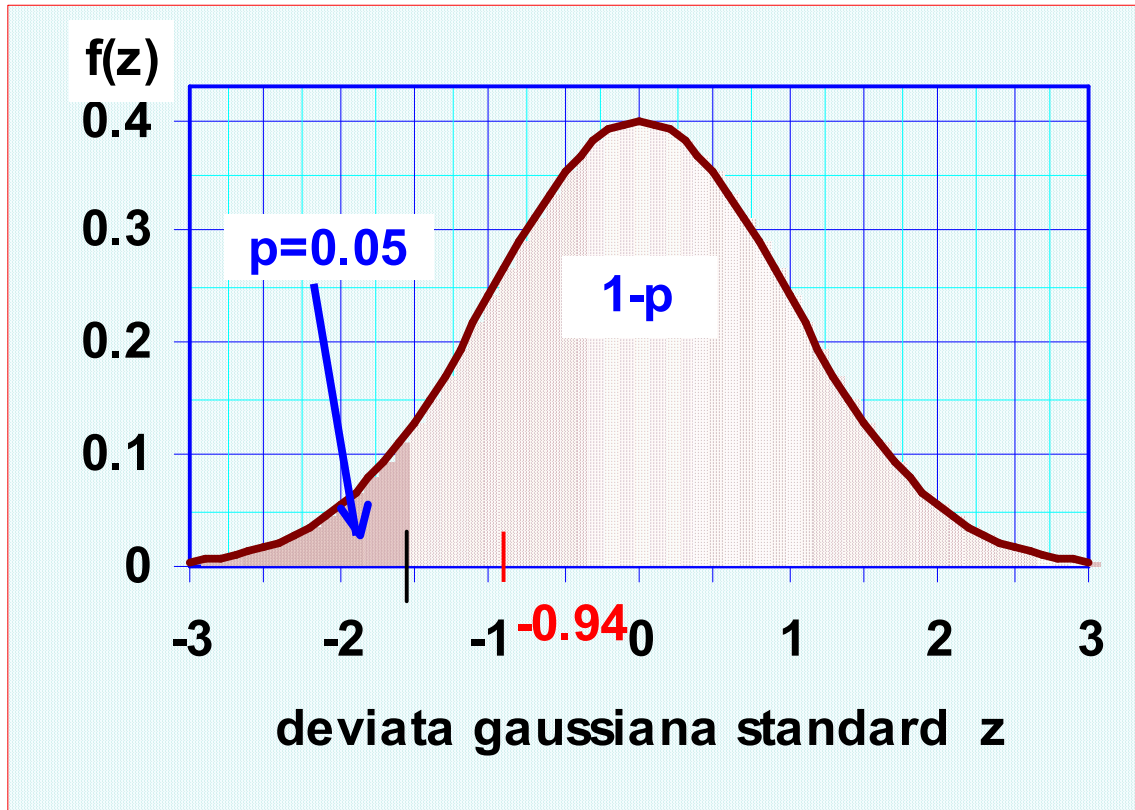
$$p = \frac{272}{290} = 0.938 \quad \text{è la stima della sensibilità vera del test}$$

In base a questa osservazione posso rifiutare  $H_0$ ?

$$z = \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} = \frac{0.938 - 0.950}{\sqrt{0.95 \times 0.05/290}} = \frac{-0.012}{0.0128} = -0.94$$

Il valore di  $z$  cade nella regione di rifiuto?

# Conclusione (1)



Il valore della statistica  $z$  è maggiore del 5° percentile della deviana gaussiana standard (-1.64)  
Non cade quindi nella regione di rifiuto

Non rifiuto l'ipotesi nulla che la sensibilità *vera* del test sia pari a 0.95

# Conclusione (2)

Non rifiuto l'ipotesi nulla, ovvero concludo, in base ai dati campionari, che non ho evidenza per dire che il test ha minore sensibilità di quella dichiarata (0.95).

Posso anche calcolare il p-value relativo al risultato del test.

