Last updated Aprile 15, 2024

# Il Test di Ipotesi
# Lezione 6

G. Bacaro

Statistica
CdL in Scienze e Tecnologie per l'Ambiente e la Natura
I anno, II semestre

# Il test di ipotesi

Cuore della statistica inferenziale!

Fino a questo momento abbiamo solo descritto popolazioni o campioni

o

stimato parametri della popolazione usando campioni (e.g. intervalli di confidenza)

Ora siamo pronti per testare vere ipotesi sulle popolazioni!

# Logica del processo inferenziale: i 5 passi

1. Costruire un'ipotesi (questione oggetto della ricerca)

E.g. Problema: gli studenti maschi hanno dei voti più alti delle femmine?

2. Scegliere l'analisi statistica

E.g. Analisi per testare differenze fra medie

3. Pianificare ed eseguire il campionamento

E.g. Selezionare un campione di M e F e raccogliere i dati

4. Eseguire il test

5. Rifiutare o accettare l'ipotesi di partenza

E.g. Maschi e femmine non sono diversi

**Errore comune**

Eseguire il campionamento prima di aver costruito l'ipotesi e aver scelto l'analisi

# 1. Costruire e testare un’ipotesi

Ipotesi: affermazione che ha come oggetto accadimenti nel mondo reale, che si presta ad essere confermata o smentita dai dati osservati sperimentalmente

Esempio: gli studenti maschi e femmine presentano gli stessi voti

# 1. Costruire e testare un'ipotesi

**Ipotesi nulla ($H_0$)**: è un'affermazione riguardo alla popolazione che si assume essere vera fino a che non ci sia una prova evidente del contrario (status quo, mancanza di effetto etc.)

**Ipotesi alterantiva ($H_a$)**: è un'affermazione riguardo alla popolazione che è contraria all'ipotesi nulla e che viene accettata solo nel caso in cui ci sia una prova evidente in suo favore

# 1. Costruire e testare un’ipotesi

Test di ipotesi consiste in una decisione fra $H_0$ e $H_a$

1. Rifiutare $H_0$ (e quindi accettare $H_a$)

2. Accettare $H_0$ (e quindi rifiutare $H_a$)

# 1. Costruire e testare un'ipotesi

1. Rifiutare $H_0$ ? 2. Accettare $H_0$

La statistica inferenziale ci permette di quantificare delle probabilità per decidere se accettare o rifiutare l'ipotesi nulla:
Quanto attendibile è $H_0$?

# Testare un’ipotesi su una popolazione con σ e μ note!

# Testare un'ipotesi su una popolazione

$H_0$: Il campione ha una media=μ

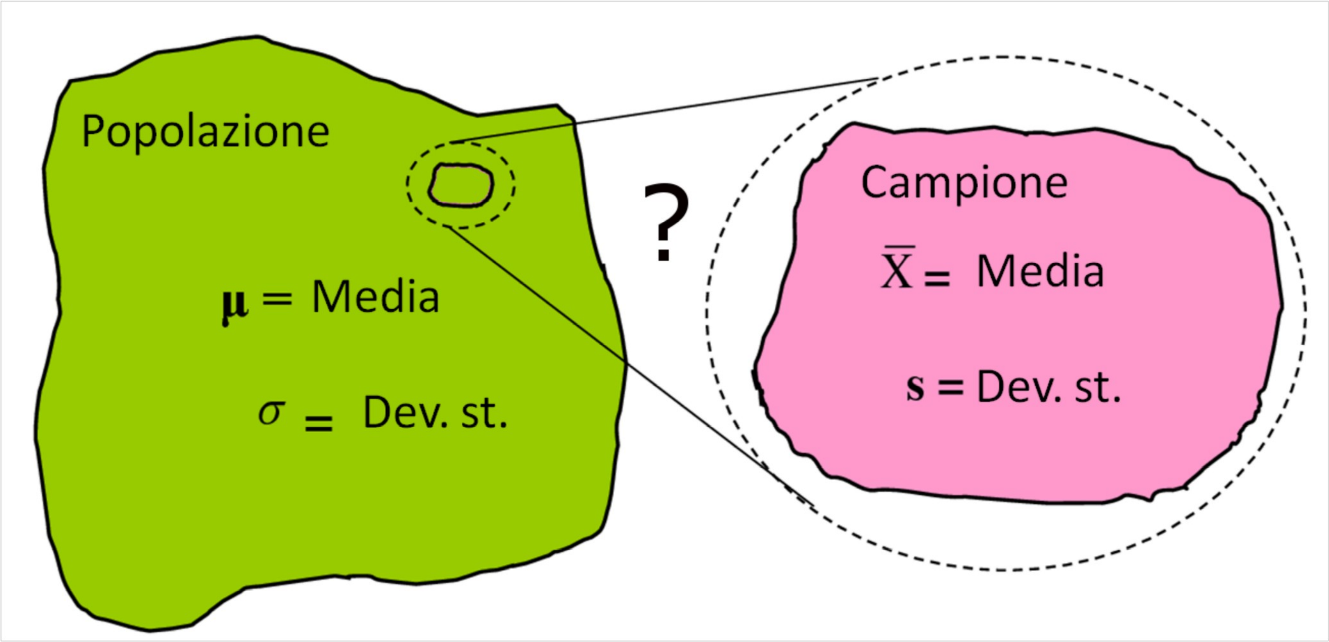
TRE IPOTESI ALTERNATIVE POSSIBILI:

$H_a$: Il campione ha una media ≠ μ

$H_a$: Il campione ha una media > μ

$H_a$: Il campione ha una media < μ

Come quantificare la probabilità per decidere se accettare o rifiutare l'ipotesi nulla?

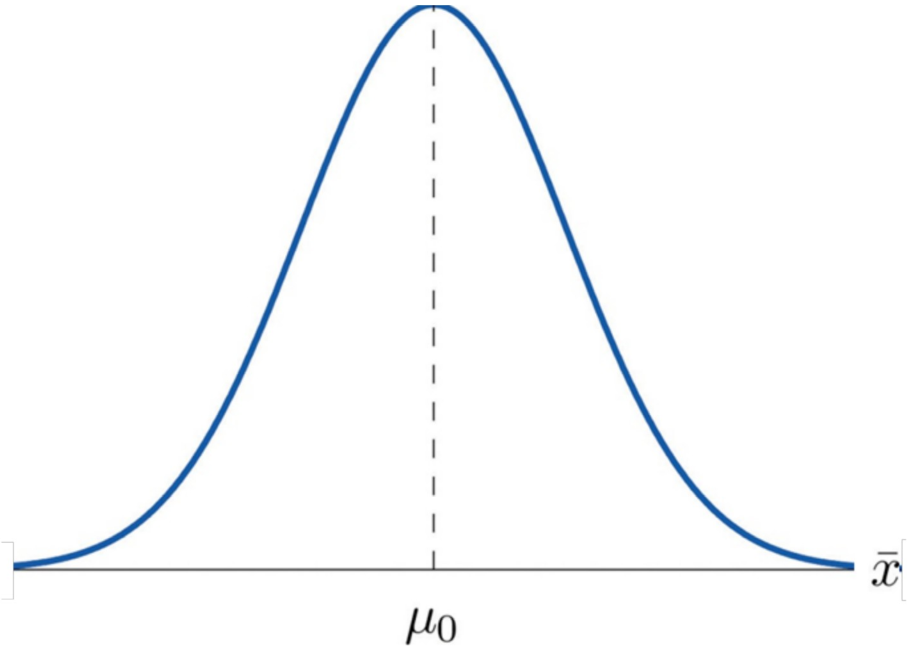


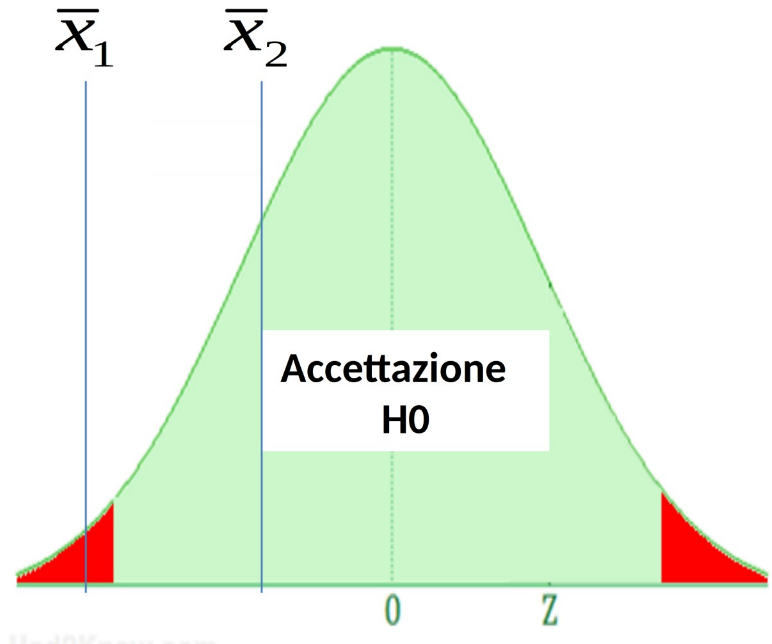
# Testare un'ipotesi su una popolazione

Popolazione con media μ e deviazione standard σ nota:

$H_0$: Il campione ha una media= μ

$H_a$: Il campione ha una media ≠ μ (o >μ o <μ)




$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

# Livello di significatività (alpha)

Devo definire a priori una probabilità (alpha) per rifiutare l'ipotesi nulla

