# Confronto tra due medie, test, IC e dimensione del campione

## Varianza nota

## Parte I

# Esempio

Uno studio randomizzato vuole valutare un nuovo (N) farmaco per la diminuzione della pressione sanguigna con uno già in uso (V).
E' nota l'entità della variabilità ($\sigma$ = 10 mmHg)

Ci si chiede se il nuovo farmaco abbia un effetto diverso dal vecchio sulla pressione sistolica.

Si reclutano 240 soggetti con pressione alta che vengono randomizzati ai due trattamenti:

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

Dopo 6 mesi nei due gruppi si osservano le seguenti riduzioni di pressione (pressione alla randomizzazione-pressione a 6 mesi):

| | | |
|---|---|---|
| Farmaco V | $n_V$=120 | $\bar{x}_V$=11.9 |
| Farmaco N | $n_N$=118* | $\bar{x}_N$=15.8 |

*2 soggetti del gruppo N si trasferiscono durante il periodo dello studio

# Test per il confronto tra due medie

**Si vuole saggiare se i due gruppi hanno medie di riduzione di pressione diverse.**
**Ciò significa scegliere tra le due ipotesi:**

$$\begin{cases} H_0 : \mu_N = \mu_V \\ H_1 : \mu_N \neq \mu_V \end{cases}$$

$H_0 : \mu_N = \mu_V$ — Le medie vere dei due gruppi coincidono, quindi i farmaci hanno la stessa efficacia

$H_1 : \mu_N \neq \mu_V$ — Le medie vere dei due gruppi differiscono, quindi i farmaci non hanno la stessa efficacia

# Test per il confronto tra due medie

**Sotto l'ipotesi nulla di uguaglianza delle medie (assumendo che gli errori di misura siano indipendenti) si ha che:**

**la variabile "differenza tra due medie campionarie"**

$$(\bar{x}_N - \bar{x}_V)$$

**ha distribuzione Gaussiana con** ***media 0*** **e** ***varianza pari alla somma delle varianze*** **delle due medie campionarie**

**In sintesi, sotto $H_0$ :** $\bar{x}_V \sim N\left(\mu, \frac{\sigma^2}{n_V}\right)$ $\quad \bar{x}_N \sim N\left(\mu, \frac{\sigma^2}{n_N}\right)$

$$(\bar{x}_N - \bar{x}_V) \sim N\left(0, \sigma^2\left(\frac{1}{n_N} + \frac{1}{n_V}\right)\right)$$

# Test per il confronto tra due medie

**Si ricordi, infatti, che date due variabili casuali indipendenti**

$x_1$ con media $\mu_1$ e varianza $\sigma_1^2$

$x_2$ con media $\mu_2$ e varianza $\sigma_2^2$

la nuova variabile ottenuta come somma o differenza

$x_1 \pm x_2$

ha media $\mu_1 \pm \mu_2$ e varianza $\sigma_1^2 + \sigma_2^2$

# Test per il confronto tra due medie

Il concetto alla base della verifica di ipotesi è il seguente:

se i dati campionari forniscono due valori di $\bar{x}_V$ e $\bar{x}_N$ la cui differenza è "molto lontana" dal valore atteso zero

**allora si confuterà $H_0$**

**Per applicare il test ci riferiamo alla deviata gaussiana standardizzata che valuta la "lontananza" da zero commisurata all'errore standard**

$$z = \frac{(\bar{x}_N - \bar{x}_V) - 0}{\sqrt{\sigma^2 \left( \frac{1}{n_N} + \frac{1}{n_V} \right)}}$$

# Test per il confronto tra due medie

**Nell'esempio:**

$$z = \frac{\bar{x}_N - \bar{x}_V}{\sqrt{\sigma^2\left(\frac{1}{n_N} + \frac{1}{n_V}\right)}} = \frac{15.8 - 11.9}{\sqrt{10^2\left(\frac{1}{120} + \frac{1}{118}\right)}} = \frac{3.9}{1.2965} = 3.008$$

**Ammettendo una probabilità $\alpha$ di compiere un errore di tipo I (rifiutare $H_0$ quando essa è vera)**

si determina la zona di rifiuto dalla distribuzione Gaussiana standard, individuando la soglia z* che cumula a destra una probabilità pari a $\alpha/2$

se $|z| > z^*$ => si confuterà $H_0$

se $|z| <= z^*$ => non si confuterà $H_0$

# Test per il confronto tra due medie

$$z = \frac{\bar{x}_N - \bar{x}_V}{\sqrt{\sigma^2\left(\frac{1}{n_N} + \frac{1}{n_V}\right)}} = \frac{15.8 - 11.9}{\sqrt{10^2\left(\frac{1}{120} + \frac{1}{118}\right)}} = \frac{3.9}{1.2965} = 3.008$$

