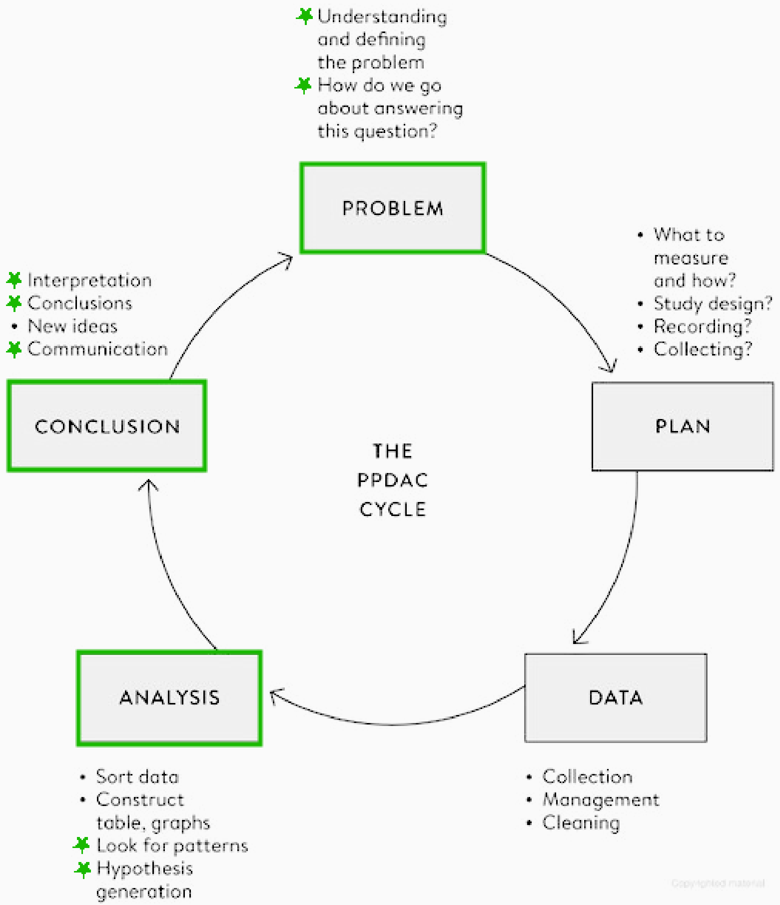
La statistica inferenziale

(Parte II: Test di ipotesi)

Obiettivi di apprendimento

- Formulare e verificare ipotesi
- Interpretare i P-value (e la loro relazione con i CI)
- Conoscere la differenza tra significatività statistica e clinica
- Saper distinguere tra errori del primo e del secondo tipo
- Interpretare la potenza di uno studio

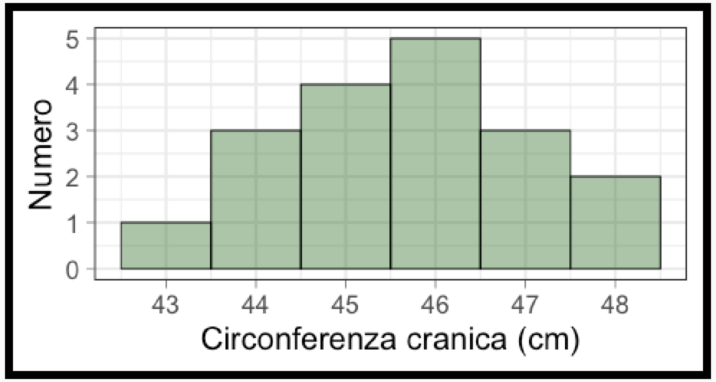
Le fasi della ricerca

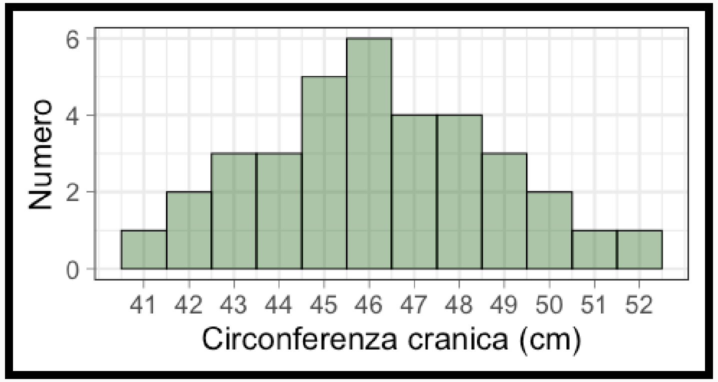


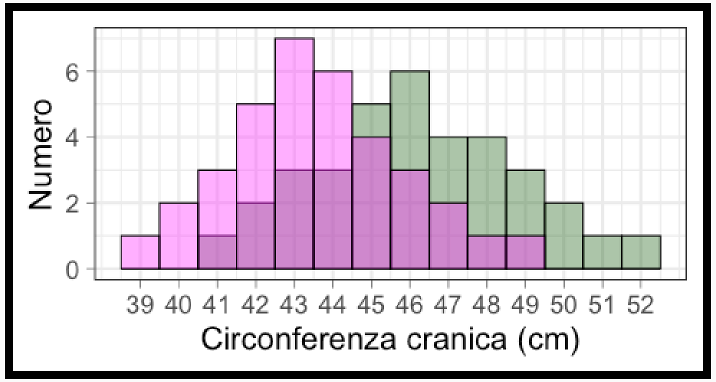
Spiegelhalter, D., *The Art of Statistics: Learning From Data*, Pelican, 2019

Perché c'è variabilità nelle osservazioni?

🎯 Vedremo come determinare se la variabilità è generata dalle condizioni sperimentali o se è generata da differenze individuali e/o da errori di misurazione







⚠️ Attenzione ⚠️

Questa parte continua a essere complessa, ma non demordete: siamo alla fine!

Cos'è un'ipotesi?

🎯 Una possibile spiegazione per un fenomeno, che non rappresenta la verità assoluta, ma una congettura provvisoria

Esempi di ipotesi

- L'esito di un trattamento è diverso nel gruppo di trattamento e di controllo
- La proporzione di un evento è diversa nel gruppo di trattamento e di controllo

Esercizio #1

❓ Supponiamo che la nostra ipotesi sia che tutti coloro che vivono più di 90 anni siano non fumatori

Per indagare questa ipotesi è più facile...

a) Dimostrare l'ipotesi trovando ogni singola persona di 90 anni o più e verificare che siano tutti non fumatori
b) Confutare l'ipotesi trovando una sola persona di 90 anni o più che sia un fumatore

01:00

Esercizio #1 -- Soluzione

❓ Supponiamo che la nostra ipotesi sia che tutti coloro che vivono più di 90 anni siano non fumatori

Per indagare questa ipotesi è più facile...

a) Dimostrare l'ipotesi trovando ogni singola persona di 90 anni o più e verificare che siano tutti non fumatori
b) Confutare l'ipotesi trovando una sola persona di 90 anni o più che sia un fumatore ✅

Il principio di falsificabilità e l'ipotesi nulla

- L'esito di un trattamento è ~~diverso~~ **uguale** nel gruppo di trattamento e di controllo
- La proporzione di un evento è ~~diversa~~ **uguale** nel gruppo di trattamento e di controllo

Esercizio #2

Objective To determine whether intravenous dexamethasone increases the number of ventilator-free days among patients with COVID-19–associated ARDS.

Design, Setting, and Participants Multicenter, randomized, open-label, clinical trial conducted in 41 intensive care units (ICUs) in Brazil. Patients with COVID-19 and moderate to severe ARDS, according to the Berlin definition, were enrolled from April 17 to June 23, 2020. Final follow-up was completed on July 21, 2020. The trial was stopped early following publication of a related study before reaching the planned sample size of 350 patients.

? Qual è l'ipotesi nulla di questo studio

a) Dexamethasone e standard care sono **più efficaci** che lo standard care da solo
b) Dexamethasone e standard care sono **meno efficaci** che lo standard care da solo
c) Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo
d) Dexamethasone e standard care **non** sono **tanto efficaci quanto** lo standard care da solo

02:00

Tomazini, B.M., *et al.*, "Effect of dexamethasone on days alive and ventilator-free in patients with moderate or severe acute respiratory distress syndrome and COVID-19: the CoDEX randomized clinical trial.", JAMA, 2020, doi:10.1001/jama.2020.17021

Esercizio #2 -- Soluzione

Objective To determine whether intravenous dexamethasone increases the number of ventilator-free days among patients with COVID-19–associated ARDS.

Design, Setting, and Participants Multicenter, randomized, open-label, clinical trial conducted in 41 intensive care units (ICUs) in Brazil. Patients with COVID-19 and moderate to severe ARDS, according to the Berlin definition, were enrolled from April 17 to June 23, 2020. Final follow-up was completed on July 21, 2020. The trial was stopped early following publication of a related study before reaching the planned sample size of 350 patients.

❓ Qual è l'ipotesi nulla di questo studio

a) Dexamethasone e standard care sono **più efficaci** che lo standard care da solo
b) Dexamethasone e standard care sono **meno efficaci** che lo standard care da solo
c) Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo ✅
d) Dexamethasone e standard care **non** sono **tanto efficaci quanto** lo standard care da solo

Esercizio #3

Objective To determine whether intravenous dexamethasone increases the number of ventilator-free days among patients with COVID-19–associated ARDS.

Design, Setting, and Participants Multicenter, randomized, open-label, clinical trial conducted in 41 intensive care units (ICUs) in Brazil. Patients with COVID-19 and moderate to severe ARDS, according to the Berlin definition, were enrolled from April 17 to June 23, 2020. Final follow-up was completed on July 21, 2020. The trial was stopped early following publication of a related study before reaching the planned sample size of 350 patients.

? Come formuleresti operativamente l’ipotesi nulla di questo studio?

a) $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0$
b) $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} \neq 0$
c) $\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} = 0$
d) $\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} \neq 0$

02:00

Esercizio #3 -- Soluzione

Objective To determine whether intravenous dexamethasone increases the number of ventilator-free days among patients with COVID-19–associated ARDS.

Design, Setting, and Participants Multicenter, randomized, open-label, clinical trial conducted in 41 intensive care units (ICUs) in Brazil. Patients with COVID-19 and moderate to severe ARDS, according to the Berlin definition, were enrolled from April 17 to June 23, 2020. Final follow-up was completed on July 21, 2020. The trial was stopped early following publication of a related study before reaching the planned sample size of 350 patients.

❓ Come formuleresti operativamente l'ipotesi nulla di questo studio?

a) $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0$ ✅
b) $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} \neq 0$
c) $\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} = 0$
d) $\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} \neq 0$

Formulare ipotesi

📌 Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo

$$\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0 \\ \rightarrow \text{Ipotesi nulla } (\mathcal{H}_0)$$

Formulare ipotesi

📌 Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo

$$\mu_\mathrm{c} - \mu_\mathrm{i} = 0 \quad \rightarrow \text{Ipotesi nulla } (\mathcal{H}_0)$$

$$\mu_\mathrm{c} - \mu_\mathrm{i} \neq 0 \quad \rightarrow \text{Ipotesi alternativa } (\mathcal{H}_1/\mathcal{H}_A)$$

Esercizio #4

❓ Il fatto che l'ipotesi nulla e l'ipotesi alternativa siano mutualmente esclusive significa che se l'ipotesi nulla è vera, l'ipotesi alternativa...

a) deve anche essere vera
b) può essere sia vera sia falsa
c) deve essere falsa
d) dipende dall'ipotesi alternativa

01:00

Esercizio #4 -- Soluzione

❓ Il fatto che l'ipotesi nulla e l'ipotesi alternativa siano mutualmente esclusive significa che se l'ipotesi nulla è vera, l'ipotesi alternativa...

a) deve anche essere vera
b) può essere sia vera sia falsa
c) deve essere falsa ✅
d) dipende dall'ipotesi alternativa

Verificare ipotesi

📌 Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo

Interventions Twenty mg of dexamethasone intravenously daily for 5 days, 10 mg of dexamethasone daily for 5 days or until ICU discharge, plus standard care (n=151) or standard care alone (n=148).

Results A total of 299 patients (mean [SD] age, 61 [14] years; 37% women) were enrolled and all completed follow-up. Patients randomized to the dexamethasone group had a mean 6.6 ventilator-free days (95% CI, 5.0-8.2) during the first 28 days vs 4.0 ventilator-free days (95% CI, 2.9-5.4) in the standard care group

$$
\begin{aligned}
n_\mathrm{i} &= 151, \quad \bar{x}_\mathrm{i} = 6.6, \quad s_\mathrm{i} = 10.0 \\
n_\mathrm{c} &= 148, \quad \bar{x}_\mathrm{c} = 4.0, \quad s_\mathrm{c} = 8.7
\end{aligned}
$$

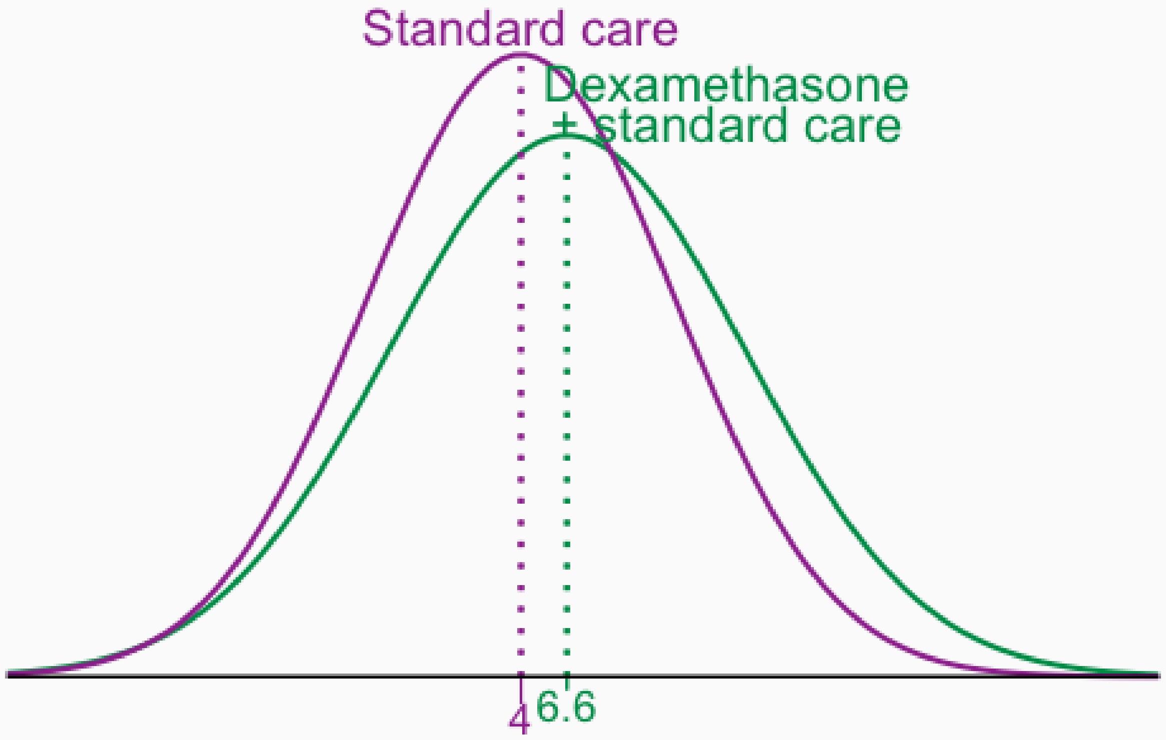
Tomazini B.M., *et al.*, *Effect of Dexamethasone on Days Alive and Ventilator-Free in Patients With Moderate or Severe Acute Respiratory Distress Syndrome and COVID-19*, JAMA, 2020