Il livello di significatività di un test: probabilità di rifiutare $H_0$, quando in realtà è vera (quanto confidenti siamo nelle nostre conclusioni?)

Più piccola è alpha maggiore sarà la certezza nel rifiutare l'ipotesi nulla

**Valori usuali sono 10%, 5%, 1%, 0.1%**

I valori più comuni

# Esempio

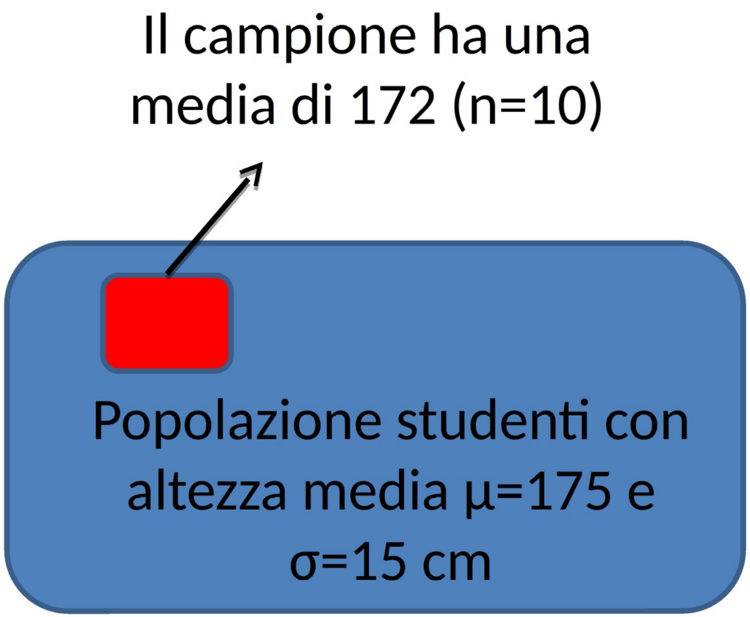
$H_0$: Il campione ha una media = 175

TRE IPOTESI ALTERNATIVE POSSIBILI:

$H_a$: Il campione ha una media ≠ 175

$H_a$: Il campione ha una media > 175

$H_a$: Il campione ha una media < 175




$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Quale alpha volete?

