

Dimensionamento del Campione





Perché si calcola la numerosità campionaria?

Per assicurare che il numero di soggetti coinvolti nello studio clinico permetta di rispondere adeguatamente al quesito di interesse, considerando che:

- x uno studio di dimensioni limitate avrà una probabilità elevata di non riuscire a riconoscere un trattamento promettente.*
- x uno studio di dimensioni eccessive è economicamente oneroso e rischia di sottoporre un eccessivo numero di soggetti ad un trattamento non efficace.*

Come si calcola la numerosità campionaria?

1) Si sceglie il metodo da utilizzare per il calcolo della numerosità campionaria in funzione:

- *dell'obiettivo dello studio: studio di superiorità, di differenza (di non inferiorità, di equivalenza);*
- *del disegno dello studio: a gruppi paralleli, fattoriale, etc.. ;*
- *della natura dell'end-point primario: percentuale di "successo" (ad esempio di remissioni complete), probabilità di sopravvivenza libera da evento, livello medio di piastrine.*

2) Si determina il numero di soggetti da coinvolgere nello studio dopo aver definito il valore di alcuni parametri:

- *clinici (entità dell'effetto del trattamento, percentuale di drop-out, variabilità del fenomeno)*
- *statistici (tipo di test, livello di significatività, potenza)*

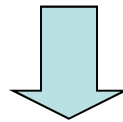
In generale

*Di quanti pazienti abbiamo bisogno
per avviare una sperimentazione clinica?*

Di un numero tale per cui si abbia:

- 1) una elevata probabilità di "vincere" quando si dovrebbe "vincere"*
- 2) una bassa probabilità di "vincere" quando non si dovrebbe "vincere"*

Cosa si intende per "vincere" ?



Dipende dal tipo di studio!

Studi di Superiorità - 1

Quesito dello studio:

Il trattamento sperimentale fornisce un risultato superiore a quello standard?

Definizione di superiorità:

Dipende dalla minima differenza di interesse clinico

Probabilità di "vincere"

La probabilità di dimostrare la superiorità del trattamento sperimentale

Esempio

Si confrontano due tipi di trattamenti di induzione che differiscono fra loro per combinazioni e dosi di farmaci. Si vuole verificare se il trattamento sperimentale garantisce il 50% (p_s) di remissione completa invece del 40% (p_c) usualmente ottenuto con il trattamento standard..

Beneficio del tratt. : sperim. - standard $p_s - p_c$	Probabilità di indicare la superiorità del trat. sperimentale
10	elevata
0	piccola

Studi di Superiorità - 2

Beneficio del tratt.: sperim. - standard $p_s - p_c$	Probabilità di indicare la superiorità del trat. sperimentale
10	elevata \longrightarrow potenza $(1-\beta)$
0	piccola \longrightarrow livello di significatività (α)

\Rightarrow La probabilità di ritenere superiore un trattamento quando lo è si chiama:
potenza = $(1 - \text{errore di secondo tipo})$
si vuole sia "elevata" e usualmente la si fissa pari a $(1-\beta) = 0.80$

\Rightarrow La probabilità di ritenere superiore un trattamento quando non lo è si chiama:
livello di significatività = $(\text{errore di primo tipo})$
si vuole sia "piccola" e usualmente la si fissa pari a $\alpha=0.05$

Studi di Differenza

Sono una importante generalizzazione degli studi di superiorità. In questo caso, nel confronto fra due strategie di trattamento, il ricercatore si chiede: vi è una differenza rilevante nell'effetto dei due trattamenti?

Il ricercatore si pone in una condizione di prudenza: ammette che possa verificarsi l'eventualità, inattesa ma pur possibile, che il trattamento sperimentale sia peggiore di quello standard.

A parità di minima differenza rilevante, di livelli di significatività e potenza, saranno necessari più soggetti rispetto ad uno studio di superiorità: il calcolo della dimensione campionaria corrisponde infatti a quello di uno studio di superiorità (o inferiorità) con un livello di significatività dimezzato (e ugual potenza).