Verificare ipotesi

📌 Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo

$$n_{\mathrm{i}} = 151, \bar{x}_{\mathrm{i}} = 6.6, s_{\mathrm{i}} = 10.0$$
$$n_{\mathrm{c}} = 148, \bar{x}_{\mathrm{c}} = 4.0, s_{\mathrm{c}} = 8.7$$

$$\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0 \qquad \leftarrow$$
$$\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} = 6.6 - 4.0 = 2.6$$



Esercizio #5

❓ Anche se l'ipotesi nulla fosse vera, la differenza delle medie potrebbe non essere esattamente zero a causa...

a) dell'ipotesi nulla, che è stata formulata in modo impreciso
b) di differenze individuali
c) di errori di misurazione
d) se l'ipotesi nulla è vera, la differenza è sempre zero

01:00

Esercizio #5 -- Soluzione

❓ Anche se l'ipotesi nulla fosse vera, la differenza delle medie potrebbe non essere esattamente zero a causa...

a) dell'ipotesi nulla, che è stata formulata in modo impreciso
b) di differenze individuali ✅
c) di errori di misurazione ✅
d) se l'ipotesi nulla è vera, la differenza è sempre zero

Errore di campionamento

Verificare ipotesi

📌 Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo

$$n_{\mathrm{i}} = 151, \bar{x}_{\mathrm{i}} = 6.6, s_{\mathrm{i}} = 10.0$$
$$n_{\mathrm{c}} = 148, \bar{x}_{\mathrm{c}} = 4.0, s_{\mathrm{c}} = 8.7$$

$$\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0 \qquad \leftarrow$$
$$\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} = 6.6 - 4.0 = 2.6$$

❓ Qual è la probabilità di osservare una differenza di 2.6 giorni se $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0$?

Facciamo un paio di passi indietro

1. La Normale è definita dalla sua media e deviazione standard e corrisponde a una distribuzione di probabilità
   $$\rightarrow \quad \text{Area sottesa a } Z \equiv \text{probabilità } \mathcal{P}$$
2. Il teorema del limite centrale ci dice che le distribuzioni campionarie (incluso la differenza delle medie) tendono alla Normale

Facciamo un paio di passi indietro

1. La Normale è definita dalla sua media e deviazione standard e corrisponde a una distribuzione di probabilità
   $\rightarrow$ Area sottesa a $Z \equiv$ probabilità $\mathcal{P}$
2. Il teorema del limite centrale ci dice che le distribuzioni campionarie (incluso la differenza delle medie) tendono alla Normale

Per la differenza tra due medie

$\mathcal{N} = (\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$ con

$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \rightarrow$ standard error

Verificare ipotesi

📌 Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo

$$n_{\mathrm{i}} = 151, \bar{x}_{\mathrm{i}} = 6.6, s_{\mathrm{i}} = 10.0$$
$$n_{\mathrm{c}} = 148, \bar{x}_{\mathrm{c}} = 4.0, s_{\mathrm{c}} = 8.7$$

$$\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0 \qquad \leftarrow$$
$$\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} = 6.6 - 4.0 = 2.6$$

$$\mathcal{N} = \left(\mu_{\mathrm{c}} - \mu_{\mathrm{i}}, \frac{\sigma_c^2}{n_c} + \frac{\sigma_i^2}{n_i}\right) \rightarrow \mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0$$
$$\hat{\mathrm{SE}}^{(*)} = \sqrt{\frac{s_{\mathrm{c}}^2}{n_{\mathrm{c}}} + \frac{s_{\mathrm{i}}^2}{n_{\mathrm{i}}}} = 1.08$$

$^{(*)}$ In realtà non conosciamo σ ma solo s, quindi quello che usiamo è una t di Student con $(n_{\mathrm{c}} + n_{\mathrm{i}} - 2)$ gradi di libertà

Verificare ipotesi

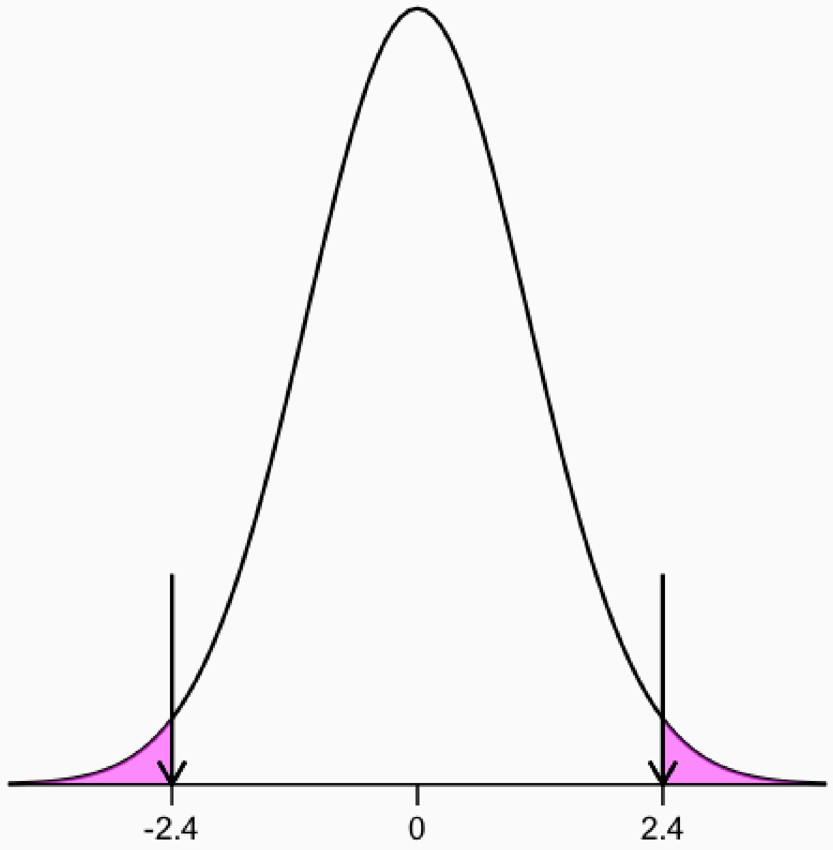
📌 Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo

$$\mu_\mathrm{c} - \mu_\mathrm{i} = 0$$
$$\hat{\mathrm{SE}} = 1.08$$

$$\bar{x}_\mathrm{c} - \bar{x}_\mathrm{i} = 6.6 - 4.0 = 2.6$$

❓ Qual è la probabilità di osservare una differenza di 2.6 giorni se $\mu_\mathrm{c} - \mu_\mathrm{i} = 0$?

$$t = \frac{(\bar{x}_\mathrm{c} - \bar{x}_\mathrm{i}) - (\mu_\mathrm{c} - \mu_\mathrm{i})}{\hat{SE}} = \frac{2.6 - 0}{1.08} = 2.4$$



Verificare ipotesi

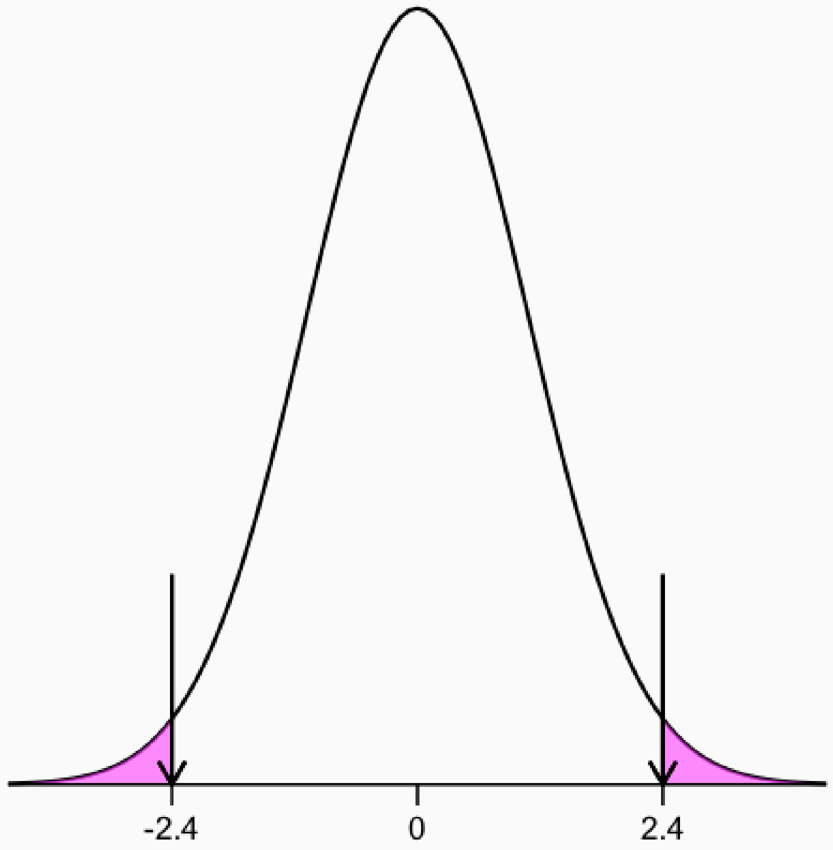
📌 Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo

$$\mu_{\text{c}} - \mu_{\text{i}} = 0$$
$$\hat{\text{SE}} = 1.08$$

$$\bar{x}_{\text{c}} - \bar{x}_{\text{i}} = 6.6 - 4.0 = 2.6$$

❓ Qual è la probabilità di osservare una differenza di 2.6 giorni se $\mu_{\text{c}} - \mu_{\text{i}} = 0$?

$$t = 2.4 \rightarrow \mathcal{P} = 2 \times 0.008 = 0.016$$



P-value

🎯 Il P-value misura la discrepanza tra i dati e $\mathcal{H}_0$ e corrisponde alla probabilità di ottenere un risultato tanto estremo quanto quello ottenuto se l'ipotesi nulla fosse vera.

P-value

🎯 Il P-value misura la discrepanza tra i dati e $\mathcal{H}_0$ e corrisponde alla probabilità di ottenere un risultato tanto estremo quanto quello ottenuto se l'ipotesi nulla fosse vera.

$$
\begin{aligned}
&\text{P-value} = 0.5 \rightarrow 50\% \rightarrow 1 \text{ campione su } 2 \\
&\text{P-value} = 0.1 \rightarrow 10\% \rightarrow 1 \text{ campione su } 10 \\
&\text{P-value} = 0.05 \rightarrow 5\% \rightarrow 1 \text{ campione su } 20 \\
&\text{P-value} = 0.01 \rightarrow 1\% \rightarrow 1 \text{ campione su } 100 \\
&\text{P-value} = 0.005 \rightarrow 0.5\% \rightarrow 1 \text{ campione su } 200
\end{aligned}
$$

P-value e significatività statistica

🎯 Il P-value misura la discrepanza tra i dati e $\mathcal{H}_0$ e corrisponde alla probabilità di ottenere un risultato tanto estremo quanto quello ottenuto se l'ipotesi nulla fosse vera.

Se il P-value è minore di una soglia critica (o livello di significatività) α, possiamo dire che il risultato è statisticamente significativo

$$\alpha = 0.05 \text{ oppure } 0.01$$

Perché?

- È conveniente decidere quando uno può dire "o c'è qualcosa nel trattamento o c'è una coincidenza che avviene più di 1 volta su 20"
- Il valore per cui $\alpha = 0.05$, o 1 in 20, è 1.96 o circa 2 [...] Deviazioni oltre due volte la deviazione standard sono considerate formalmente come significative
- Se 1/20 non è abbastanza, e se lo preferiamo, possiamo usare 1/50 (2%) o 1/100 (1%)

Esercizio #6

❓ Quando in uno studio si dice che il risultato è "statisticamente significativo" significa che...

a) l'ipotesi nulla è stata rifiutata
b) l'ipotesi nulla **non** è stata rifiutata
c) il risultato osservato è probabilmente dovuto a errori di campionamento
d) il risultato osservato **non** è probabilmente dovuto a errori di campionamento
e) il p-value è inferiore al livello di significatività α
f) il p-value è superiore al livello di significatività α

02:00

Esercizio #6 -- Soluzione

❓ Quando in uno studio si dice che il risultato è "statisticamente significativo" significa che...

a) l'ipotesi nulla è stata rifiutata ✅
b) l'ipotesi nulla **non** è stata rifiutata
c) il risultato osservato è probabilmente dovuto a errori di campionamento
d) il risultato osservato **non** è probabilmente dovuto a errori di campionamento ✅
e) il p-value è inferiore al livello di significatività α ✅
f) il p-value è superiore al livello di significatività α

Esercizio #7

❓ In uno studio clinico randomizzato, il P-value associato alla variabile "Sex" è pari a 0.48. Con un livello di significatività del 5%, ci sono differenze statisticamente significative nella distribuzione maschi/femmine nei due gruppi?

a) Sì, perché il P-value è maggiore del livello di significatività
b) Sì, perché il P-value è minore del livello di significatività
c) No, perché il P-value è maggiore del livello di significatività
d) No, perché il P-value è minore del livello di significatività

01:00

Esercizio #7 -- Soluzione

❓ In uno studio clinico randomizzato, il P-value associato alla variabile "Sex" è pari a 0.48. Con un livello di significatività del 5%, ci sono differenze statisticamente significative nella distribuzione maschi/femmine nei due gruppi?

a) Sì, perché il P-value è maggiore del livello di significatività
b) Sì, perché il P-value è minore del livello di significatività
c) No, perché il P-value è maggiore del livello di significatività ✅
d) No, perché il P-value è minore del livello di significatività

Esercizio #8

? Se **non** rifiuto l'ipotesi nulla significa che

a) ho provato che l'ipotesi nulla sia vera
b) ho provato che l'ipotesi nulla sia falsa
c) le mie osservazioni sono compatibili con l'ipotesi nulla
d) le mie osservazioni non sono compatibili con l'ipotesi nulla
e) dipende dalla domanda di ricerca
f) nessuno dei precedenti

01:00

Esercizio #8 -- Soluzione

❓ Se **non** rifiuto l'ipotesi nulla significa che

a) ho provato che l'ipotesi nulla sia vera
b) ho provato che l'ipotesi nulla sia falsa
c) le mie osservazioni sono compatibili con l'ipotesi nulla ✅
d) le mie osservazioni non sono compatibili con l'ipotesi nulla
e) dipende dalla domanda di ricerca
f) nessuno dei precedenti

Test di ipotesi, un passo alla volta

Test di ipotesi, un passo alla volta

1. Definisco la mia ipotesi nulla $(\mathcal{H}_0)$

    Dexamethasone e standard care sono **tanto efficaci quanto** lo standard care da solo
    $\mathcal{H}_0 : \mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0$

Test di ipotesi, un passo alla volta

1. Definisco la mia ipotesi nulla $(\mathcal{H}_0)$
2. Scelgo un test statistico che stimi qualcosa che, se abbastanza estremo, mi faccia dubitare di $\mathcal{H}_0$

t-test$^{(*)}$ della differenza di due medie campionarie

$^{(*)}$ Stiamo usando il t-test della differenza di due medie campionarie e non lo z-test perché non conosciamo la deviazione standard σ delle popolazioni (e stiamo usando s, quelle dei campioni).