# Testare un'ipotesi su una popolazione

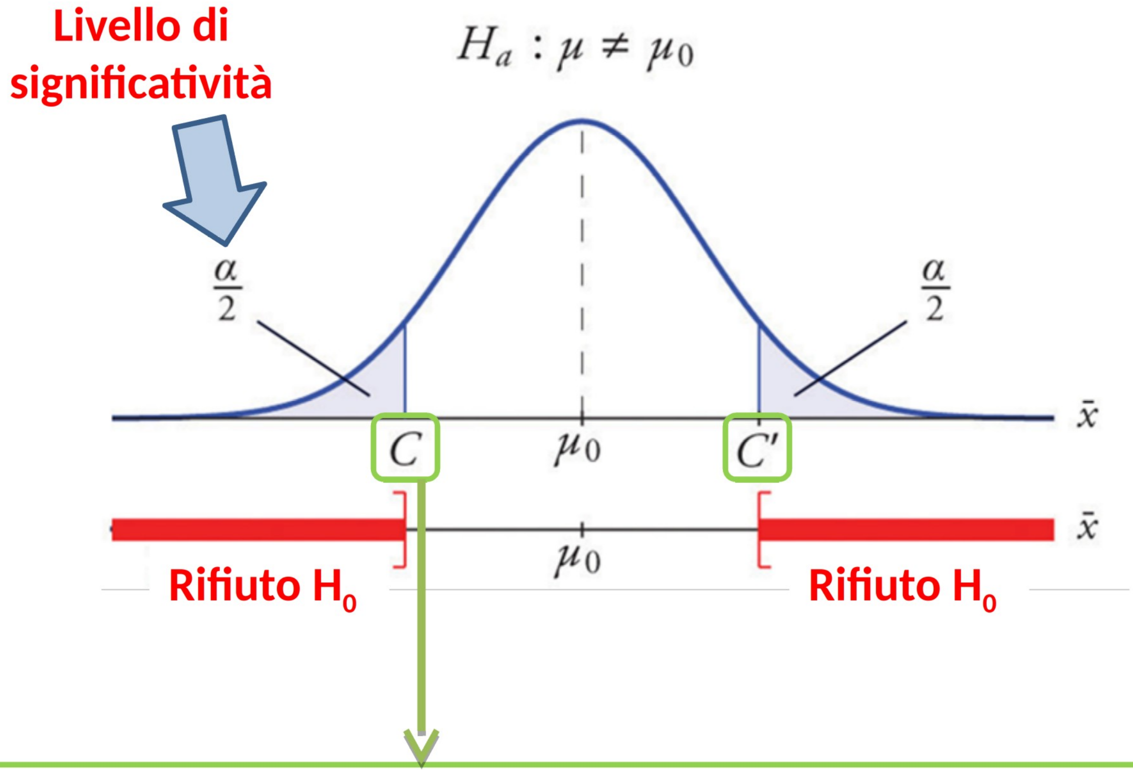
Passi:

1. Definisco alpha
2. Controllo il z-critico a seconda che sia una o due code dalla tavola
3. Calcolo il valore di z-calcolato

$$z_{calcolato} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

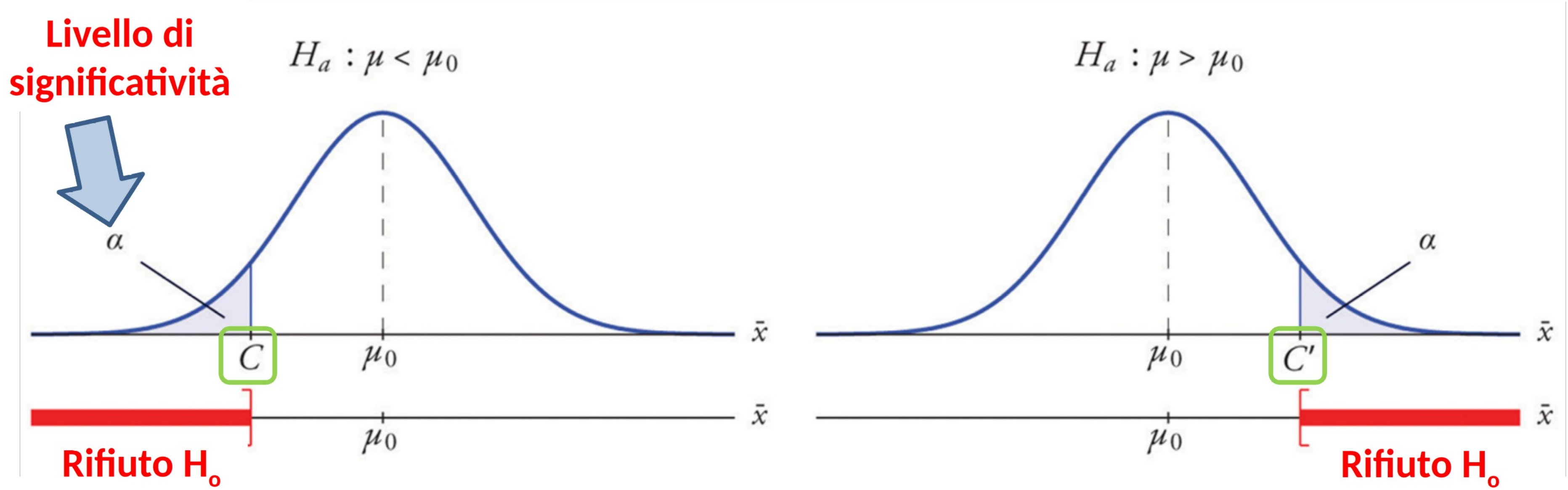
5. Se z-calcolato è più estremo del valore critico rifiuto H0