Scegliamo $\alpha$=0.01 da cui deriva che la soglia critica per un test a due code è z* = 2.58

quindi essendo 3.008 > z* => si rifiuta $H_0$

**Quindi i farmaci V ed N non hanno la medesima efficacia, poiché il farmaco N tende a ridurre maggiormente la pressione.**

**Consultando le tavole si ricava un p-value p ≈ 0.0027**

(0.00135*2)

# Significato di p=0.0027 (livello di probabilità sotto $H_0$)

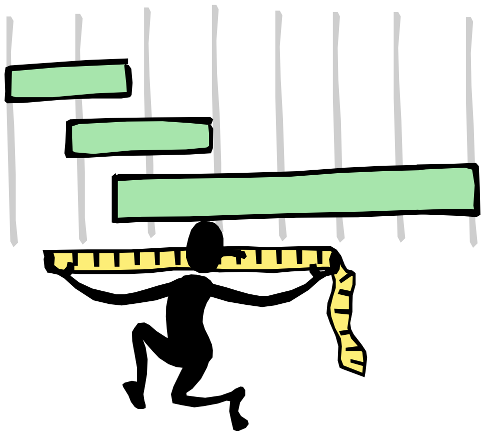
Se i due farmaci avessero la stessa efficacia, un risultato campionario uguale o più estremo (nella coda della distribuzione) di quello osservato nel campione ($\bar{d} = 3.9\ mmHg$) si verificherebbe 27 volte su 10000.

***Questo esprime la forza dell'evidenza contro l'ipotesi nulla***

- È **possibile** che, nonostante la differenza osservata di 3.9 mmHg, i due farmaci abbiano la stessa efficacia, tuttavia …

- … è **poco probabile** ($p \approx 0.0027$)

- È più plausibile che il farmaco N tenda a diminuire maggiormente la pressione



***l'esperimento suggerisce che i farmaci hanno diversa efficacia***

# IC per la differenza tra due medie

Se il fine è valutare di quanto differisca l'efficienza dei due farmaci:

1. si esegue il test sull'uguaglianza delle medie $\mu_V$ e $\mu_N$
2. si calcola anche l'intervallo di confidenza per la vera differenza tra le medie dei due metodi ($\delta=\mu_N-\mu_V$)

$$I.C._{(1-\alpha)} = (\bar{x}_N - \bar{x}_V) \pm z_{\alpha/2}\sqrt{\sigma^2\left(\frac{1}{n_N}+\frac{1}{n_V}\right)}$$

Nell'esempio, la confidenza corrispondente ad un errore di tipo I pari a $\alpha = 0.01$ è $1 - \alpha = 0.99$ :

$$I.C._{99\%} = 3.9 \pm 2.58 \cdot 1.2965 = [0.56;7.24]$$

$$\approx [1;7]$$

# IC per la differenza tra due medie

Si può quindi affermare che la vera differenza (δ) di efficacia tra i due metodi è un qualunque valore incluso tra 1 e 7 mmHg (in altri termini che la differenza non è minore di 1 ma non è maggiore di 7 mmHg).

Tuttavia questa affermazione non è certa, ma la probabilità (confidenza) che sia vera è del 99%

**Si noti che l'intervallo di confidenza non contiene lo 0:**

- **quindi la probabilità che la differenza sia nulla è inferiore all' 1%**
- **questo è coerente con l'esito del test di ipotesi**

# Confronto tra due medie, test, IC e dimensione del campione

## Varianza nota

## Parte II

# 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 (rifiutare $H_0$ vera) $\alpha = 0.01$

**si potesse evidenziare una minima differenza clinicamente rilevante $\delta = 5$ mmHg**

- con una prob. di errore di tipo II (non rifiutare $H_0$ falsa) $\beta = 0.10$

$1 - \beta = 0.90$ è la prob. di rifiutare $H_0$ quando è falsa

**$1 - \beta$ è la potenza del test**

## Rischio di errore di tipo I ($\alpha$)

**Probabilità di rifiutare $H_0$ quando è vera $H_0$**

es. si conclude che N è meglio (o peggio) di V quando in realtà non lo è (i trattamenti N e V non differiscono).

Di solito si fissa $\leq 5\%$

## Potenza del test ($1-\beta$):

**Probabilità di rifiutare $H_0$ quando è vera una specifica $H_1$**

es. si conclude che N differisce da V quando effettivamente N è meglio (o peggio) di V.