Test di ipotesi, un passo alla volta

1. Definisco la mia ipotesi nulla $(\mathcal{H}_0)$
2. Scelgo un test statistico che stimi qualcosa che, se abbastanza estremo, mi faccia dubitare di $\mathcal{H}_0$
3. Genero la distribuzione campionaria del test scelto, assumendo $\mathcal{H}_0$ vera

$$\mathcal{N} = (\mu_{\mathrm{c}} - \mu_{\mathrm{i}}, \mathrm{SE}), \text{ con } \mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0 \text{ e } \hat{\mathrm{SE}} = \sqrt{\frac{s_{\mathrm{c}}^2}{n_{\mathrm{c}}} + \frac{s_{\mathrm{i}}^2}{n_{\mathrm{i}}}}$$

Test di ipotesi, un passo alla volta

1. Definisco la mia ipotesi nulla $(\mathcal{H}_0)$
2. Scelgo un test statistico che stimi qualcosa che, se abbastanza estremo, mi faccia dubitare di $\mathcal{H}_0$
3. Genero la distribuzione campionaria del test scelto, assumendo $\mathcal{H}_0$ vera
4. Verifico se la statistica osservata si trovi sulla coda di questa distribuzione e assegno una probabilità (P-value) a questo evento

$$\mathcal{P} = 2 \times 0.0082 = 0.0164$$

Test di ipotesi, un passo alla volta

1. Definisco la mia ipotesi nulla $(\mathcal{H}_0)$
2. Scelgo un test statistico che stimi qualcosa che, se abbastanza estremo, mi faccia dubitare di $\mathcal{H}_0$
3. Genero la distribuzione campionaria del test scelto, assumendo $\mathcal{H}_0$ vera
4. Verifico se la statistica osservata si trovi sulla coda di questa distribuzione e assegno una probabilità (P-value) a questo evento
5. Dichiaro il risultato come statisticamente significativo se il P-value è inferiore a una soglia critica α

$$\text{P-value} = 2 \times 0.0082 = 0.0164 < \alpha = 0.05 \quad \rightarrow \quad \text{rifiuto } \mathcal{H}_0$$

Comunicare il risultato

📌 Dexamethasone e standard care non sono tanto efficaci quanto lo standard care da solo. Osserviamo una differenza statisticamente significativa di 2.6 giorni tra i due trattamenti ($P = 0.016$).

Qual è l'incertezza di questa stima?

Esercizio #9

? Calcoliamo il 95% CI?

$$\bar{x}_{\text{c}} - \bar{x}_{\text{i}} = 2.6 \quad \rightarrow \quad \mu_{\text{c}} - \mu_{\text{i}} = 2.6$$

$$\hat{\text{SE}} = \sqrt{\frac{s_{\text{c}}^2}{n_{\text{c}}} + \frac{s_{\text{i}}^2}{n_{\text{i}}}} = 1.08$$

05:00

Esercizio #9 -- Soluzione

❓ Calcoliamo il 95% CI?

$$\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} = 2.6 \quad \rightarrow \quad \mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 2.6$$

$$\hat{\mathrm{SE}} = \sqrt{\frac{s_{\mathrm{c}}^2}{n_{\mathrm{c}}} + \frac{s_{\mathrm{i}}^2}{n_{\mathrm{i}}}} = 1.08$$

$$95\%\ \mathrm{ME} = 2 \times \hat{\mathrm{SE}} = 2 \times 1.08 = 2.16$$

Esercizio #9 -- Soluzione

? Calcoliamo il 95% CI?

$$\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} = 2.6 \quad \rightarrow \quad \mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 2.6$$

$$\hat{\mathrm{SE}} = \sqrt{\frac{s_{\mathrm{c}}^2}{n_{\mathrm{c}}} + \frac{s_{\mathrm{i}}^2}{n_{\mathrm{i}}}} = 1.08$$

$$95\%\ \mathrm{ME} = 2 \times \hat{\mathrm{SE}} = 2 \times 1.08 = 2.16$$

$$95\%\ \mathrm{CI} = (\bar{x}_{\mathrm{i}} - \bar{x}_{\mathrm{c}}) - 95\%\ \mathrm{ME}\ ;\ (\bar{x}_{\mathrm{i}} - \bar{x}_{\mathrm{c}}) + 95\%\ \mathrm{ME} =$$
$$= (2.6 - 2.16\ ;\ 2.6 + 2.16) = (0.44\ ;\ 4.78)$$

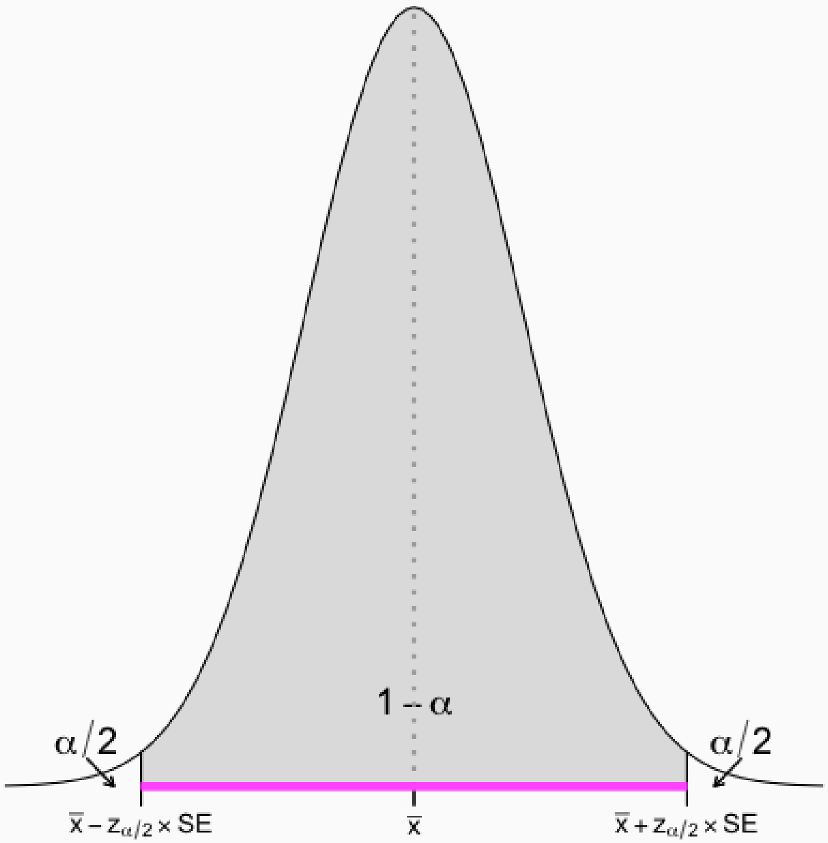
Comunicare il risultato

📌 Dexamethasone e standard care non sono tanto efficaci quanto lo standard care da solo. Osserviamo una differenza statisticamente significativa di 2.6 giorni (95% CI = 0.44 ; 4.78) tra i due trattamenti (P = 0.016).

Test di ipotesi & intervallo di confidenza

🎯 L'intervallo di confidenza del 95% è l'insieme delle ipotesi nulle che non sono rifiutate con $\alpha = 0.05$

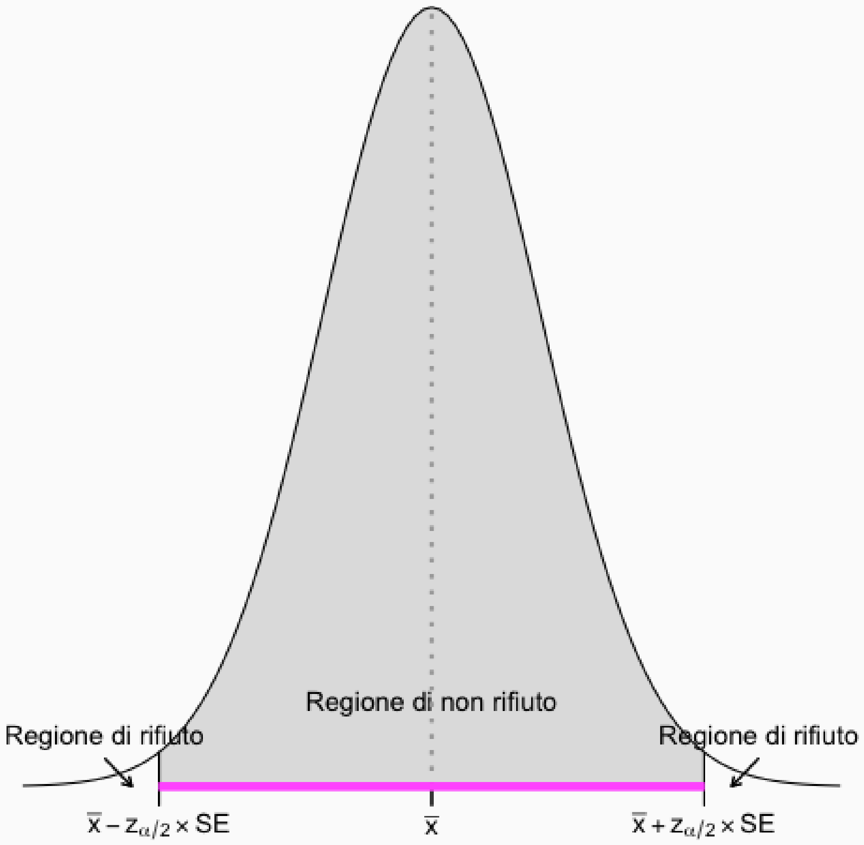
Livello di confidenza	α	$\alpha/2$
95%	5%	2.5%



Test di ipotesi & intervallo di confidenza

🎯 L'intervallo di confidenza del 95% è l'insieme delle ipotesi nulle che non sono rifiutate con $\alpha = 0.05$

In un test a due code, $\mathrm{P} < 0.05$ se il 95% CI non include l'ipotesi nulla (solitamente zero)



Test di ipotesi & intervallo di confidenza

🎯 L'intervallo di confidenza del 95% è l'insieme delle ipotesi nulle che non sono rifiutate con $\alpha = 0.05$

In un test a due code, $\mathrm{P} < 0.05$ se il 95% CI non include l'ipotesi nulla (solitamente zero)

RESULTS

Of the 355 children and adolescents who underwent screening, 290 were enrolled. A total of 146 participants were assigned to the oxytocin group and 144 to the placebo group; 139 and 138 participants, respectively, completed both the baseline and at least one postbaseline ABC-mSW assessments and were included in the modified intention-to-treat analyses. The least-squares mean change from baseline in the ABC-mSW score (primary outcome) was −3.7 in the oxytocin group and −3.5 in the placebo group (least-squares mean difference, −0.2; 95% confidence interval, −1.5 to 1.0; P=0.61). Secondary outcomes generally did not differ between the trial groups. The incidence and severity of adverse events were similar in the two groups.

Sikich, L. *et al.*, *Intranasal Oxytocin in Children and Adolescents with Autism Spectrum Disorder*, NEJM, 2021

Significatività statistica e significatività clinica

📌 Dexamethasone e standard care non sono tanto efficaci quanto lo standard care da solo. Osserviamo una differenza statisticamente significativa di 2.6 giorni (95% CI = 0.44 ; 4.78) tra i due trattamenti (P = 0.016).