P-value = 0.17

# Intervallo di confidenza per una proporzione

Sappiamo che, se  $n$  è sufficientemente grande (e il valore *più piccolo* tra  $n\pi$  e  $n(1-\pi)$  è maggiore di 5) la distribuzione di una proporzione (variabile binomiale  $B(\pi, n)$ ) è approssimata dalla gaussiana

$$p \sim N [\pi, \pi(1-\pi)/n]$$

L'espressione dell'**intervallo di confidenza di  $\pi$**  si ottiene con la seguente approssimazione:

$$\text{I.C.}_{1-\alpha} = p \pm z_{\alpha/2} \times \sqrt{p(1-p)/n}$$



# ATTENZIONE

L'errore standard (ES) usato nella statistica **test**  
(calcolato sotto  $H_0$  )

**è diverso**

da quello usato

per il calcolo dell'**intervallo di confidenza**

**(che prescinde da  $H_0$  )**

$$\sqrt{\pi_0(1-\pi_0)/n} \neq \sqrt{p(1-p)/n}$$

Tuttavia questo non comporta di solito, in pratica, risultati contraddittori (perché  $p(1-p)$  varia abbastanza lentamente al variare di  $p$ )

- In un campione di 290 soggetti affetti da malattia
- 272 risultano positivi al test diagnostico, allora la sensibilità stimata è  $p=272/290=0.938$
- L'intervallo di confidenza al 90% sarà:

$$\begin{aligned} \text{I.C.}_{90\%} &= 0.938 \pm 1.64 \times \sqrt{0.938 \times 0.062 / 290} \\ &= 0.938 \pm 1.64 \times 0.014 = [0.914; 0.962] \end{aligned}$$

- **La vera sensibilità ( $\pi$ ) del test diagnostico è un qualunque valore incluso tra 0.914 e 0.962.** La probabilità che tale affermazione corrisponda a verità è di **circa il 90%**.

# Test d'ipotesi per la differenza tra due proporzioni

Si vuole saggiare se due farmaci antiepilettici (A e B) hanno differente efficacia. L'indice di efficacia adottato (end-point dello studio) è la proporzione di pazienti che, dopo un mese dall'inizio della terapia, hanno ridotto almeno della metà la frequenza delle crisi.

Indichiamo con  $\pi_A$  e  $\pi_B$  la *vera* efficacia dei trattamenti A e B.

**H<sub>0</sub>** :  $\pi_A = \pi_B = \pi$  (i farmaci A e B hanno uguale efficacia)

**H<sub>1</sub>** :  $\pi_A \neq \pi_B$  (i farmaci A e B hanno diversa efficacia)

Fissiamo  $\alpha=0.05$  (due code)

# Il test d'ipotesi

**Sotto ipotesi nulla** la differenza tra le proporzioni campionarie  $p_A$  e  $p_B$  ha distribuzione che approssima la distribuzione gaussiana

$$(p_A - p_B) \sim N \left[ 0, \pi(1 - \pi) \times \left( \frac{1}{n_A} + \frac{1}{n_B} \right) \right]$$

e il rapporto

$$z = \frac{p_A - p_B}{\sqrt{\pi(1 - \pi) \times \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

è una devinata gaussiana standard

# Esempio

Sono stati campionati 330 soggetti per gruppo.

Per motivi non dipendenti né dalla comparsa di effetti collaterali né dall'efficacia del farmaco, si sono ritirati dallo studio 18 pazienti trattati con il **farmaco A** e 24 pazienti trattati con il **farmaco B**.

Sui rimanenti soggetti (*soggetti valutabili*) si sono calcolati i seguenti indici di efficacia:

$$p_A = 240/312 = 0.769 \quad p_B = 210/306 = 0.686$$

Sotto ipotesi nulla, la stima ( $p$ ) del parametro di efficacia  $\pi$  comune ai due farmaci è data dal rapporto tra il numero di pazienti in cui la terapia si è rivelata efficace ed il numero totale dei pazienti valutabili:

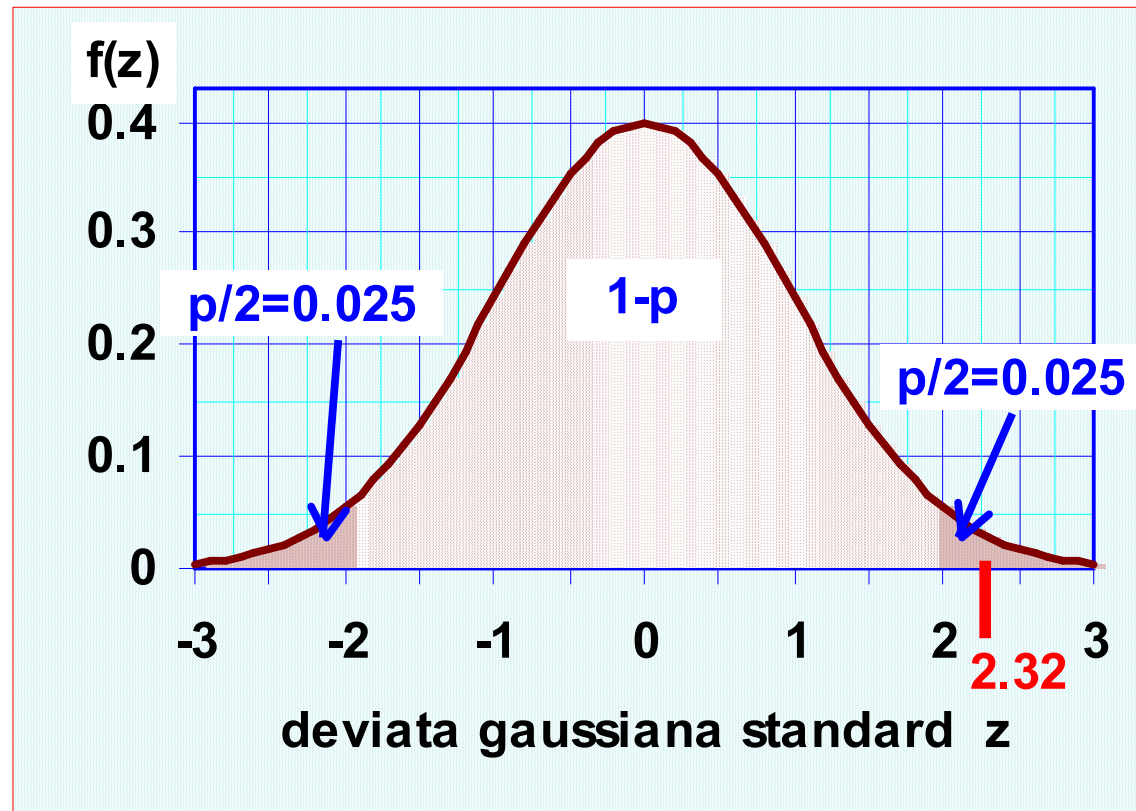
$$p = (240+210)/(312+306) = 450/618 = 0.728$$

Tale valore sostituisce  $\pi$  nella stima dell'errore standard della differenza ( $p_A - p_B$ )

$$z = \frac{0.769 - 0.686}{\sqrt{0.728 \times 0.272 \times \left( \frac{1}{312} + \frac{1}{306} \right)}} = \frac{0.083}{0.0358} = 2.318 \quad p = 2 * 0.01 = 0.02$$

Rifiuto  $H_0$  : l'efficacia del **farmaco A** è significativamente maggiore di quella del **farmaco B**

Se veramente non vi fosse differenza tra i trattamenti, un risultato (in valore assoluto) uguale o più estremo (nelle code della distribuzione) di quello osservato si verificherebbe meno di 5 volte su 100 (2 volte su 100 dato il p-value)



# Intervallo di confidenza

Per il calcolo dell'**intervallo di confidenza**, l'errore standard della differenza tra due proporzioni viene calcolato **a prescindere dall'ipotesi nulla** e si basa sulle proporzioni campionarie  $p_A$  e  $p_B$ :

$$\text{I.C.}_{1-\alpha} = (p_A - p_B) \pm z_{\alpha/2} \times \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

Quindi, sostituendo le proporzioni campionarie  $p_A$  e  $p_B$  ai parametri  $\pi_A$  e  $\pi_B$  si ottiene:

$$\text{ES}(p_A - p_B) = \sqrt{\frac{0.769 \times 0.231}{312} + \frac{0.686 \times 0.314}{306}} = 0.0357$$



Nel nostro *esempio*,

fissata la confidenza del 95%, si ha:

$$\text{I.C.95\%} = 0.083 \pm 1.96 \times 0.0357 = [0.013; 0.153]$$

Si può affermare che:

il valore vero della differenza ( $\delta$ ) tra i trattamenti è incluso tra i valori dell'intervallo [0.013; 0.153]

**che non comprende lo zero!**

(con confidenza 95%, ovvero con la probabilità di sbagliare 5 volte su 100)