Di solito si fissa $\geq 80\%$

# Dimensione del campione per il confronto tra due medie

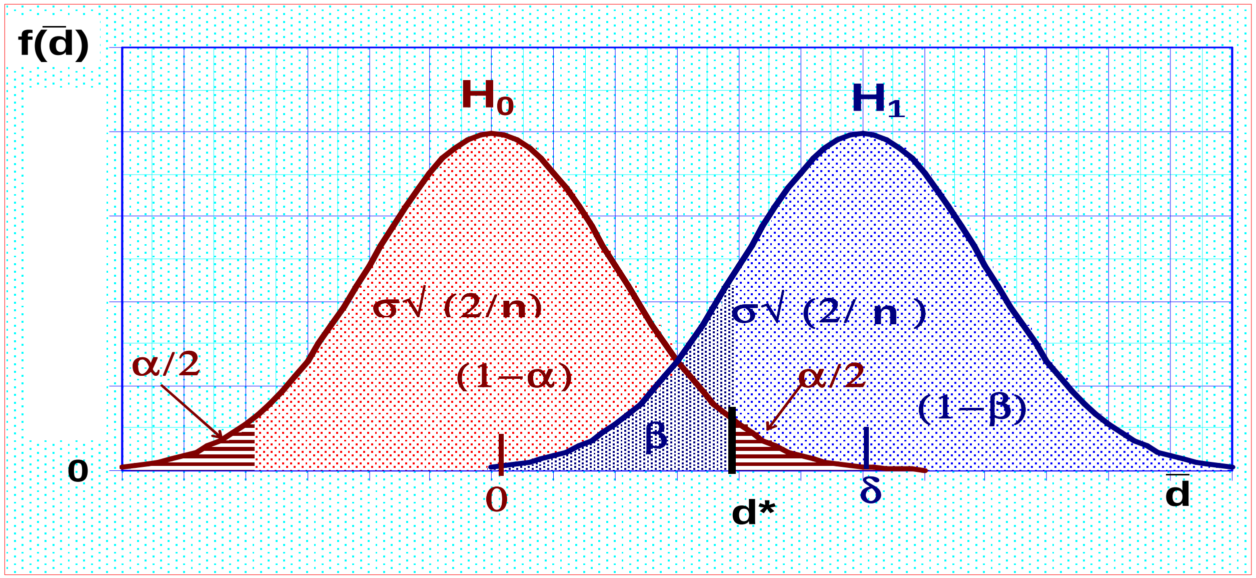
Due campioni di 120 soggetti garantiscono quanto segue:

- 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*** $\delta$ con una probabilità del 90%.

# Dimensione del campione per il confronto tra due medie

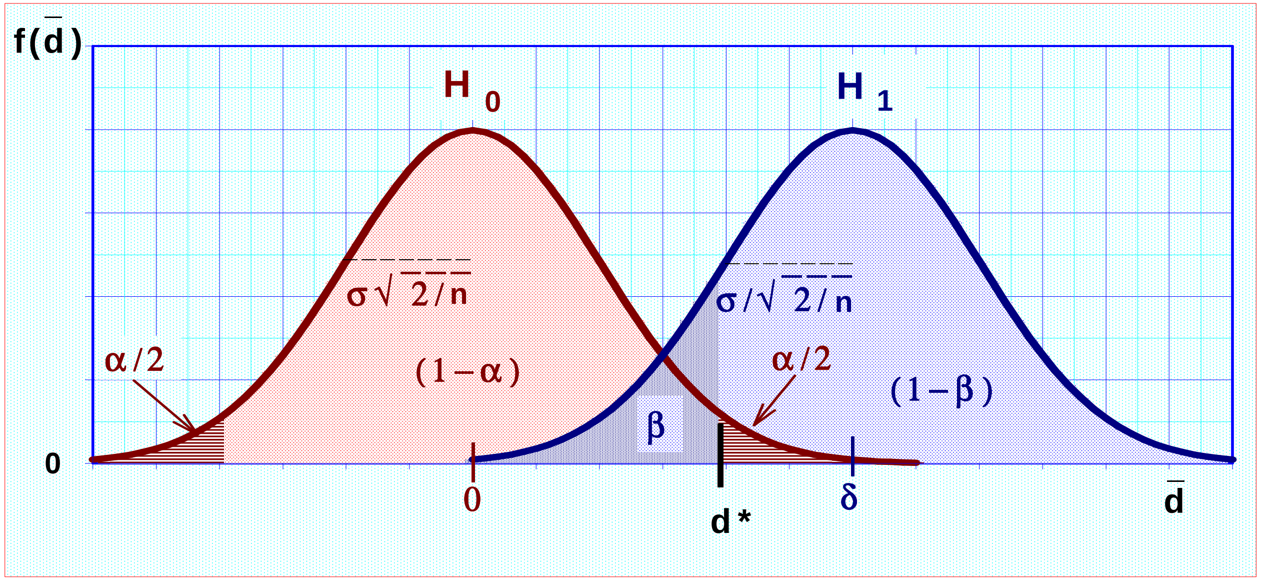
$\delta$ è nella scala originale, quindi consideriamo la distribuzione della differenza $(\bar{x}_N - \bar{x}_V)$ (non commisurata all'errore standard) che per 2 campioni di dimensione *n* che è Gaussiana con