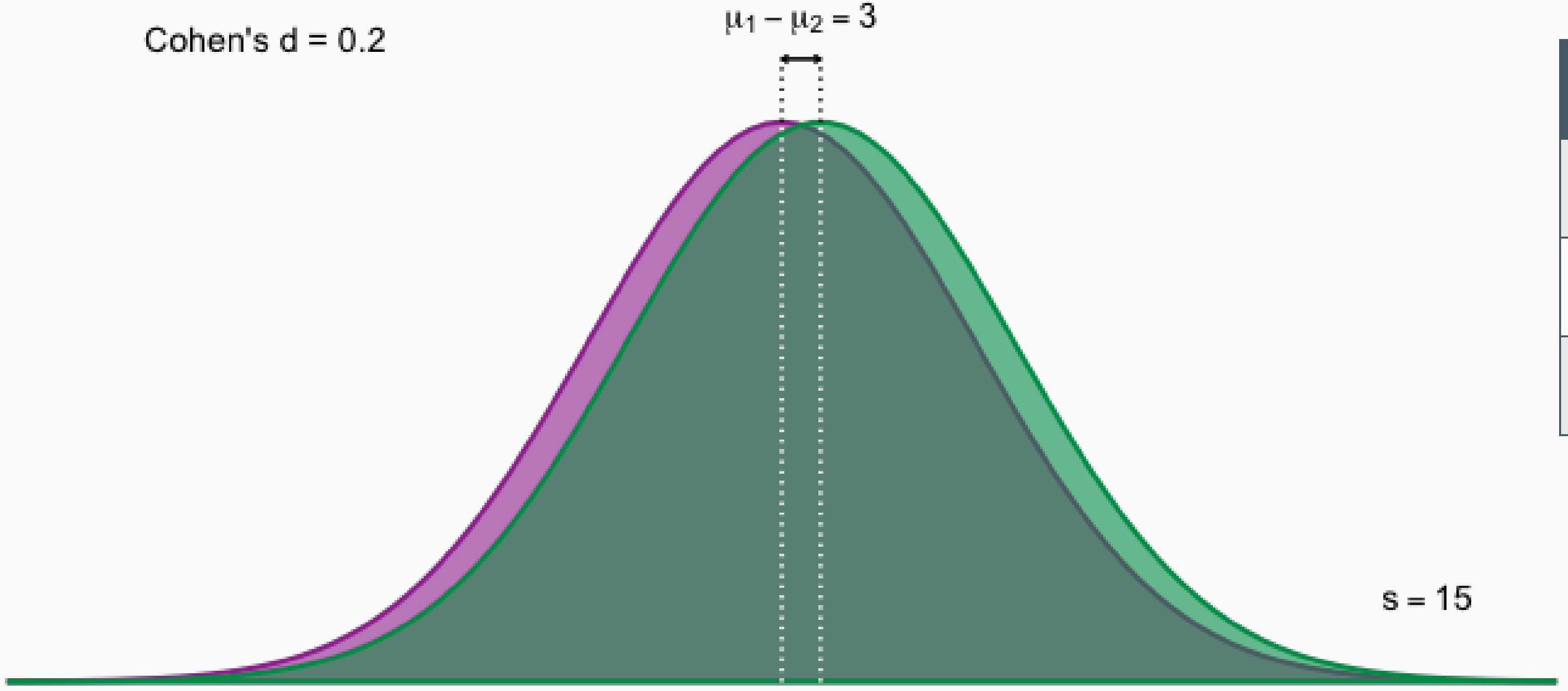
Qual è la significatività clinica del trattamento, tenendo conto che la differenza nella popolazione potrebbe essere di soli di 0.44 giorni?

La dimensione dell'effetto

📌 d di Cohen $= \left|\frac{\bar{x}_\mathrm{c} - \bar{x}_\mathrm{i}}{s_p}\right|$ con $s_p = \sqrt{\frac{(n_i - 1) \times s_i + (n_c - 1) \times s_c}{(n_i - 1) + (n_c - 1)}}$

La dimensione dell'effetto

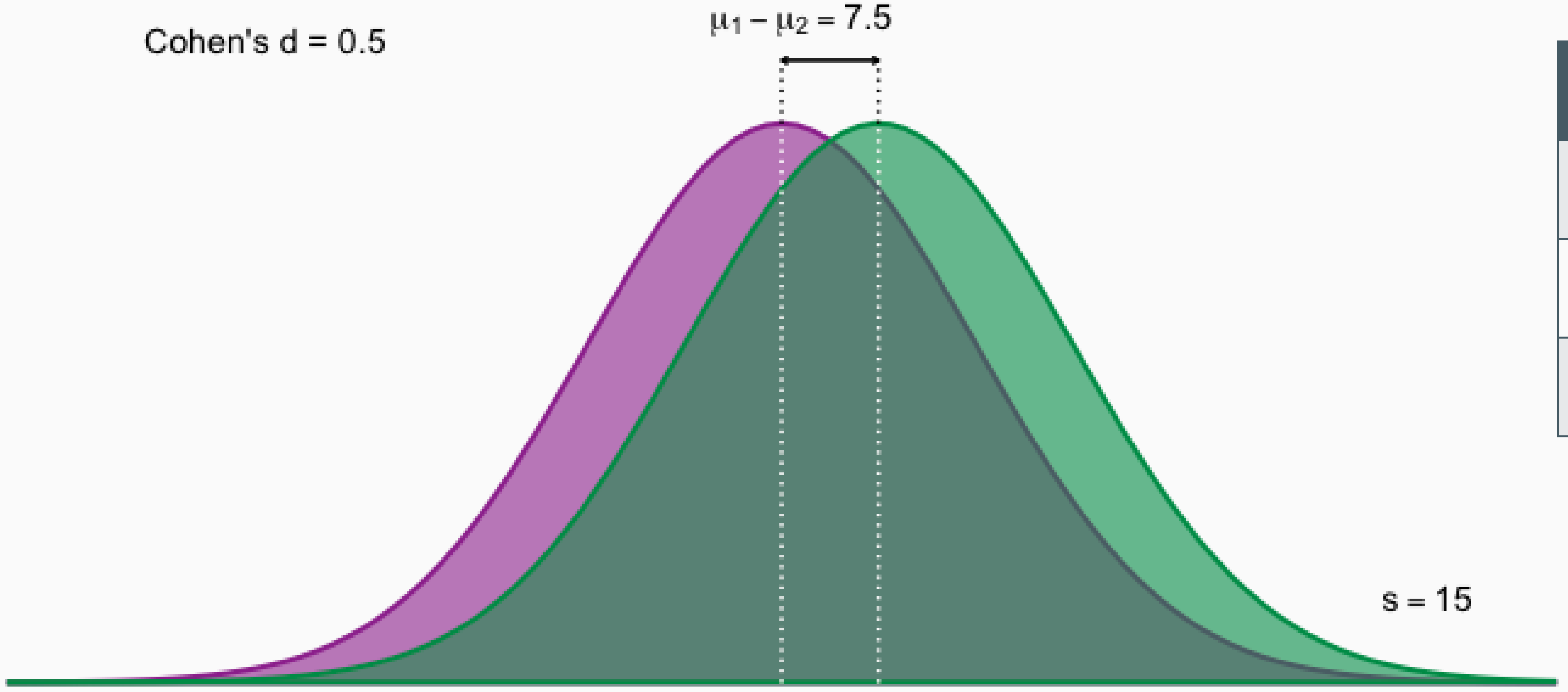
📌 $d \text{ di Cohen} = \left|\frac{\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}}}{s_p}\right|$ con $s_p = \sqrt{\frac{(n_i - 1) \times s_i + (n_c - 1) \times s_c}{(n_i - 1) + (n_c - 1)}}$



d	Interpretazione
0.2	Piccolo

La dimensione dell'effetto

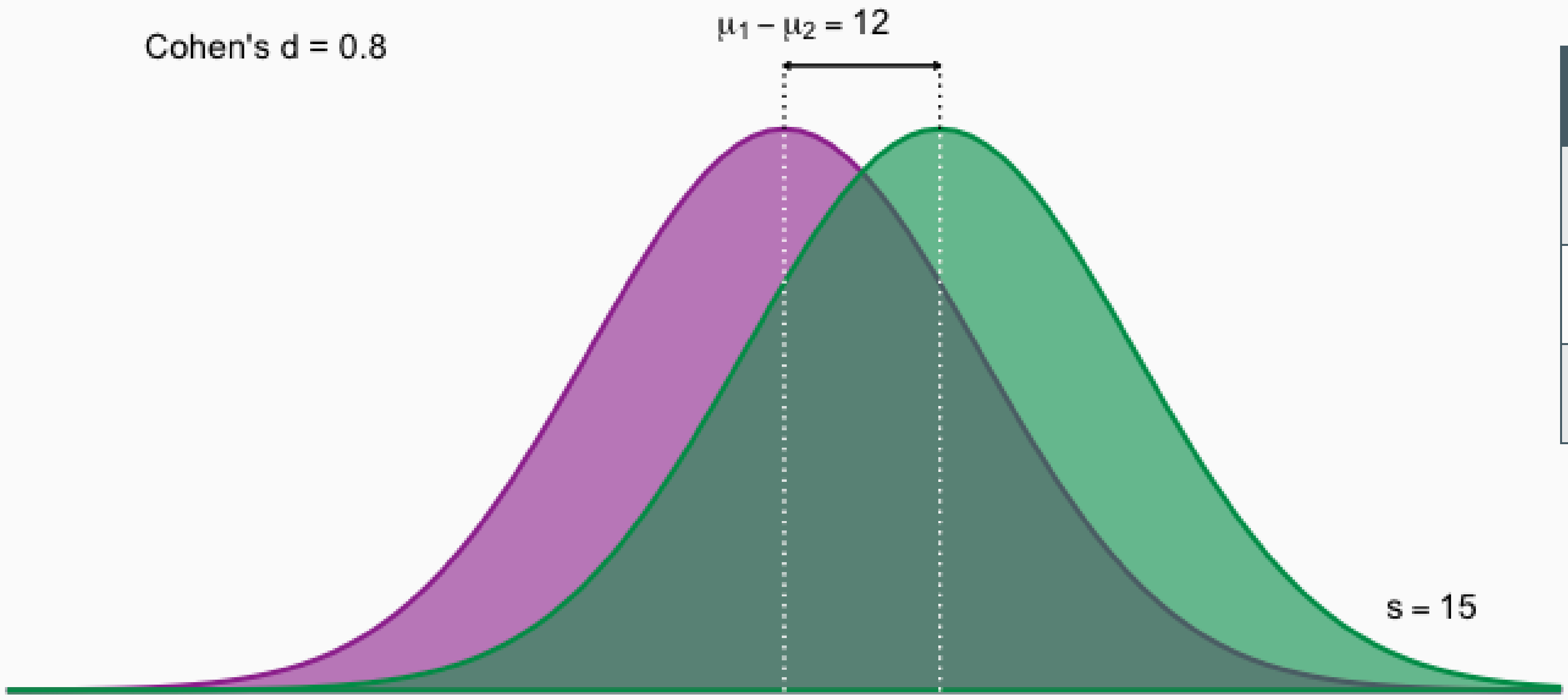
📌 d di Cohen $= \left|\frac{\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}}}{s_p}\right|$ con $s_p = \sqrt{\frac{(n_i-1)\times s_i + (n_c-1)\times s_c}{(n_i-1)+(n_c-1)}}$



d	Interpretazione
0.2	Piccolo
0.5	Medio

La dimensione dell'effetto

📌 $d \text{ di Cohen} = \left|\frac{\bar{x}_c - \bar{x}_i}{s_p}\right|$ con $s_p = \sqrt{\frac{(n_i-1)\times s_i + (n_c-1)\times s_c}{(n_i-1)+(n_c-1)}}$



d	Interpretazione
0.2	Piccolo
0.5	Medio
0.8	Grande

Esercizio #10

❓ Se in uno studio osservo $d = 0.65$, la dimensione dell'effetto è...

a) Piccola
b) Medio-piccola
c) Media
d) Medio-grande
e) Grande

01:00

Esercizio #10 -- Soluzione

❓ Se in uno studio osservo $d = 0.65$, la dimensione dell'effetto è...

a) Piccola
b) Medio-piccola
c) Media
d) Medio-grande ✅
e) Grande

La dimensione dell'effetto

📌 Dexamethasone e standard care **non sono tanto efficaci quanto** lo standard care da solo. Osserviamo una differenza statisticamente significativa di 2.6 giorni (95% CI = 0.44 ; 4.78) tra i due trattamenti (P = 0.016).

$$d \text{ di Cohen} = \left|\frac{\bar{x}_\mathrm{c} - \bar{x}_\mathrm{i}}{s_p}\right| \quad \text{con} \quad s_p = \sqrt{\frac{(n_i-1)\times s_i + (n_c-1)\times s_c}{(n_i-1)+(n_c-1)}}$$

$$s_p = \sqrt{\frac{(151-1)\times 10 + (148-1)\times 8.7}{(151-1)+(148-1)}} = 3$$

$$d = \frac{2.6}{3} = 0.85 \quad \rightarrow \quad \text{Grande}$$

Comunicare il risultato

📌 Dexamethasone e standard care **non sono tanto efficaci quanto** lo standard care da solo. Osserviamo una differenza statisticamente significativa di 2.6 giorni (95% CI = 0.44 ; 4.78) tra i due trattamenti (P = 0.016), con un effetto grande (d di Cohen = 0.85).

Comunicare (e interpretare) il risultato

- **Test di ipotesi:** è la procedura che valuta la probabilità che un’ipotesi sia supportata dai dati osservati
- **Intervallo di confidenza:** identifica l'incertezza di una statistica, l'intervallo di valori plausibili se il risultato in un campione fosse applicato all'intera popolazione
- **Dimensione dell'effetto:** la magnitudine dei risultati di uno studio, che determina se i risultati sono grandi abbastanza per essere utili nel mondo reale

Comunicare (e interpretare) il risultato

Scenario 1

- Test di ipotesi: P value $< \alpha$
- Intervallo di confidenza: molto stretto
- Dimensione dell'effetto: medio o grande

Abbiamo tre evidenze che supportano la significatività del risultato

Comunicare (e interpretare) il risultato

Scenario 2

- Test di ipotesi: P value $< \alpha$
- Intervallo di confidenza: molto stretto
- Dimensione dell'effetto: molto piccolo o piccolo

Abbiamo due evidenze che supportano la significatività statistica del risultato, ma la significatività clinica è minima

Comunicare (e interpretare) il risultato

Scenario 3

- Test di ipotesi: P value $> \alpha$
- Intervallo di confidenza: molto largo
- Dimensione dell'effetto: molto piccolo o piccolo

Abbiamo tre evidenze che supportano la mancanza di significatività del risultato

Comunicare (e interpretare) il risultato

Scenario 4

- Test di ipotesi: P value $> \alpha$
- Intervallo di confidenza: molto largo
- Dimensione dell'effetto: grande

Probabilmente abbiamo un campione troppo piccolo per decidere con sicurezza se rifiutare o meno l'ipotesi nulla

Esercizio #11

❓ In uno studio sono stati raccolti i voti di maturità di 1.5M di studenti, osservando che i ragazzi e le ragazze raggiungono risultati diversi ($P < 0.001$). Che informazione servirebbe per decidere che la differenza osservata sia effettivamente importante?

a) L'intervallo di confidenza della differenza delle medie
b) La dimensione dell'effetto
c) Nessuna, lo posso concludere dal P-value
c) Nessuna, lo posso concludere dalla dimensione campionaria