# Esercizio

- Studio randomizzato
- Follow-up medio di 7.4 anni
- Evento: problema coronarico

Risultati	Placebo	Farmaco	Totali
N° di soggetti con evento	187	155	342
N° totale di soggetti	1900	1906	3806

**Il farmaco determina un numero diverso di eventi?**

# Test a due code

$$H_0: \pi_p = \pi_f \text{ oppure } \pi_p - \pi_f = 0$$

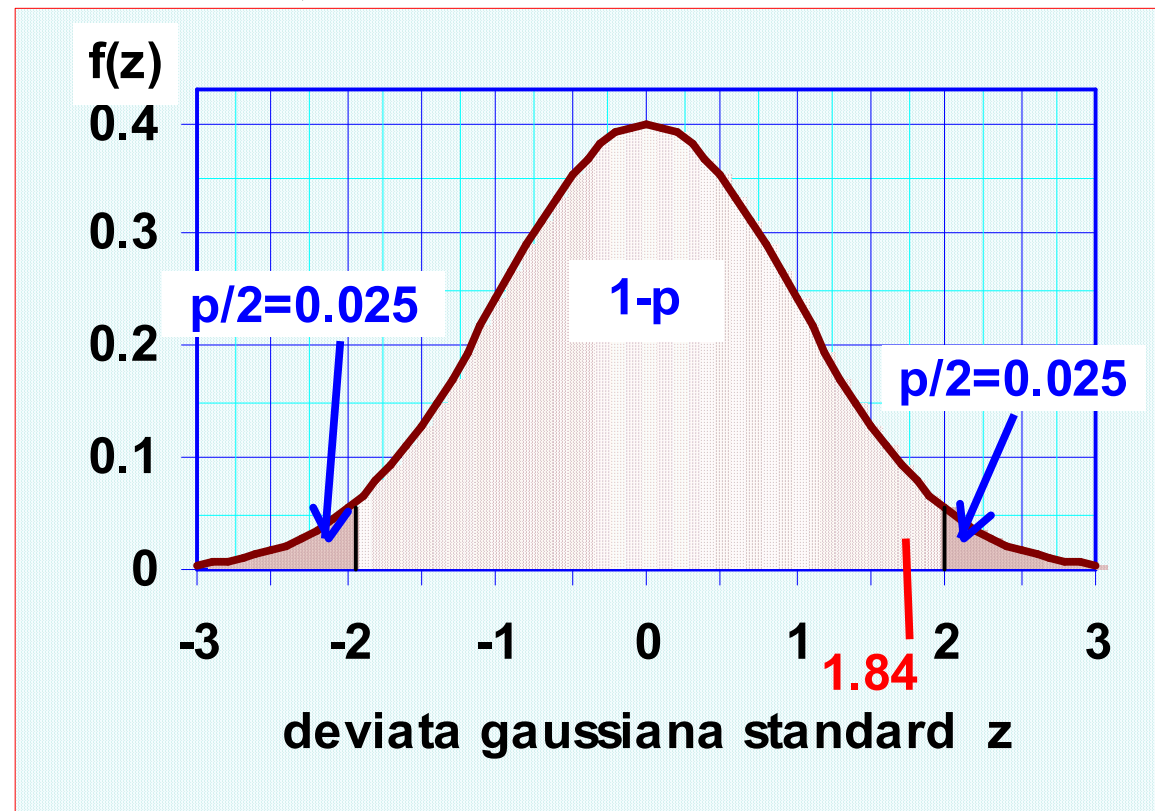
$$H_1: \pi_p \neq \pi_f \text{ oppure } \pi_p - \pi_f \neq 0$$

$$z = \frac{(0.0984 - 0.0813) - 0}{\sqrt{0.0899(1 - 0.0899) \left( \frac{1}{1900} + \frac{1}{1906} \right)}} = 1.84$$