- media $\delta$ 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 $\mathbf{H_0}$ $$d^* = 0 + z_{\alpha/2} \cdot \sigma\sqrt{2/n}$$

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

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

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

# Considerazioni

Nella programmazione dello studio illustrato nel nostro ***esempio***, ci eravamo proposti di seguire un totale di 240 soggetti (120 con V e 120 con N): questa ripartizione dei soggetti nei due gruppi è la più efficiente, nel senso che si ottiene l'errore standard minimo possibile per la differenza tra i farmaci N e V:

$$E.S.(\bar{x}_N - \bar{x}_V) = \sqrt{\sigma^2\left(\frac{1}{n_N} + \frac{1}{n_V}\right)} = \sqrt{10^2\left(\frac{1}{120} + \frac{1}{120}\right)} = 1.29$$

Se si fossero assegnati 60 soggetti al farmaco N e 180 al V, si sarebbe fatta la stessa quantità di lavoro, ma si sarebbe ottenuto un errore standard maggiore:

$$E.S.(\bar{x}_N - \bar{x}_V) = \sqrt{\sigma^2\left(\frac{1}{n_N} + \frac{1}{n_V}\right)} = \sqrt{10^2\left(\frac{1}{60} + \frac{1}{180}\right)} = 1.49$$

# Esercizio

Di due metodi analitici per la determinazione dell'uricemia, uno già in uso (V) e l'altro nuovo (N), sono note:

- la forma della distribuzione degli errori (Gaussiana)
- l'entità dell'imprecisione ($\sigma$ = 0.3 mg/dl)

Ci si chiede se "in media" i due metodi tendano a fornire il medesimo valore ed abbiano quindi la medesima "accuratezza"

1) Fissati $\alpha$ = 0.01 e $\beta$=0.1, per evidenziare una minima differenza tecnicamente rilevante $\delta$ = 0.45 mg/dl con una potenza 1 - $\beta$ =0.90, quanti soggetti dovrei reclutare?
2) Si eseguono su uno stesso "standard": 14 misure con il metodo V, 14 con il metodo N. A causa della scarsa dimestichezza con il metodo N, il tecnico di laboratorio ha eseguito correttamente solamente 12 misure, ottenendo i seguenti risultati: $\bar{x}_V = 5.21$ e $\bar{x}_N = 5.03$

Cosa posso concludere?

$$\begin{cases} H_0 : \mu_N = \mu_V \\ H_1 : \mu_N \neq \mu_V \end{cases}$$

1) Data una imprecisione di entrambi i metodi: σ = 0.30 mg/dl e fissato α = 0.01 per evidenziare una minima differenza tecnicamente rilevante δ = 0.45 mg/dl con una potenza 1 - β =0.90 si ottiene una dimensione del singolo campione pari a

$$n = 2 \cdot (z_{\alpha/2} + z_\beta)^2 \cdot (\sigma/\delta)^2 = 2 \cdot (2.58 + 1.28)^2 \cdot (0.30/0.45)^2 \cong 14$$

2a) test d'ipotesi:

$$z = \frac{\bar{x}_V - \bar{x}_N}{\sqrt{\sigma^2 \left( \frac{1}{n_V} + \frac{1}{n_N} \right)}} = \frac{5.21 - 5.03}{\sqrt{0.3^2 \left( \frac{1}{14} + \frac{1}{12} \right)}} = \frac{0.18}{0.118} = 1.53$$

La soglia critica per un test a due code è z* = 2.58 =>non si rifiuta $H_0$

Quindi i metodi V ed N non mostrano un diverso grado di accuratezza, la differenza osservata è probabilmente dovuta al caso.

2b) p-value: Consultando le tavole si ricava un p-value $p \approx 0.126$.

Se l'accuratezza fosse la stessa, un risultato campionario uguale o più estremo di quello osservato nel campione ($\bar{d} = 0.18$) si verificherebbe circa 13 volte su 100. E' possibile che l'accuratezza sia diversa ma i dati non lo evidenziano.

2c) Intervallo di confidenza $I.C._{99\%} = 0.18 \pm 2.58 \cdot 0.118 = [-0.12; 0.48]$

Si può quindi affermare che la vera differenza (δ) di accuratezza tra i due metodi è un qualunque valore incluso tra -0.12 e 0.48 (in altri termini che la differenza non è minore di -0.12 ma non è maggiore di 0.48 mg/dl).

Tuttavia questa affermazione non è certa, ma la probabilità (confidenza) che sia vera è del 99%.

Siccome una differenza di 0 (stessa accuratezza) non è esclusa dall'intervallo non possiamo affermare che l'accuratezza dei due metodi sia diversa.