01:00

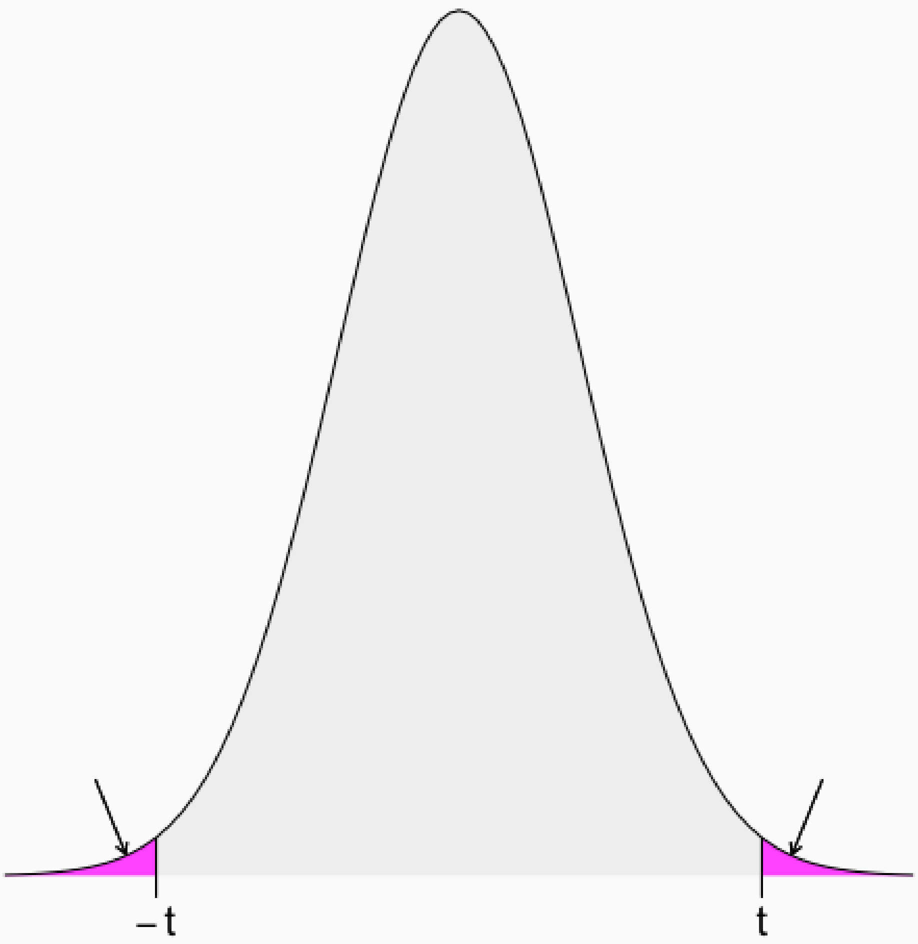
Esercizio #11 -- Soluzione

❓ In uno studio sono stati raccolti i voti di maturità di 1.5M di studenti, osservando che i ragazzi e le ragazze raggiungono risultati diversi ($P < 0.001$). Che informazione servirebbe per decidere che la differenza osservata sia effettivamente importante?

a) L'intervallo di confidenza della differenza delle medie
b) La dimensione dell'effetto ✅
c) Nessuna, lo posso concludere dal P-value
c) Nessuna, lo posso concludere dalla dimensione campionaria

Test a una e due code

🎯 $\mathcal{H}_1: \quad \mu_c - \mu_i \neq 0$
$\mathcal{H}_0: \quad \mu_c - \mu_i = 0$
$\rightarrow$ test a due code

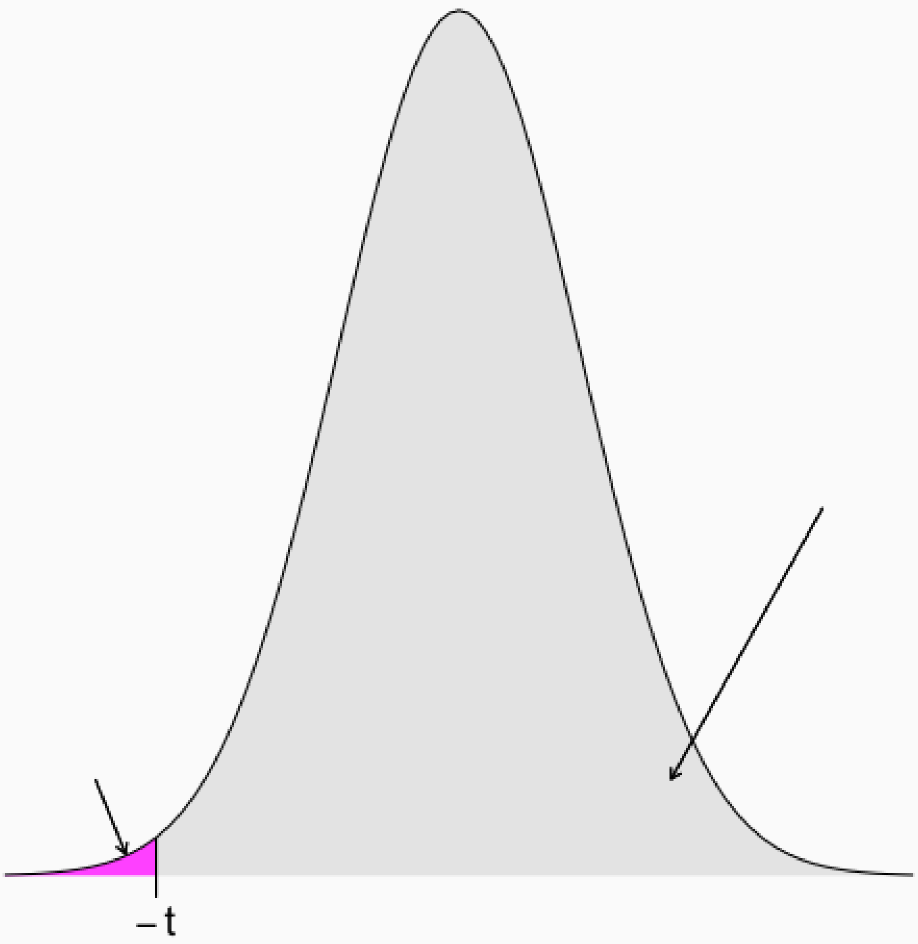


Test a una e due code

🎯 $\mathcal{H}_1$: $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} \neq 0$
$\mathcal{H}_0$: $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0$
$\rightarrow$ test a due code

$\mathcal{H}_1$: $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} < 0$
$\mathcal{H}_0$: $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} \geq 0$

$\rightarrow$ test a una coda



Test a una e due code

🎯 $\mathcal{H}_1$: $\mu_c - \mu_i \neq 0$

$\mathcal{H}_0$: $\mu_c - \mu_i = 0$

$\rightarrow$ test a due code

$\mathcal{H}_1$: $\mu_c - \mu_i < 0$

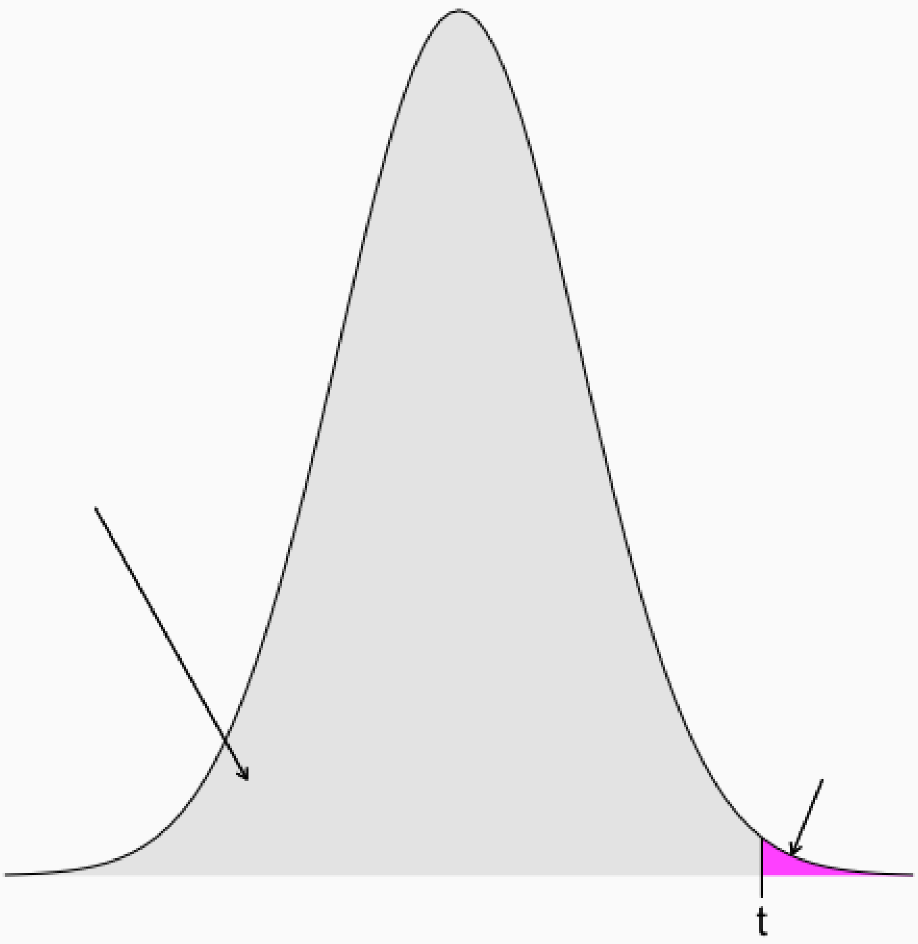
$\mathcal{H}_0$: $\mu_c - \mu_i \geq 0$

oppure

$\mathcal{H}_1$: $\mu_c - \mu_i > 0$

$\mathcal{H}_0$: $\mu_c - \mu_i \leq 0$

$\rightarrow$ test a una coda



Esercizio #12

❓ Un test a una coda deve essere usato quando l'ipotesi alternativa suppone che...

a) l'effetto del trattamento sia positivo
b) l'effetto del trattamento sia negativo
c) l'effetto del trattamento sia indifferentemente positivo o negativo
d) dipende dalla domanda di ricerca

01:00

Esercizio #12 -- Soluzione

❓ Un test a una coda deve essere usato quando l'ipotesi alternativa suppone che...

a) l'effetto del trattamento sia positivo ✅
b) l'effetto del trattamento sia negativo ✅
c) l'effetto del trattamento sia indifferentemente positivo o negativo
d) dipende dalla domanda di ricerca

Esercizio #13

? Quali delle seguenti formulazioni operative rappresenta l'ipotesi nulla in un test a una coda?

a) $\mu_1 \geq \mu_2$
b) $\mu_1 > \mu_2$
c) $\mu_1 \neq \mu_2$
d) nessuna delle precedenti

01:00

Esercizio #13 -- Soluzione

❓ Quali delle seguenti formulazioni operative rappresenta l'ipotesi nulla in un test a una coda?

a) $\mu_1 \geq \mu_2$ ✅
b) $\mu_1 > \mu_2$
c) $\mu_1 \neq \mu_2$
d) nessuna delle precedenti

t-test

📌 Per la differenza di medie:

$$\mathcal{N} = (\mu_\mathrm{c} - \mu_\mathrm{i}, \frac{\sigma_c^2}{n_c} + \frac{\sigma_i^2}{n_i})$$

$$\hat{\mathrm{SE}} = \sqrt{\frac{s_\mathrm{c}^2}{n_\mathrm{c}} + \frac{s_\mathrm{i}^2}{n_\mathrm{i}}}$$

📌 Per la differenza di proporzioni:

$$\mathcal{N} = (\pi_\mathrm{c} - \pi_\mathrm{i}, \frac{\pi_\mathrm{c} \times (1 - \pi_\mathrm{c})}{n_\mathrm{c}} + \frac{\pi_\mathrm{i} \times (1 - \pi_\mathrm{i})}{n_\mathrm{i}})$$

$$\hat{\mathrm{SE}} = \sqrt{\frac{\bar{p}_\mathrm{c} \times (1 - \bar{p}_\mathrm{c})}{n_\mathrm{c}} + \frac{\bar{p}_\mathrm{i} \times (1 - \bar{p}_\mathrm{i})}{n_\mathrm{i}}}$$

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è diverso rispetto ad altri ospedali britannici

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è diverso rispetto ad altri ospedali britannici

$\mathcal{H}_0$: Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

Nell'ospedale di Bristol, sono stati effettuati 143 interventi cardiochirugici e sono stati registrati 41 decessi (27.8%). Negli altri ospedali britannici, a fronte di 3176 interventi sono stati registrati 356 decessi (10.7%).

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

Nell'ospedale di Bristol, sono stati effettuati 143 interventi cardiochirugici e sono stati registrati 41 decessi (27.8%). Negli altri ospedali britannici, a fronte di 3176 interventi sono stati registrati 356 decessi (10.7%).

1. Definisco la mia ipotesi nulla $(\mathcal{H}_0)$

$$\mathcal{H}_0 : \pi_{\mathrm{B}} - \pi_{\mathrm{H}} = 0$$
$$\mathcal{H}_1 : \pi_{\mathrm{B}} - \pi_{\mathrm{H}} \neq 0$$

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

Nell'ospedale di Bristol, sono stati effettuati 143 interventi cardiochirugici e sono stati registrati 41 decessi (27.8%). Negli altri ospedali britannici, a fronte di 3176 interventi sono stati registrati 356 decessi (10.7%).

2. Scelgo un test statistico che stimi qualcosa che, se abbastanza estremo, mi faccia dubitare di $\mathcal{H}_0$

    Pearson's χ^2 test per dati categorici

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

Nell'ospedale di Bristol, sono stati effettuati 143 interventi cardiochirugici e sono stati registrati 41 decessi (27.8%). Negli altri ospedali britannici, a fronte di 3176 interventi sono stati registrati 356 decessi (10.7%).

3. Genero la distribuzione campionaria del test scelto, assumendo $\mathcal{H}_0$ vera

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

Nell'ospedale di Bristol, sono stati effettuati 143 interventi cardiochirugici e sono stati registrati 41 decessi (27.8%). Negli altri ospedali britannici, a fronte di 3176 interventi sono stati registrati 356 decessi (10.7%).

❓ Come completiamo questa tabella di contingenza?

Ospedale/Deceduti	Si	No	Totale
Bristol			
Altri			
Totale			

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

Nell'ospedale di Bristol, sono stati effettuati 143 interventi cardiochirugici e sono stati registrati 41 decessi (27.8%). Negli altri ospedali britannici, a fronte di 3176 interventi sono stati registrati 356 decessi (10.7%).