Il livello critico del test a 2 code con  $\alpha=0.05$  è 1.96

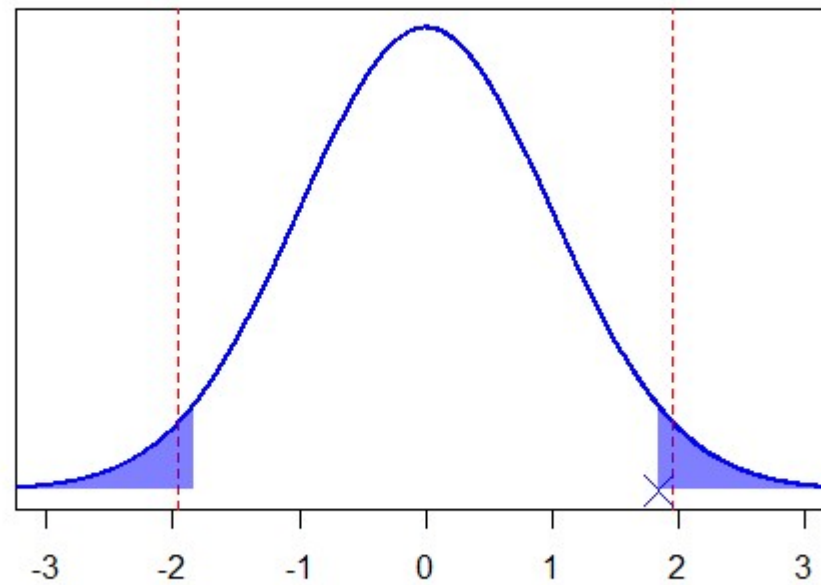
$p > 0.05$

Il farmaco **non produce** un numero significativamente diverso di eventi



# P-value

- $P(Z > 1.84) = 0.03288$ 
  - $\rightarrow$  p-value =  $0.03288 * 2 = 0.06576$



# Intervallo di confidenza

Definiamo:

$\delta = \pi_p - \pi_f =$  differenza “vera”

$\mathbf{d} = p_p - p_f =$  diff. osservata  $= 0.0984 - 0.0813 = 0.0171$

$\mathbf{SE(d)} =$  errore standard di  $\mathbf{d} = \sqrt{\frac{0.0984(1-0.0984)}{1900} + \frac{0.0813(1-0.0813)}{1900}} = 0.00927$

$d - z_{1-\alpha/2} \mathbf{SE(d)} \leq \delta \leq d + z_{1-\alpha/2} \mathbf{SE(d)}$

$0.0171 - 1.96 * 0.00927 \leq \delta \leq 0.0171 + 1.96 * 0.00927$

$-0.0011 \leq \delta \leq 0.0353$  cioè  $-0.1\% \leq \delta \leq 3.5\%$

N.B.  $\delta=0$  è incluso nell'intervallo di confidenza 95%  
( $p > 0.05$  nel test a due code)

L'intervallo di confidenza sposta l'attenzione  
dalla **SIGNIFICATIVITÀ STATISTICA**  
alla **RILEVANZA CLINICA**

IC permette di dire, con accettabili margini di confidenza, che **l'aumento degli eventi** attribuibile al farmaco **non è maggiore di 1 ogni mille pazienti** e che il **vantaggio** legato al farmaco potrebbe essere **la prevenzione di 3.5 eventi ogni 100 pazienti**

$$- 0.1\% \leq \delta \leq 3.5\%$$

# Dimensione del campione

**End-point binario:  
confronto di proporzioni**

## Tipi di studio

### Ipotesi di interesse e calcolo della dimensione

Differenza	$H_0: \pi_S - \pi_C = 0$ (o $\pi_S = \pi_C$ ) versus $H_1: \pi_S \neq \pi_C$ (o $ \pi_S - \pi_C  > \delta$ )
Superiorità	$H_0: \pi_S - \pi_C = 0$ (o $\pi_S = \pi_C$ ) versus $H_1: \pi_S - \pi_C > 0$ (o $\pi_S - \pi_C > \delta$ )

$\pi_S$  è la probabilità di successo nel trattamento sperimentale  
 $\pi_C$  è la probabilità di successo nel trattamento di controllo



## *Il significato di $\delta$*

- *La scelta di  $\delta$  è cruciale e influenza la dimensione del campione molto più di  $\alpha$  e  $\beta$*
- *$\delta$  deve avere senso dal punto di vista clinico:*
  - *Studi di differenza/ superiorità  $\Rightarrow$  minima differenza ritenuta rilevante*