# Regione di accettazione e rifiuto




Definito il valore critico (C) della statistica possiamo individuare una zona di rifiuto e una zona di accettazione di $H_0$

# Regione di accettazione e rifiuto




Definito il valore critico della statistica (C) possiamo individuare una zona di rifiuto e una zona di accettazione di $H_o$

# Testare un’ipotesi su una popolazione con $\mu$ nota ma $\sigma$ incognita

Caso più comune!

# Testare un'ipotesi su una popolazione

Se non conosciamo la deviazione standard devo usare la distibuzione di t!

$H_0$: Il campione ha una media = μ

$H_a$: Il campione ha una media ≠ μ (o >μ o <μ)

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

t si ditribuisce come una variabile causale con n-1 gradi di libertà

# Testare un'ipotesi su una popolazione

Passi:

1. Definisco alpha
2. Determino n-1 (d.g.)
3. Controllo il t-critico a seconda che sia una o due code dalla tavola
4. Calcolo il t-calcolato

$$t_{calcolato} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

5. Se t-calcolato è più estremo del critico rifiuto H0

# Tipi di errori nel test di ipotesi

Tutte le decisioni sono basate su probabilità

Esiste sempre un livello di incertezza sia nel rifiutare sia nell'accettare l'ipotesi nulla

Il test di per se non prova nulla in modo assoluto!!!

# Tipo di errori

**Decisione statistica**

**Realtà**

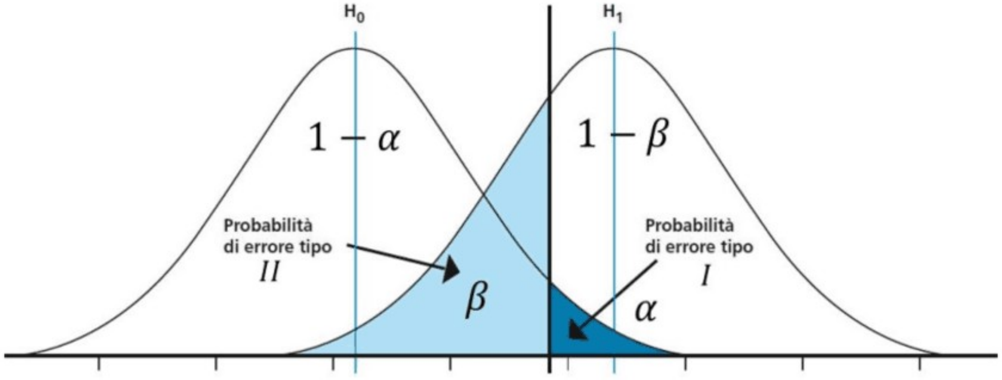
| | Rifiuto $H_0$ | Accetto $H_0$ |
|---|---|---|
| Effetto | **Corretta**<br>Effetto visto | **Errore tipo II (Βετα)**<br>Effetto non visto, ma in realtà esiste |
| No effetto | **Errore tipo I (Αλφα)**<br>Effetto visto, ma in realtà non esiste<br>(valore di P) | **Corretta**<br>Effetto non visto,<br>Effetto non esiste<br>(Power) |

# Errore di Tipo I e II

- *Errore di I tipo "α": rifiuto un'ipotesi vera*
- *Errore di II tipo "β": accetto ipotesi falsa*

| Realtà | Decisione | |
|---|---|---|
| | *Non rifiuto* $H_0$ | *Rifiuto* $H_0$ |
| $H_0$ *vera* | Decisione corretta $\boldsymbol{1-\alpha}$ | $\boldsymbol{\alpha}$ |
| $\mathrm{H}_0$ falsa | $\boldsymbol{\beta}$ | Decisione corretta $\boldsymbol{1-\beta}$ |

La probabilità $\boldsymbol{1-\beta}$ è detta ***potenza del test.*** Un test è tanto "più potente" quanto più è atto a respingere ipotesi false.