Valori osservati

Ospedale/Deceduti	Si	No	Totale
Bristol	41	102	143
Altri	356	2820	3176
Totale	397	2922	3319

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

$$\Pi = \frac{tot_{\text{decessi}}}{tot_{\text{interventi}}} = \frac{397}{3319} = 0.1196$$

Valori osservati

Ospedale/Deceduti	Si	No	Totale
Bristol	41	102	143
Altri	356	2820	3176
Totale	397	2922	3319

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

$$\Pi = \frac{tot_{\text{decessi}}}{tot_{\text{interventi}}} = \frac{397}{3319} = 0.1196$$

Valori osservati

Ospedale/Deceduti	Si	No	Totale
Bristol	41	102	143
Altri	356	2820	3176
Totale	397	2922	3319

Valori attesi

Ospedale/Deceduti	Si	No	Totale
Bristol	143*0.1196		143
Altri	3176*0.1196		3176
Totale	397	2922	3319

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

$$\Pi = \frac{tot_{\text{decessi}}}{tot_{\text{interventi}}} = \frac{397}{3319} = 0.1196$$

Valori osservati

Ospedale/Deceduti	Si	No	Totale
Bristol	41	102	143
Altri	356	2820	3176
Totale	397	2922	3319

Valori attesi

Ospedale/Deceduti	Si	No	Totale
Bristol	17.1		143
Altri	379.9		3176
Totale	397	2922	3319

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

$$\Pi = \frac{tot_{\text{decessi}}}{tot_{\text{interventi}}} = \frac{397}{3319} = 0.1196$$

Valori osservati

Ospedale/Deceduti	Si	No	Totale
Bristol	41	102	143
Altri	356	2820	3176
Totale	397	2922	3319

Valori attesi

Ospedale/Deceduti	Si	No	Totale
Bristol	17.1	125.9	143
Altri	379.9	2796.1	3176
Totale	397	2922	3319

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

Valori osservati

Ospedale/Deceduti	Si	No	Totale
Bristol	41	102	143
Altri	356	2820	3176
Totale	397	2922	3319

Valori attesi

Ospedale/Deceduti	Si	No	Totale
Bristol	17.1	125.9	143
Altri	379.9	2796.1	3176
Totale	397	2922	3319

$$\chi^2 = \sum \frac{(Osservati - Attesi)^2}{Attesi} = \frac{(41-17.1)^2}{17.1} + \frac{(102-125.9)^2}{125.9} + \frac{(356-379.9)^2}{379.9} + \frac{(2820-2796.1)^2}{2796.1} = 39.65$$

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici

Valori osservati

Ospedale/Deceduti	Si	No	Totale
Bristol	41	102	143
Altri	356	2820	3176
Totale	397	2922	3319

Valori attesi

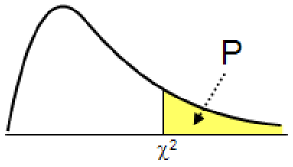
Ospedale/Deceduti	Si	No	Totale
Bristol	17.1	125.9	143
Altri	379.9	2796.1	3176
Totale	397	2922	3319

$$\chi^2 = \sum \frac{(Osservati - Attesi)^2}{Attesi} = \frac{(41-17.1)^2}{17.1} + \frac{(102-125.9)^2}{125.9} + \frac{(356-379.9)^2}{379.9} + \frac{(2820-2796.1)^2}{2796.1} = 39.65$$

$$\mathrm{df} = (n_{\mathrm{righe}} - 1) \times (n_{\mathrm{colonne}} - 1) = 1$$

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici



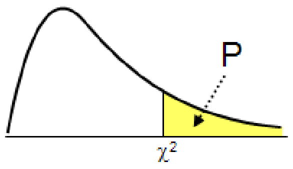
	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458

4. Verifico se la statistica osservata si trovi sulla coda di questa distribuzione e assegno una probabilità (P-value) a questo evento

$$\chi^2 = 39.65 \qquad \mathrm{df} = 1$$

Pearson's χ^2 test

📌 Il numero di bambini deceduti a seguito di interventi cardiochirugici a Bristol tra il 1984 e il 1995 è **lo stesso** degli altri ospedali britannici



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458

5. Dichiaro il risultato come statisticamente significativo se il P-value è inferiore a una soglia critica α

$$\chi^2 = 39.65 \quad \mathrm{df} = 1 \;\rightarrow\; \mathrm{P} < 0.001 = 3 \times 10^{-10} < \alpha = 0.05 \;\rightarrow\; \text{rifiuto } \mathcal{H}_0$$

Pearson's χ^2 test

Il livello di istruzione influenza la frequenza dell'esercizio fisico?

Valori osservati

	Nessuno	Sporadico	Regolare	Totale
Licenza elementare	$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	ΣRiga_1
Licenza media	$f_{1,2}$	...	...	ΣRiga_2
Diploma	$f_{1,3}$	...	...	ΣRiga_3
Laurea	$f_{1,4}$	...	...	ΣRiga_4
Totale	$\Sigma \text{Colonna}_1$	$\Sigma \text{Colonna}_2$	$\Sigma \text{Colonna}_3$	Totale

Pearson's χ^2 test

📌 Il livello di istruzione influenza la frequenza dell'esercizio fisico?

Valori attesi

	Nessuno	Sporadico	Regolare	Totale
Licenza elementare	$\frac{\Sigma\text{Riga}_1 \times \Sigma\text{Colonna}_1}{\text{Totale}}$	$\frac{\Sigma\text{Riga}_1 \times \Sigma\text{Colonna}_2}{\text{Totale}}$	$\frac{\Sigma\text{Riga}_1 \times \Sigma\text{Colonna}_3}{\text{Totale}}$	ΣRiga_1
Licenza media	$\frac{\Sigma\text{Riga}_2 \times \Sigma\text{Colonna}_1}{\text{Totale}}$	...	...	ΣRiga_2
Diploma	$\frac{\Sigma\text{Riga}_3 \times \Sigma\text{Colonna}_1}{\text{Totale}}$	...	...	ΣRiga_3
Laurea	$\frac{\Sigma\text{Riga}_4 \times \Sigma\text{Colonna}_1}{\text{Totale}}$	...	...	ΣRiga_4
Totale	$\Sigma\text{Colonna}_1$	$\Sigma\text{Colonna}_2$	$\Sigma\text{Colonna}_3$	Totale

❓ $\text{df} = \ ?$

Pearson's χ^2 test

📌 Il livello di istruzione influenza la frequenza dell'esercizio fisico?

Valori attesi

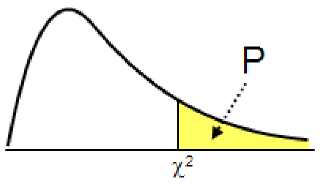
	Nessuno	Sporadico	Regolare	Totale
Licenza elementare	$\frac{\Sigma\text{Riga}_1 \times \Sigma\text{Colonna}_1}{\text{Totale}}$	$\frac{\Sigma\text{Riga}_1 \times \Sigma\text{Colonna}_2}{\text{Totale}}$	$\frac{\Sigma\text{Riga}_1 \times \Sigma\text{Colonna}_3}{\text{Totale}}$	ΣRiga_1
Licenza media	$\frac{\Sigma\text{Riga}_2 \times \Sigma\text{Colonna}_1}{\text{Totale}}$	...	...	ΣRiga_2
Diploma	$\frac{\Sigma\text{Riga}_3 \times \Sigma\text{Colonna}_1}{\text{Totale}}$	...	...	ΣRiga_3
Laurea	$\frac{\Sigma\text{Riga}_4 \times \Sigma\text{Colonna}_1}{\text{Totale}}$	...	...	ΣRiga_4
Totale	$\Sigma\text{Colonna}_1$	$\Sigma\text{Colonna}_2$	$\Sigma\text{Colonna}_3$	Totale

❓ $\text{df} = (n_{\text{righe}} - 1) \times (n_{\text{colonne}} - 1) = (4 - 1) \times (3 - 1) = 3 \times 2 = 6$

Esercizio #14

❓ L'essere astemio dipende dal reparto ospedaliero in cui si lavora?

Tra 279, 230 e 130 professionisti sanitari che lavorano nei reparti di medicina, chirurgia o altro (per esempio laboratori o altri servizi ospedalieri) sono stati individuati 122, 107, e 51 astemi.



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458

15:00

Albano, L. *et al.*, *Alcohol consumption in a sample of Italian healthcare workers: A cross-sectional study*, Archives of Environmental & Occupational Health, 2020

Esercizio #14 -- Soluzione

❓ L'essere astemio dipende dal reparto ospedaliero in cui si lavora?

Tra 279, 230 e 130 professionisti sanitari che lavorano nei reparti di medicina, chirurgia o altro (per esempio laboratori o altri servizi ospedalieri) sono stati individuati 122, 107, e 51 astemi.

$\mathcal{H}_0$: L'essere astemio **non** dipende dal reparto ospedaliero in cui si lavora

Esercizio #14 -- Soluzione

❓ L'essere astemio dipende dal reparto ospedaliero in cui si lavora?

Tra 279, 230 e 130 professionisti sanitari che lavorano nei reparti di medicina, chirurgia o altro (per esempio laboratori o altri servizi ospedalieri) sono stati individuati 122, 107, e 51 astemi.

Valori osservati

Reparto/Astemio	Si	No	Totale
Medicina	122	157	279
Chirurgia	107	123	230
Altro	51	79	130
Totale	280	359	639

Esercizio #14 -- Soluzione

❓ L'essere astemio dipende dal reparto ospedaliero in cui si lavora?

Valori osservati

Reparto/Astemio	Si	No	Totale
Medicina	122	157	279
Chirurgia	107	123	230
Altro	51	79	130
Totale	280	359	639

Valori attesi

Reparto/Astemio	Si	No	Totale
Medicina	$\frac{279 \times 280}{639}$	$\frac{279 \times 359}{639}$	279
Chirurgia	$\frac{230 \times 280}{639}$	$\frac{230 \times 359}{639}$	230
Altro	$\frac{130 \times 280}{639}$	$\frac{130 \times 359}{639}$	130
Totale	280	359	639

Esercizio #14 -- Soluzione

❓ L'essere astemio dipende dal reparto ospedaliero in cui si lavora?

Valori osservati

Reparto/Astemio	Si	No	Totale
Medicina	122	157	279
Chirurgia	107	123	230
Altro	51	79	130
Totale	280	359	639

Valori attesi

Reparto/Astemio	Si	No	Totale
Medicina	122.25	156.75	279
Chirurgia	100.78	129.22	230
Altro	56.96	73.04	130
Totale	280	359	639

$$\chi^2 = \sum \frac{(Osservati - Attesi)^2}{Attesi}$$

Esercizio #14 -- Soluzione

❓ L'essere astemio dipende dal reparto ospedaliero in cui si lavora?

Valori osservati

Reparto/Astemio	Si	No	Totale
Medicina	122	157	279
Chirurgia	107	123	230
Altro	51	79	130
Totale	280	359	639

Valori attesi

Reparto/Astemio	Si	No	Totale
Medicina	122.25	156.75	279
Chirurgia	100.78	129.22	230
Altro	56.96	73.04	130
Totale	280	359	639

$$\chi^2 = \frac{(122-122.25)^2}{122.25} + \frac{(107-100.78)^2}{100.78} + \frac{(51-56.96)^2}{56.96} + \frac{(157-156.75)^2}{156.75} + \frac{(123-129.22)^2}{129.22} + \frac{(79-73.04)^2}{73.04} = 1.17$$

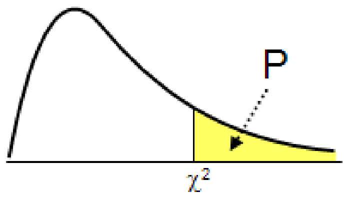
$$\mathrm{df} = (n_{\mathrm{righe}} - 1) \times (n_{\mathrm{colonne}} - 1) = 2$$

Esercizio #14 -- Soluzione

❓ L'essere astemio dipende dal reparto ospedaliero in cui si lavora?

$\mathcal{H}_0$: L'essere astemio **non** dipende dal reparto ospedaliero in cui si lavora

$$\chi^2 = 1.17, \qquad \mathrm{df} = 2$$



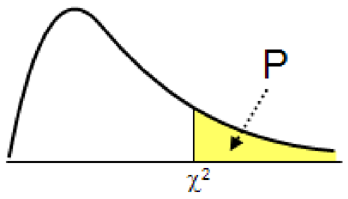
	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458

Esercizio #14 -- Soluzione

❓ L'essere astemio dipende dal reparto ospedaliero in cui si lavora?

$\mathcal{H}_0$: L'essere astemio **non** dipende dal reparto ospedaliero in cui si lavora

$$\chi^2 = 1.17, \qquad \mathrm{df} = 2 \quad \rightarrow \quad \mathrm{P} > 0.20 = 0.41 > \alpha = 0.05 \rightarrow \text{non rifiuto } \mathcal{H}_0$$



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458

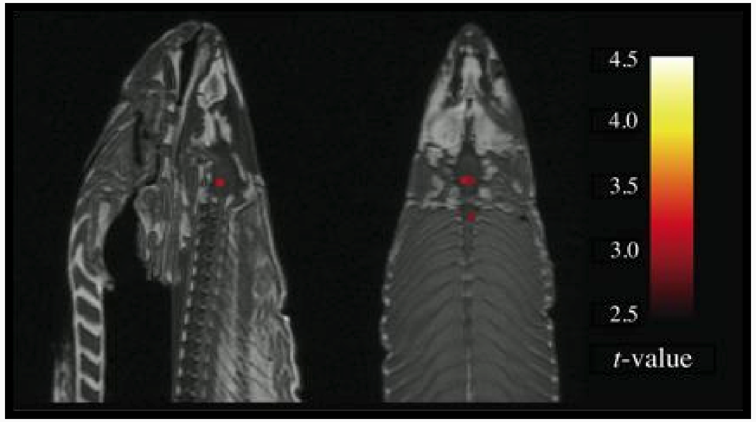
Quando facciamo più di un test alla volta?

📌 Un gruppo di ricerca ha effettuato fMRI su un singolo soggetto (*) mentre gli venivano mostrate delle fotografie in cui le persone fotografate esprimevano diverse emozioni. Sedici regioni cerebrali risultano statisticamente significative con $\mathrm{P} < 0.001$.

Quando facciamo più di un test alla volta?