Tipi di studio

Ipotesi di interesse e calcolo della dimensione

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

p_S è la probabilità di successo nel trattamento sperimentale
 p_C è la probabilità di successo nel trattamento di controllo

Il significato di δ

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

Parametri da specificare: superiorità

- *Livello di significatività (α)*
- *Potenza*
- *La differenza clinicamente rilevante tra i due gruppi*
- *Il rapporto di randomizzazione (randomizzazione bilanciata / sbilanciata)*
- *La deviazione standard attesa entro ciascun gruppo per variabili continue*

Criterio di decisione



	Se è vera H_0	Se è vera H_1
... e in base al campione <u>non rifiuto</u> H_0	Decisione giusta protezione: $(1-\alpha)$	Decisione sbagliata errore di tipo II: β
... e in base al campione <u>rifiuto</u> H_0 (preferisco H_1)	Decisione sbagliata errore di tipo I: α	Decisione giusta protezione: $(1-\beta)$

Studio di differenza (medie)

Uno studio randomizzato vuole valutare un nuovo (N) farmaco per la diminuzione della pressione sanguigna con uno già in uso (V). Sono note: la forma della distribuzione degli errori (Gaussiana) e l'entità della variabilità ($\sigma = 10$ mmHg)

Ci si chiede se "in media" il nuovo farmaco abbia un diverso effetto sulla pressione sistolica.

La minima differenza che si vuol metter in evidenza è $\delta=5$, ovvero la metà della deviazione standard.

Si calcola che occorre reclutare 240 soggetti:

- 120 soggetti assumono V per 6 mesi
- 120 soggetti assumono N per 6 mesi

Come?

Dimensione del campione per il confronto tra due medie

Nell'esempio, la dimensione campionaria $n = 120$ (per ciascun gruppo) era stata calcolata per garantire che :

- data una varibilità di entrambi i gruppi: $\sigma = 10$ mmHg
- fissata una prob. di errore di tipo I $\alpha = 0.01$

si potesse evidenziare una minima differenza

$$\delta = 5 \text{ mmHg}$$

- con una prob. di errore di tipo II $\beta = 0.10$

La potenza del test $1 - \beta = 0.90$

è la prob. di rifiutare H_0 quando è falsa

Dimensione del campione per il confronto tra due medie

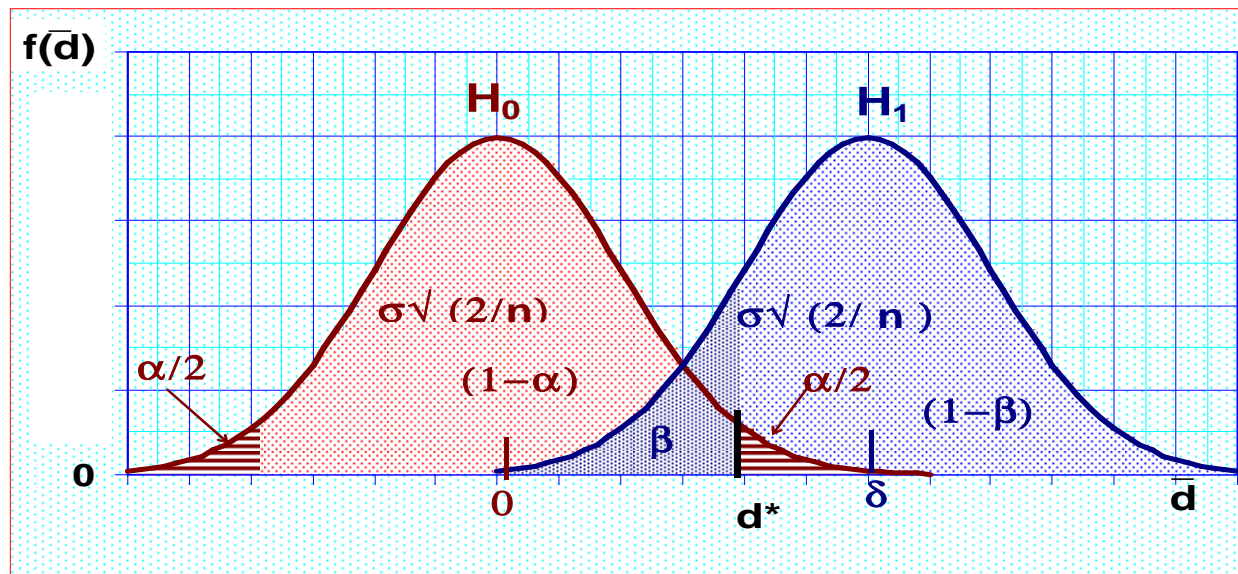
Due campioni di 120 soggetti garantiscono quanto segue:

- Non riconoscerò differenze di efficacia tra i farmaci **V** e **N** se $\mu_V = \mu_N$ con una probabilità del 99%.
- Riconoscerò differenze di efficacia pari o superiori al *minimo valore clinicamente rilevante* δ con una probabilità del 90%.

Dimensione del campione per il confronto tra due medie

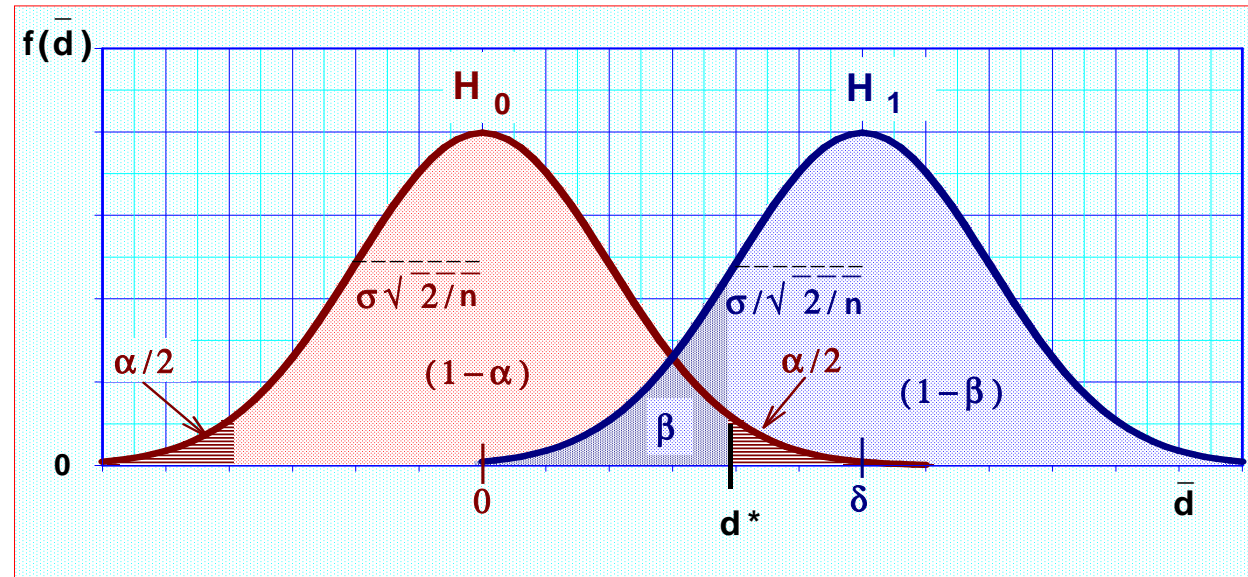
δ è nella scala originale, quindi consideriamo la distribuzione della differenza $(\bar{x}_N - \bar{x}_V)$ (non commisurata all'errore standard) che per 2 campioni di dimensione n che è Gaussiana con

- media δ e varianza $\sigma^2(2/n)$ sotto H_1
- media 0 e varianza $\sigma^2(2/n)$ sotto H_0



d^* è la soglia della zona di rifiuto nella scala originale

Dimensione del campione per il confronto tra due medie



d^* può essere espressa

sia rispetto alla media sotto H_0 $d^* = 0 + z_{\alpha/2} \cdot \sigma \sqrt{2/n}$

che rispetto alla media sotto H_1 $d^* = \delta - z_{\beta} \cdot \sigma \sqrt{2/n}$

Eguagliando le due espressioni si può ricavare la dimensione richiesta :

$$n = 2(z_{\alpha/2} + z_{\beta})^2 \cdot \frac{\sigma^2}{\delta^2}$$

Dimensione del campione per il confronto tra due medie

Nell'esempio:

- data una variabilità in entrambi i gruppi: $\sigma = 10$ mmHg
- fissato $\alpha = 0.01$

per evidenziare una minima differenza clinicamente rilevante $\delta = 5$ mmHg con una potenza $1 - \beta = 0.90$

si ottiene una dimensione del singolo campione pari a

$$n = 2 \cdot (z_{\alpha/2} + z_{\beta})^2 \cdot (\sigma / \delta)^2 = 2 \cdot (2.58 + 1.28)^2 \cdot (10 / 5)^2 = 119.2$$

Dimensione del campione per il confronto tra due medie

ESERCIZIO:

- data una variabilità in entrambi i gruppi: $\sigma = 10$ mmHg
- fissato $\alpha = 0.01$

per evidenziare una minima differenza clinicamente rilevante $\delta = 5$ mmHg **con una potenza $1 - \beta = 0.80$**

si ottiene una dimensione del singolo campione pari a

Studio di superiorità (proporzioni)

Si definisce

$$H_0: p_S - p_C = 0$$

E si "scommette" su una specifica alternativa

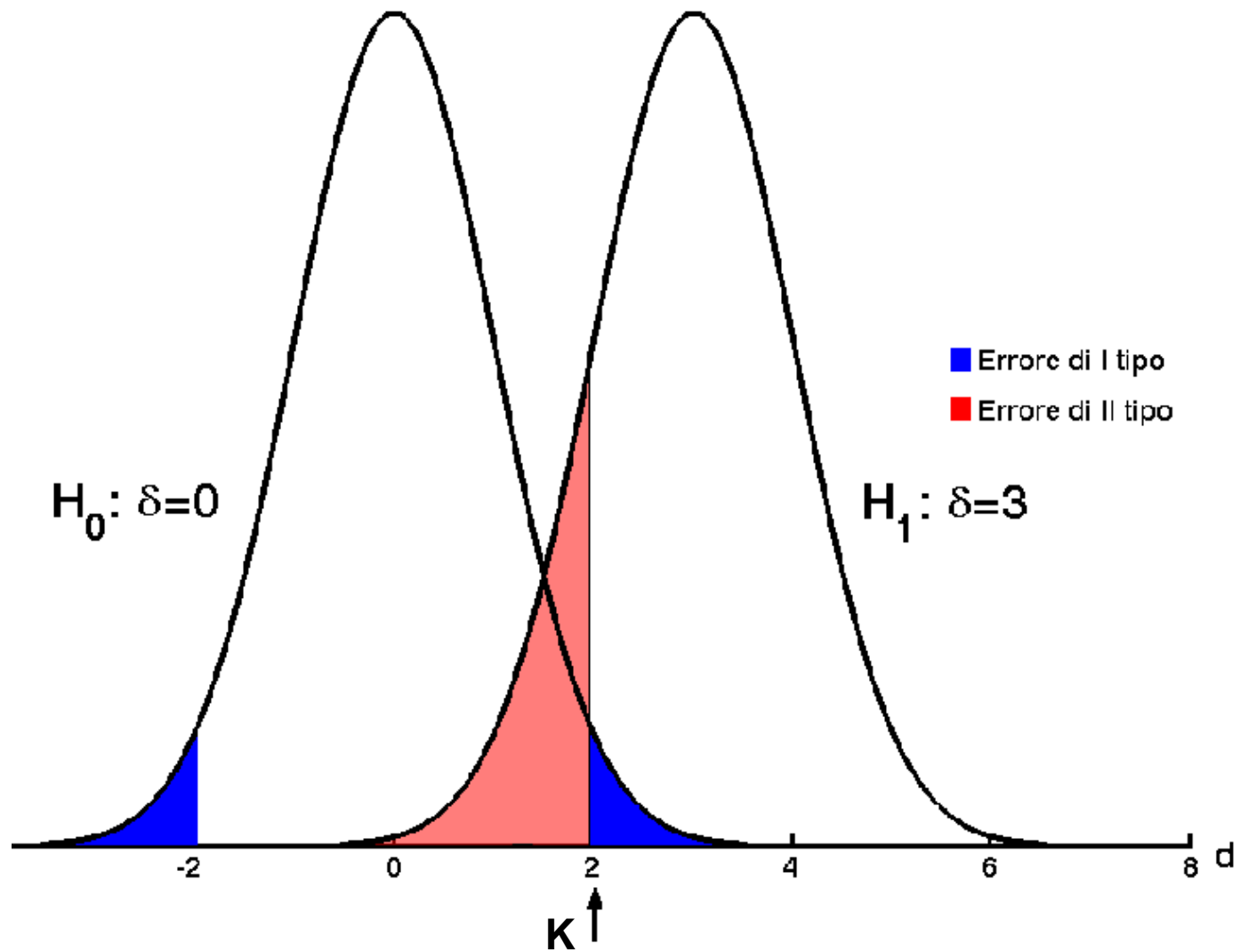
$$H_1: p_S - p_C = \delta$$

La stima campionaria sarà: $\hat{\delta} = \hat{p}_S - \hat{p}_C$

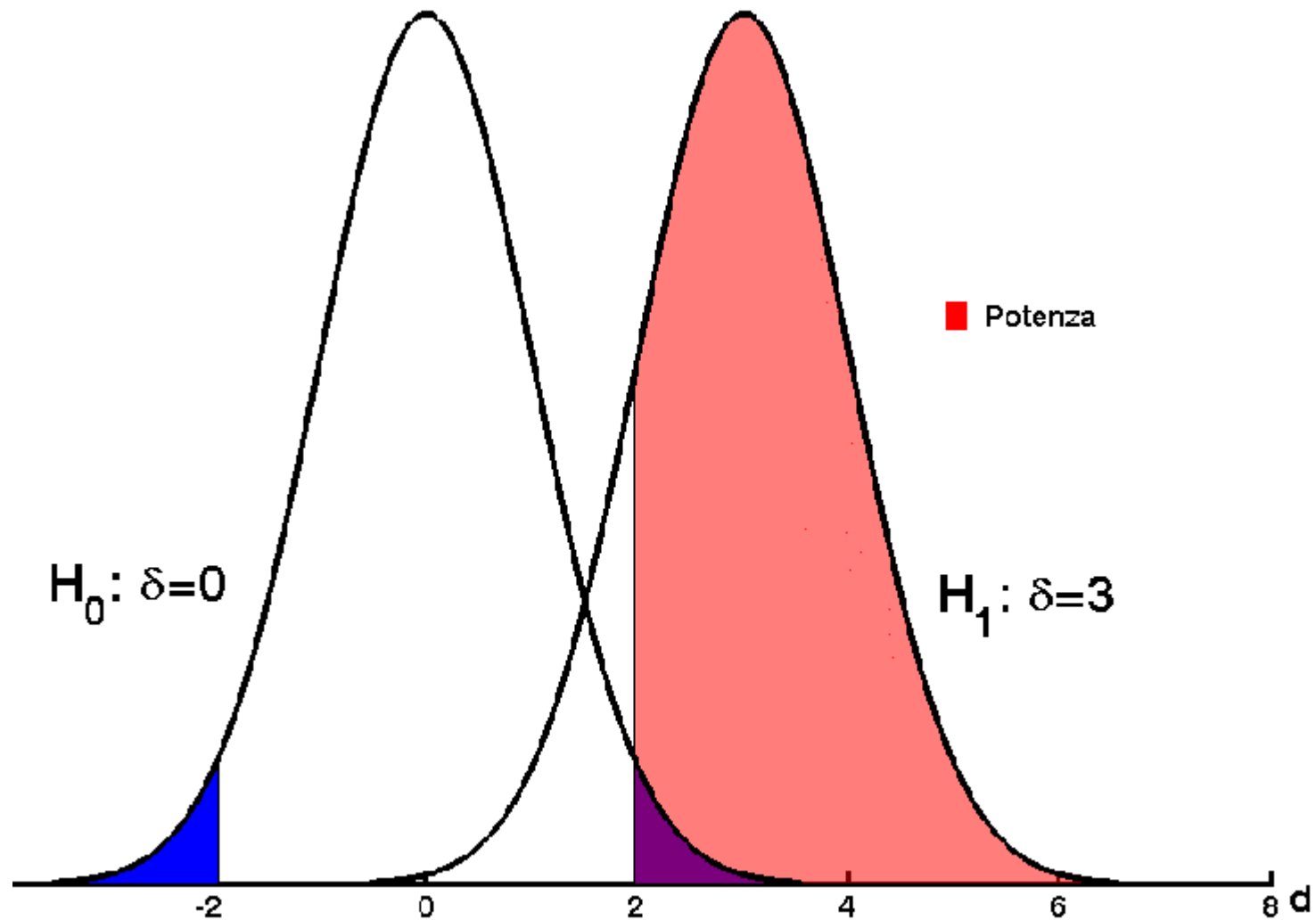
Per $n \rightarrow \infty$ la sua distribuzione sarà