## Studio di differenza (proporzioni)

Si definisce

$$H_0: \pi_S - \pi_C = 0$$

E si "scommette" su una specifica alternativa

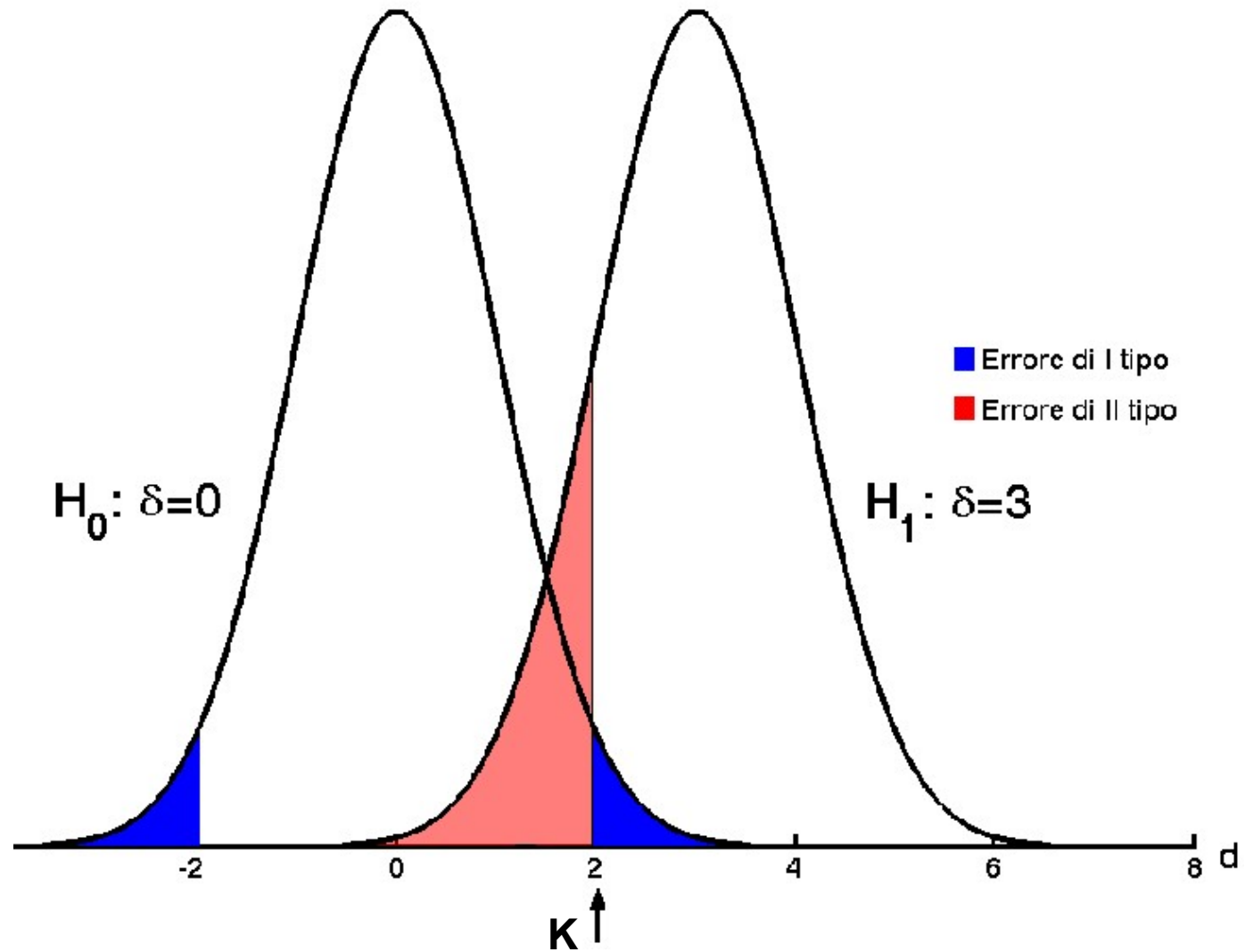
$$H_1: \pi_S - \pi_C = \delta$$

La stima campionaria sarà:  $d = p_S - p_C$

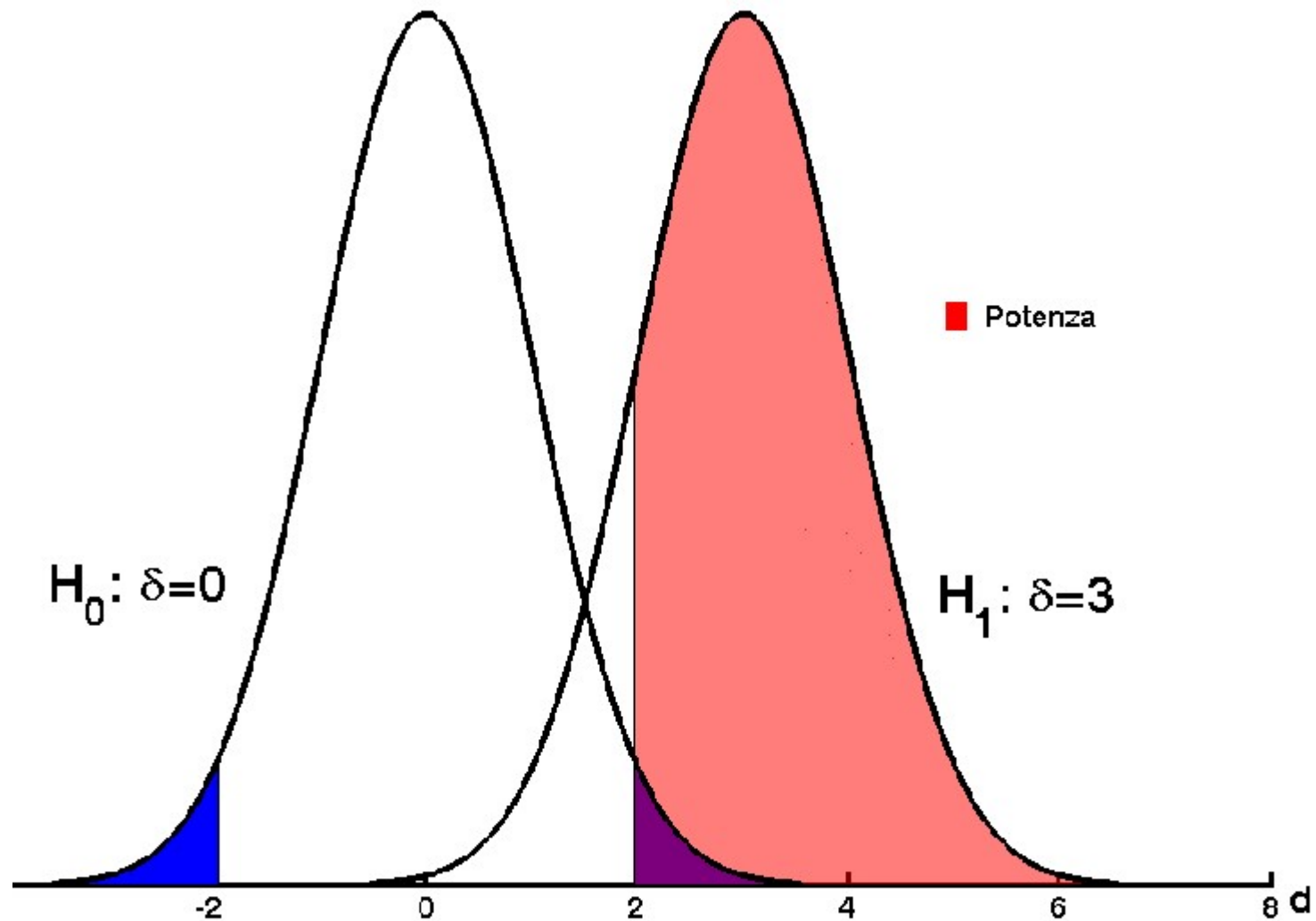
Per  $n \rightarrow \infty$  la sua distribuzione sarà

$$d \approx N\left(\delta, \frac{\pi_S(1-\pi_S)}{n_S} + \frac{\pi_C(1-\pi_C)}{n_C}\right)$$

## Le distribuzioni sotto $H_0$ e $H_1$ con $\alpha$ e $\beta$



## La potenza $1 - \beta$



Nel calcolo della dimensione campionaria, se  $n_S = n_C = n$

$\pi_0$  = probabilità di successo attesa sotto  $H_0$

$\pi_{S1}$  e  $\pi_{C1}$  = probabilità di successo attese  
sotto una specifica  $H_1$