📌 Un gruppo di ricerca ha effettuato fMRI su un singolo soggetto (*) mentre gli venivano mostrate delle fotografie in cui le persone fotografate esprimevano diverse emozioni. Sedici regioni cerebrali risultano statisticamente significative con $\mathrm{P} < 0.001$.

(*) Atlantic salmon, '*not alive at the time of scanning*'



Bennett, C. M., Miller M.B., and Wolford G.L.,. *Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction.* Neuroimage 47.Suppl 1 (2009) doi:10.1016/S1053-8119(09)71202-9

Quando facciamo più di un test alla volta?

📌 $\mathrm{P} = 0.05 \rightarrow 5\%$ rifiutiamo $\mathcal{H}_0$ anche se è vera
$\mathcal{P} = 1 - 0.95 = 0.05$

Quando facciamo più di un test alla volta?

📌 $\mathrm{P} = 0.05 \rightarrow 5\%$ rifiutiamo $\mathcal{H}_0$ anche se è vera
$\mathcal{P} = 1 - 0.95 = 0.05$

Con 2 test, averne almeno uno con $\mathrm{P} < 0.05$ è:
$\mathcal{P} = 1 - 0.95 \times 0.95 = 1 - 0.95^2 = 0.0975 \rightarrow \quad \approx 10\%$

Con 3 test, averne almeno uno con $\mathrm{P} < 0.05$ è:
$\mathcal{P} = 1 - 0.95^3 = 0.145 \rightarrow \quad \approx 14\%$

Con 10 test, averne almeno uno con $\mathrm{P} < 0.05$ è:
$\mathcal{P} = 1 - 0.95^{10} = 0.40 \rightarrow \quad \approx 40\%$

Correzione per test multipli

🎯 Quando si fanno più test, si richiede un α inferiore

Bonferroni-correction: $\alpha = \frac{0.05}{N_{\text{test}}}$

Con 10 test, averne almeno uno con $\mathrm{P} < \frac{0.05}{10} = 0.005$ è

$$\mathcal{P} = 1 - 0.995^{10} = 0.049 \rightarrow \quad \approx 5\%$$

Correzione per test multipli

🎯 Quando si fanno più test, si richiede un α inferiore
Quando si fanno più test, si fissa il numero di 'scopertè che sono false

False discovery rate (FDR, Benjamini–Hochberg procedure):

1. ordino i risultati per P value crescente
2. Rifiuto $\mathcal{H}_0$ sino a che $P_{(k)} > \alpha \times \frac{k}{N_{\text{test}}}$

Errori dei test statistici

$\mathcal{H}_0$ è	Non rifiutata	Rifiutata
Vera		
Falsa		

Errori dei test statistici

$\mathcal{H}_0$ è	Non rifiutata	Rifiutata
Vera		Falso Positivo
Falsa	Falso negativo	

Errori dei test statistici

Sospetto è	Assolto	Condannato
Innocente		Condanno un innocente
Colpevole	Assolvo un colpevole	

Errori dei test statistici

$\mathcal{H}_0$ è	Non rifiutata	Rifiutata
Vera		Errore di I tipo
Falsa	Errore di II tipo	

Esercizio #15

? In un villaggio, c'era un pastorello che faceva la guardia alle pecore. Annoiandosi, per diverse notti, si mise ad urlare "Al lupo! Al lupo!", così tutti accorrevano per aiutarlo. Una notte, un lupo venne veramente. Il pastorello cominciò a gridare: "Al lupo, al lupo!",ma nessuno venne perché tutti pensarono che fosse uno scherzo.

Che tipo di errore si sta commettendo?

a) Errore del primo tipo, poi del secondo tipo
b) Errore del secondo tipo, poi del primo tipo
c) Errore nullo, poi errore alternativo
d) Errore alternativo, poi errore nullo
e) Nessuno dei precedenti

02:00

Esercizio #15 -- Soluzione

❓ In un villaggio, c'era un pastorello che faceva la guardia alle pecore. Annoiandosi, per diverse notti, si mise ad urlare "Al lupo! Al lupo!", così tutti accorrevano per aiutarlo. Una notte, un lupo venne veramente. Il pastorello cominciò a gridare: "Al lupo, al lupo!",ma nessuno venne perché tutti pensarono che fosse uno scherzo.

Che tipo di errore si sta commettendo?

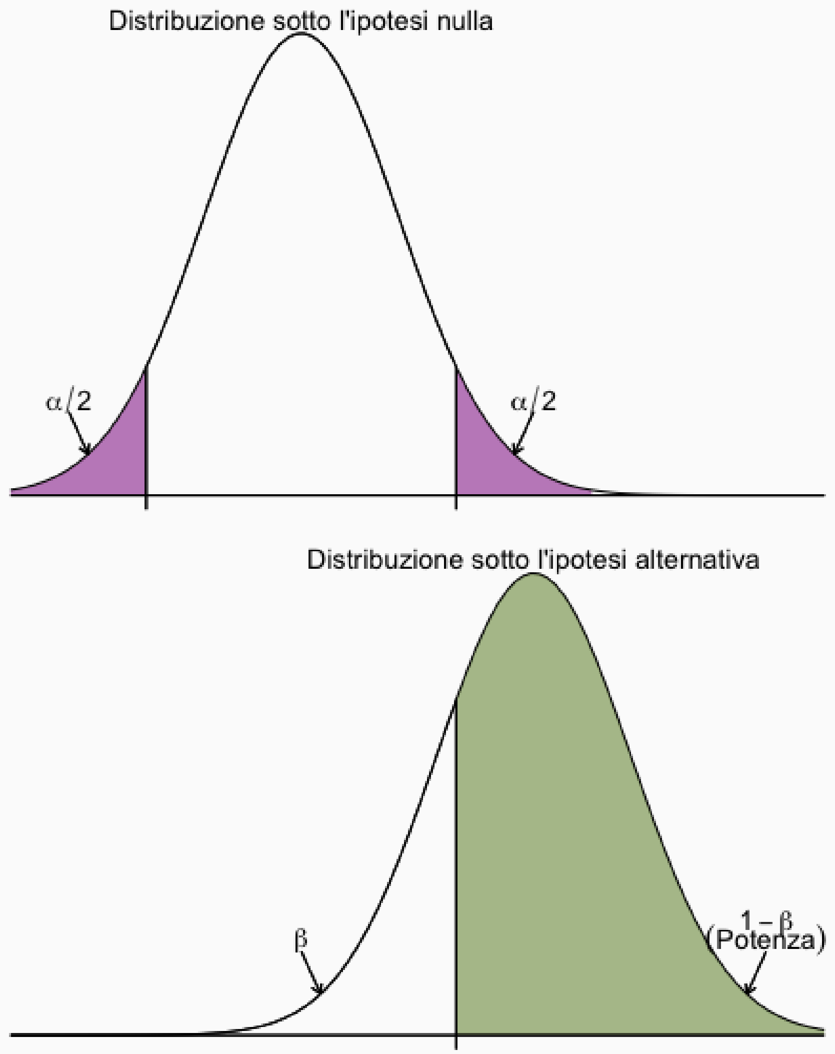
a) Errore del primo tipo, poi del secondo tipo ✅
b) Errore del secondo tipo, poi del primo tipo
c) Errore nullo, poi errore alternativo
d) Errore alternativo, poi errore nullo
e) Nessuno dei precedenti

Errori dei test statistici

$\mathcal{H}_0$ è	Non rifiutata	Rifiutata
Vera		α
Falsa	β	

La potenza di un test

$\mathcal{H}_0$ è	Non rifiutata	Rifiutata
Vera		α
Falsa	β	$1-\beta$ Potenza

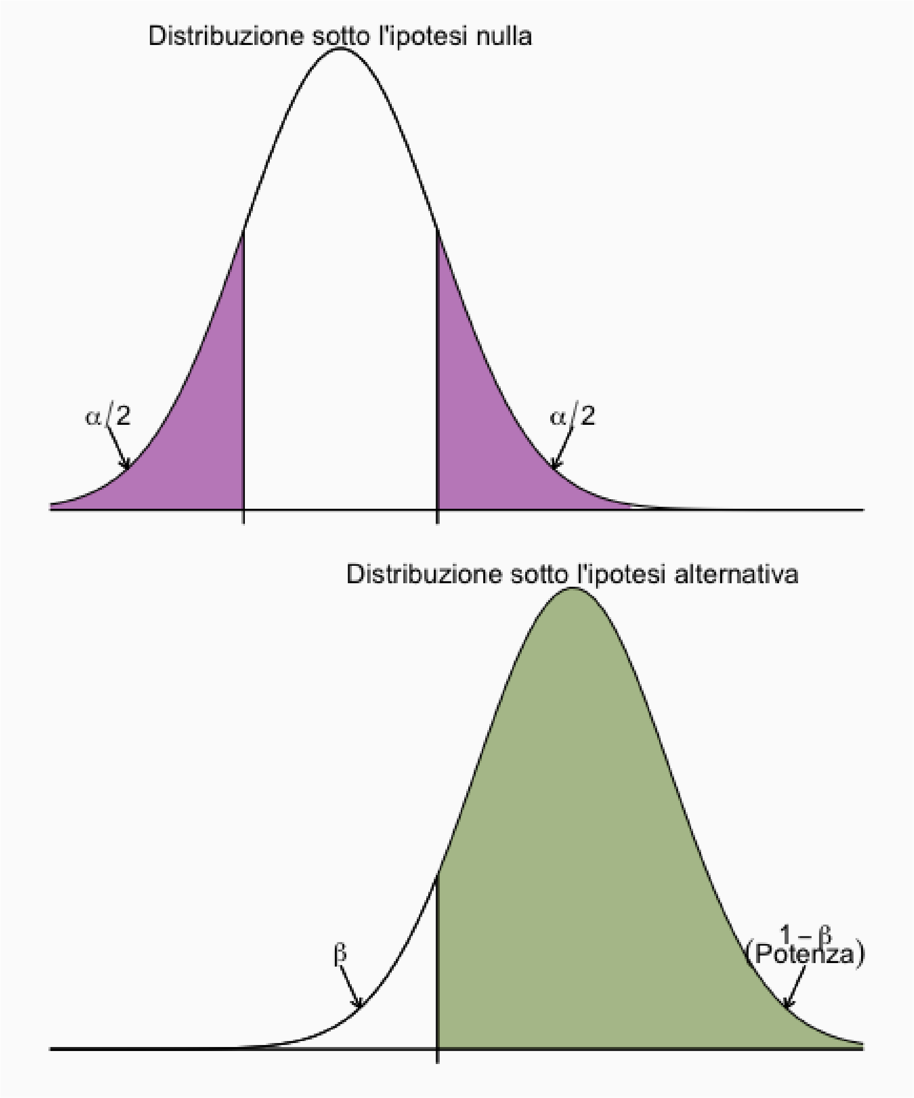


$\alpha = 0.05$
$1 - \beta = 0.8$

La potenza di un test

🎯 La potenza aumenta:

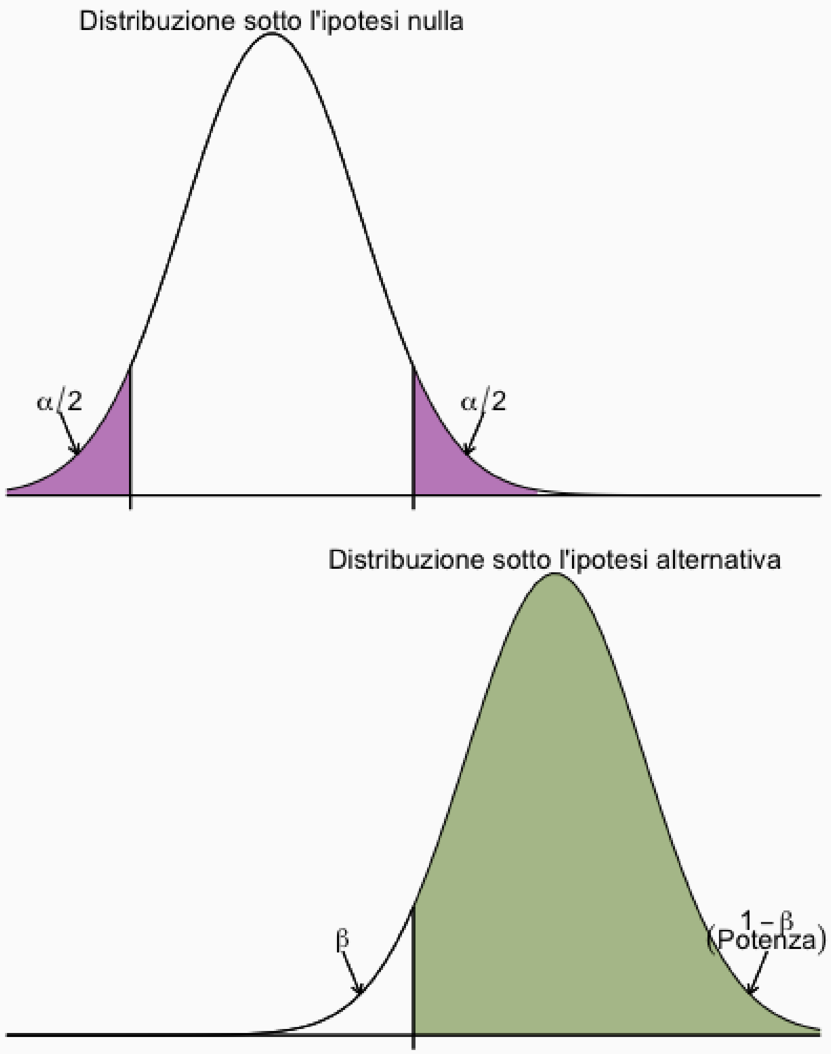
- all'aumentare del livello di significatività α



La potenza di un test

🎯 La potenza aumenta:

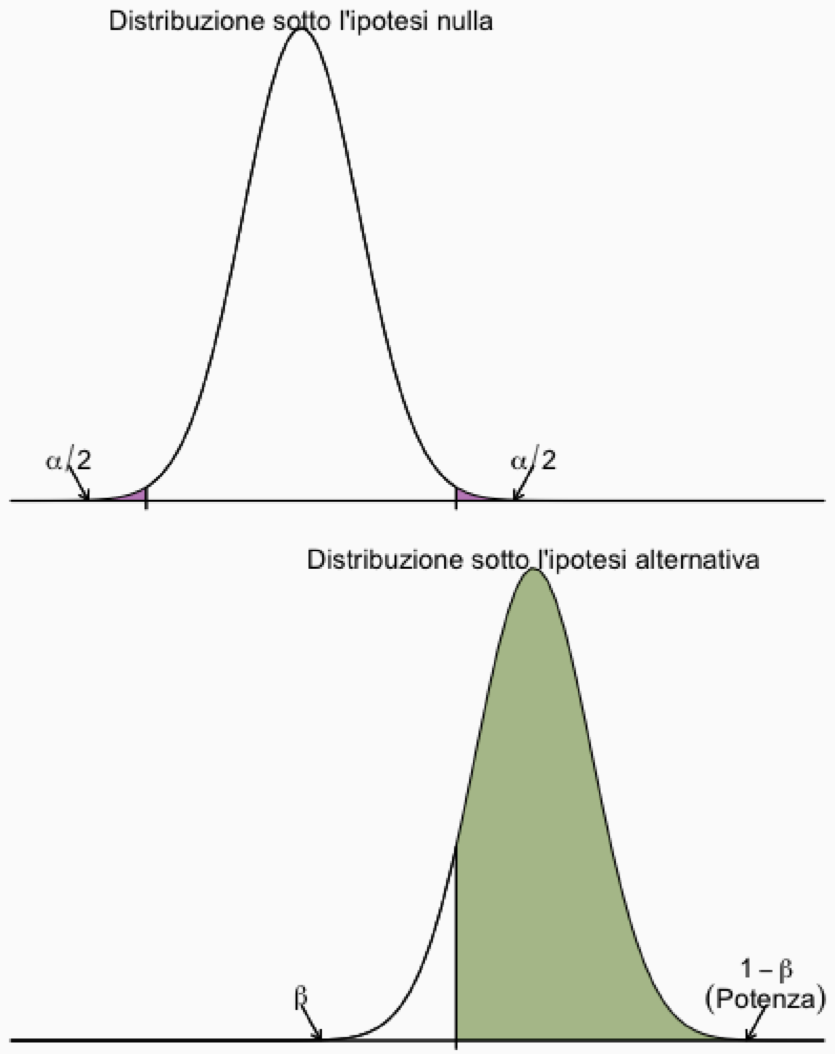
- all'aumentare del livello di significatività α
- all'aumentare della differenza $\mu_c - \mu_i$ o $\pi_c - \pi_i$



La potenza di un test

🎯 La potenza aumenta:

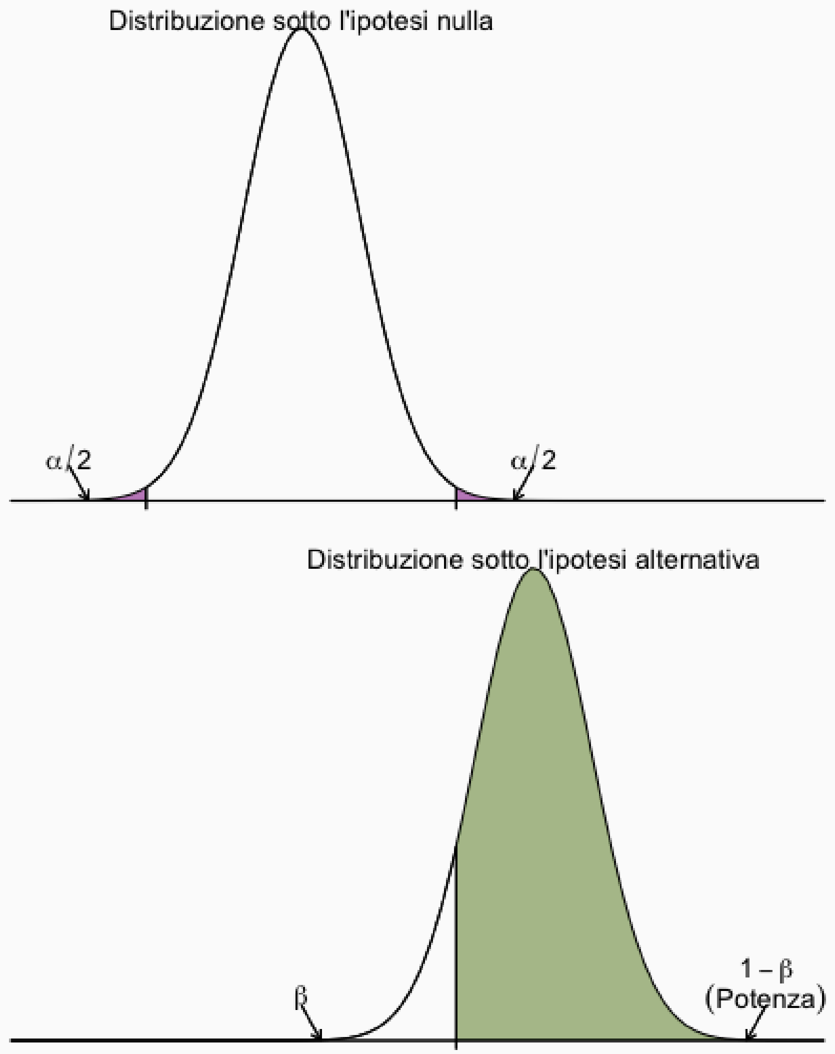
- all'aumentare del livello di significatività α
- all'aumentare della differenza $\mu_{\mathrm{c}} - \mu_{\mathrm{i}}$ o $\pi_{\mathrm{c}} - \pi_{\mathrm{i}}$
- al diminuire della deviazione standard σ^2



La potenza di un test

🎯 La potenza aumenta:

- all'aumentare del livello di significatività α
- all'aumentare della differenza $\mu_{\mathrm{c}} - \mu_{\mathrm{i}}$ o $\pi_{\mathrm{c}} - \pi_{\mathrm{i}}$
- al diminuire della deviazione standard σ^2
- all'aumentare della dimensione campionaria n



Comunicare (e interpretare) il risultato

Scenario 4

- Test di ipotesi: P value $> \alpha$
- Intervallo di confidenza: molto largo
- Dimensione dell'effetto: grande

Probabilmente abbiamo un campione troppo piccolo per decidere con sicurezza se rifiutare o meno l'ipotesi nulla

Il nostro test non ha abbastanza potenza

Esercizio #16

❓ Voglio aumentare la potenza del mio studio. Quali fattori sono effettivamente modificabili?

a) il livello di significatività α
b) la differenza $\mu_\mathrm{c} - \mu_\mathrm{i}$ o $\pi_\mathrm{c} - \pi_\mathrm{i}$
c) la deviazione standard (σ^2) dei due campioni
d) la dimensione n dei due campioni
e) nessuna delle precedenti

01:00

Esercizio #16 -- Soluzione

❓ Voglio aumentare la potenza del mio studio. Quali fattori sono effettivamente modificabili?

a) il livello di significatività α ✅
b) la differenza $\mu_{\mathrm{c}} - \mu_{\mathrm{i}}$ o $\pi_{\mathrm{c}} - \pi_{\mathrm{i}}$
c) la deviazione standard (σ^2) dei due campioni
d) la dimensione n dei due campioni ✅
e) nessuna delle precedenti

Esercizio #17

? Completate le definizioni con i seguenti termini:
Errore di I tipo, Errore di II tipo, Potenza di un testtest

a) Concludere che un trattamento funzioni quando in realtà non ha nessun effetto si dice: ....

b) Concludere che un trattamento **non** funzioni quando in realtà è efficace si dice: ....

c) Concludere correttamente che un trattamento funzioni si dice ....

02:00

Esercizio #17 -- Soluzione

❓ Completate le definizioni con i seguenti termini:
Errore di I tipo, Errore di II tipo, Potenza di un test

a) Concludere che un trattamento funzioni quando in realtà non ha nessun effetto si dice: …*Errore di I tipo*…

b) Concludere che un trattamento **non** funzioni quando in realtà è efficace si dice: …*Errore di II tipo*…

c) Concludere correttamente che un trattamento funzioni si dice …*Potenza di un test*…

Esercizio #18

? Quale α genera una probabilità maggiore di un errore di I tipo?

a) 0.01 b) 0.05

? Quale α genera una probabilità maggiore di un errore di II tipo?

a) 0.01 b) 0.05

? Quale α genera una maggiore potenza?

a) 0.01 b) 0.05

02:00

Esercizio #18 -- Soluzione

❓ Quale α genera una probabilità maggiore di un errore di I tipo?

a) 0.01 b) 0.05 ✅

❓ Quale α genera una probabilità maggiore di un errore di II tipo?

a) 0.01 b) 0.05

❓ Quale α genera una maggiore potenza?

a) 0.01 b) 0.05

Esercizio #18 -- Soluzione

❓ Quale α genera una probabilità maggiore di un errore di I tipo?

a) 0.01 b) 0.05 ✅

❓ Quale α genera una probabilità maggiore di un errore di II tipo?

a) 0.01 ✅ b) 0.05

❓ Quale α genera una maggiore potenza?

a) 0.01 b) 0.05

Esercizio #18 -- Soluzione

❓ Quale α genera una probabilità maggiore di un errore di I tipo?

a) 0.01 b) 0.05 ✅

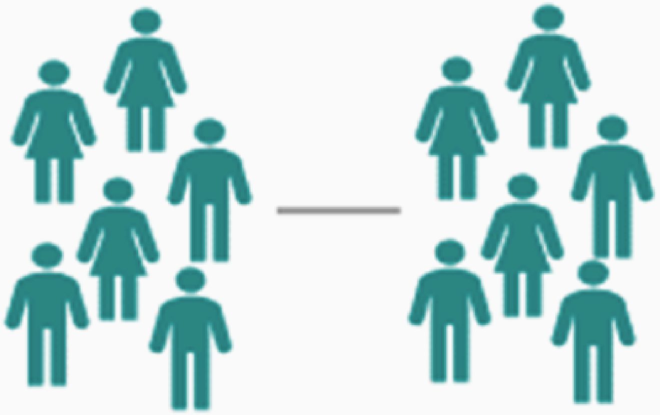
❓ Quale α genera una probabilità maggiore di un errore di II tipo?

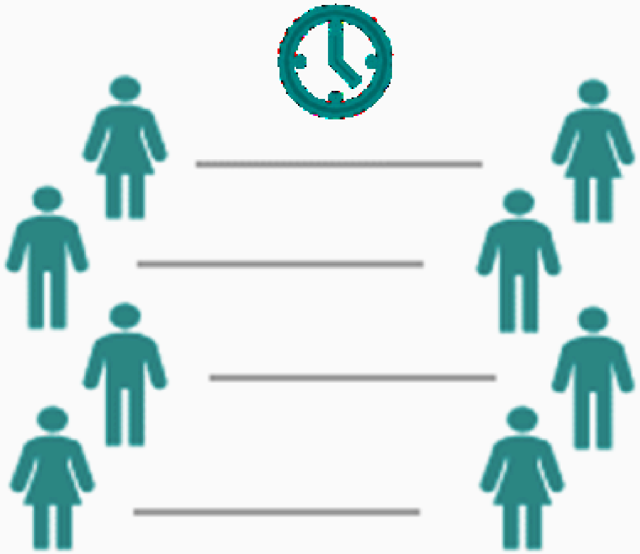
a) 0.01 ✅ b) 0.05

❓ Quale α genera una maggiore potenza?

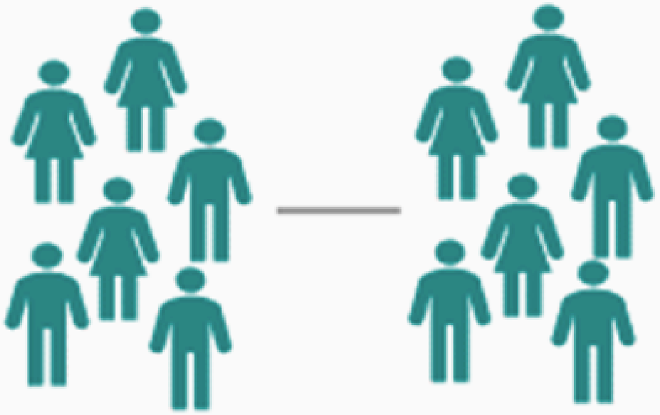
a) 0.01 b) 0.05 ✅

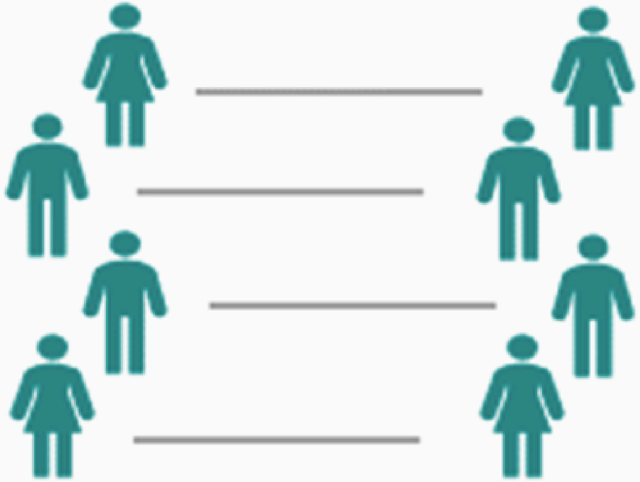
Campioni indipendenti & dipendenti





Campioni indipendenti & dipendenti





Test non-parametrici

Campione	Tipo del dato	$\mathcal{H}_0$	Test non parametrico
Indipendenti	Numerici	$\mu_1 = \mu_2$	Mann-Whitney's test
Dipendenti	Numerici	$\mu_1 = \mu_2$	Wilcoxon's test
Indipendenti	Categoriche	$\pi_1 = \pi_2$	Fisher's test
Dipendenti	Categoriche	$\pi_1 = \pi_2$	McNemar's test

Cosa abbiamo imparato in questa lezione?

- P-value misura l'incompatibilità tra i dati e la nostra ipotesi (probabilità di osservare valori così estremi se $\mathcal{H}_0$ è vera)
- Tradizionalmente, $P < 0.05$ o < 0.01 sono considerati statisticamente significativi, ma queste soglie devono essere corrette per il numero di test
- C'è una corrispondenza tra CI e P-value, e se il 95% CI non include lo zero, possiamo rifiutare $\mathcal{H}_0$ a un livello si significatività $\alpha = 0.05$
- Non sempre una significatività statistica corrisponde a una significatività clinica
- Errori del primo tipo dipendono dalla soglia di significatività α
- Esiste un legame tra errori del secondo tipo β e potenza di uno studio
- Per dati con distribuzioni non-normali possiamo usare test non parametrici