$$\hat{\delta} \approx N\left(\delta, \frac{p_S(1-p_S)}{n_S} + \frac{p_C(1-p_C)}{n_C}\right)$$

Le distribuzioni sotto H_0 e H_1 con α e β



La potenza $1 - \beta$



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

p_0 = probabilità di successo attesa sotto H_0

p_{S1} e p_{C1} = probabilità di successo attese
sotto una specifica H_1

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

$$n = \frac{[z_{\alpha/2} \sqrt{2p_0(1-p_0)} + z_{\beta} \sqrt{p_{S1}(1-p_{S1}) + p_{C1}(1-p_{C1})}]^2}{(p_{S1} - p_{C1})^2}$$

dove $z_{\alpha/2}$ e z_{β} sono 100(1- α /2) e (1- β) 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}

$2p_0(1-p_0)$ diventa $p_{S1}(1-p_{S1}) + p_{C1}(1-p_{C1})$

e quindi

$$n = (z_{\alpha/2} + z_{\beta})^2 \frac{p_{S1}(1-p_{S1}) + p_{C1}(1-p_{C1})}{(p_{S1} - p_{C1})^2}$$

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

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

Alcuni valori di $f(\alpha, 1-\beta)$

	$1-\beta$		
α	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

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

Essa può essere utilizzata per un test a una coda, a patto che si consideri il doppio del livello α (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:

$p_c = 40\%$ è la percentuale di remissione nel gruppo di controllo

$p_s = 50\%$ 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:

$$n = \frac{[p_s \times (100 - p_s) + p_c \times (100 - p_c)]}{(p_s - p_c)^2} \times f(\alpha, 1 - \beta)$$

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 , α (due code) e β 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
p_s	45	1540	2048
% di remissioni nel gruppo sperimentale	50	387	515
	55	171	228

Baseline: 40%
Alfa (2 code) =0.05

Calcolo della dimensione per studi di sopravvivenza

In questi studi è importante non solo l'evento ma quando esso si verifica nel tempo

L'attenzione si sposta dalla probabilità di evento alla valutazione nel tempo del tasso di evento o della probabilità di sopravvivenza

Non tutti i soggetti presentano l'evento nel tempo di follow-up: dati censurati

Si determina il numero di EVENTI necessari e da questi si risale al numero di soggetti

La scelta dei parametri statistici

- Un livello a meno stringente \Rightarrow *rischio falsi positivi*
- Studi con bassa potenza \Rightarrow *rischio falsi negativi*

Il calcolo della dimensione campionaria esprime
una dicotomia tra risultati
POSITIVI/NEGATIVI
perché si riferisce ai
test statistici

STIMARE GLI EFFETTI
con
INTERVALLI DI CONFIDENZA
serve a dare una indicazione
anche sulla rilevanza della differenza

L'intervallo di confidenza (IC)

- Indicano un intervallo di valori all'interno del quale il ricercatore conclude, con una specificata probabilità, che sia compreso l'effetto vero del trattamento.
- Forniscono informazioni sull'entità della stima con un livello di "confidenza" prestabilito (tipicamente 95%).
- Maggiore è l'ampiezza dell'IC, minore è la precisione della stima.

Intervalli di confidenza e rilevanza clinica

Studio di superiorità