Formula per il valore di  $n$  (numerosità in un gruppo)

$$n = \frac{\left[ \frac{z_{\alpha/2}}{2} \sqrt{2\pi_0(1-\pi_0)} + z_{\beta} \sqrt{\pi_{S1}(1-\pi_{S1}) + \pi_{C1}(1-\pi_{C1})} \right]^2}{(\pi_{S1} - \pi_{C1})^2}$$

dove  $z_{\alpha/2}$  e  $z_{\beta}$  sono 100(1- $\alpha$ /2) e (1- $\beta$ ) percentili di  $N(0,1)$

La formula si ottiene esprimendo il valore della soglia  $K$   
(valore critico) sotto  $H_0$  e  $H_1$

## Un'approssimazione molto utilizzata (Pocock 1983)

Approssima la  $var_{H0}$  con  $var_{H1}$

$$2 \pi_0(1 - \pi_0) \text{ diventa } \pi_{S1}(1 - \pi_{S1}) + \pi_{C1}(1 - \pi_{C1})$$

e quindi

$$n = \left( z_{\frac{\alpha}{2}} + z_{\beta} \right)^2 \frac{\pi_{S1}(1 - \pi_{S1}) + \pi_{C1}(1 - \pi_{C1})}{(\pi_{S1} - \pi_{C1})^2}$$

dove  $(z_{\alpha/2} + z_{\beta})^2 = f(\alpha, 1 - \beta)$

di converso, Lachin (1981) approssima  $var_{H1}$  con  $var_{H0}$

*Alcuni valori di  $f(\alpha, 1-\beta) = \left(\frac{z_\alpha}{2} + z_{1-\beta}\right)^2$*

	<i>1-<math>\beta</math></i>		
<i><math>\alpha</math></i>	0,80	0,90	0,95
0,10	6,2	8,6	10,8
0,05	7,9	10,5	13,0
0,01	11,7	14,9	17,8

Es.  $\alpha=0,10$  ,  $1 - \beta=0,8$   $f(\alpha, 1-\beta) = (z_{0,05} + z_{0,8})^2 = (1,64 + 0,84)^2 = 6,2$

Tabella relativa a  $z_{\alpha/2}$  per un test a due code e  $z_\beta$

*Essa può essere utilizzata per un test a una coda, a patto che si consideri il doppio del livello  $\alpha$  (se il test a una coda ha livello 0.025, si deve leggere la riga corrispondente a 0.05)*

## Un esempio (studio di superiorità)

*Un esempio pratico di calcolo nello studio di superiorità:*

- $\pi_c = 40\%$  è la percentuale di remissione nel gruppo di controllo  
 $\pi_s = 50\%$  è la percentuale di remissione **attesa** nel gruppo sperimentale  
 $\alpha = 0.05$  livello di significatività (una coda)  
 $1-\beta = 0.80$  potenza

*La formula approssimata (Pocock) fornisce il numero di soggetti necessari per ogni gruppo di trattamento:*

$$n = \frac{\pi_{s1}(100 - \pi_{s1}) + \pi_{c1}(100 - \pi_{c1})}{(\pi_{s1} - \pi_{c1})^2} \cdot f(\alpha, 1 - \beta)$$

nel nostro caso:

$$n = \frac{[50 \times 50 + 40 \times 60]}{(50 - 40)^2} \times 6.2 = 304$$

*Si dovranno quindi randomizzare 608 soggetti per poter mettere in evidenza una superiorità del 10% o più, con un livello di significatività del 5% ed una potenza dell'80%.*



## *Un altro esempio (studio di differenza)*

*Se l'ottica fosse quella di uno studio di differenza, con i medesimi valori di  $p_c$ ,  $p_s$ ,  $\alpha$  (due code) e  $\beta$  si otterrebbe:*

$$n = 387$$

*che è un numero lievemente superiore a quello ottenuto nell'ipotesi di studio di superiorità.*

*Perché?*

*Solo perché si dimezza il livello di significatività con cui si definisce la regione critica in ciascuna delle due direzioni.*

## Equivale a dire che.....

- Se pianifico studio di superiorità con  
Livello di sign. 0.025  
    chiedo 387 pts/gruppo
- Livello di sign. 0.05  
    chiedo 304 pts/gruppo

## *L'influenza su n della scelta dell'effetto e della potenza*

		Potenza ( $1-\beta$ )	
		0.80	0.90
$\pi_s$	45	1540	2048
% di remissioni nel gruppo sperimentale	50	387	515
	55	171	228

Baseline: 40%  
Alfa ( 2 code) =0.05