# Potenza di un test

I fattori che incidono sulla potenza di un test per la verifica dell'ipotesi nulla sono 6:

1 - il livello di significatività (alpha);

2 - la dimensione della differenza, di cui si vuole verificare la significatività;

3 - la variabilità dei dati;

4 - la direzione dell'ipotesi (una o due code);

5 - la dimensione (n) del campione;

6 - le caratteristiche del test

# Potenza di un test

1. Livello di significatività (Alpha):

Il timore di commettere errori di I tipo tende a far abbassare al ricercatore il livello di significatività alpha. Ma, riducendo il valore di alpha, egli diminuisce la probabilità di scoprire delle differenze, anche quando nella realtà esistono; in altri termini, aumenta la probabilità “beta” di commettere errori di II tipo

# Potenza di un test (Power)

2. La dimensione della differenza tra il valore osservato e il valore atteso nell'ipotesi nulla (di solito, la media)

La potenza di un test statistico è funzione crescente della differenza, considerata in valore assoluto

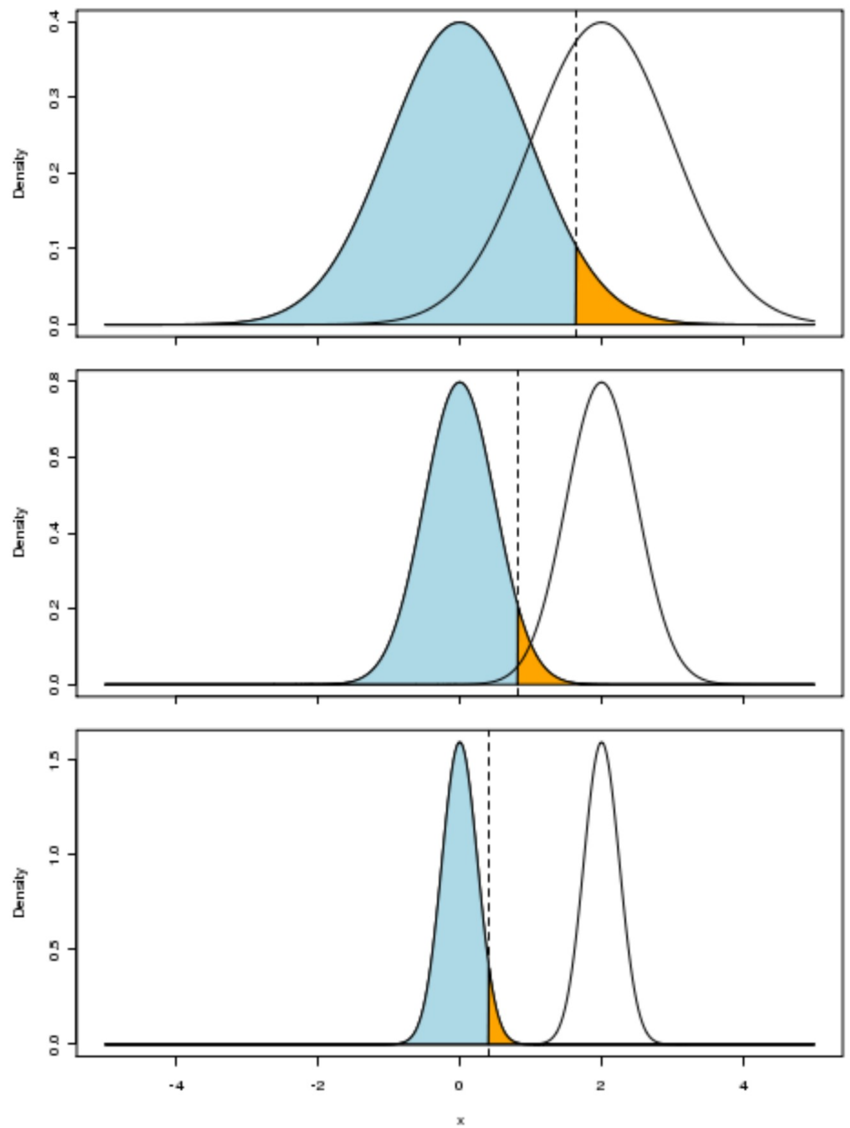
$$z = \frac{\boxed{\bar{x} - \mu}}{\frac{\sigma}{\sqrt{n}}}$$

Esempio test z

# Potenza di un test (Power)

3. Variabilità

$$z = \frac{\bar{x} - \mu}{\frac{\boxed{\sigma}}{\sqrt{n}}}$$

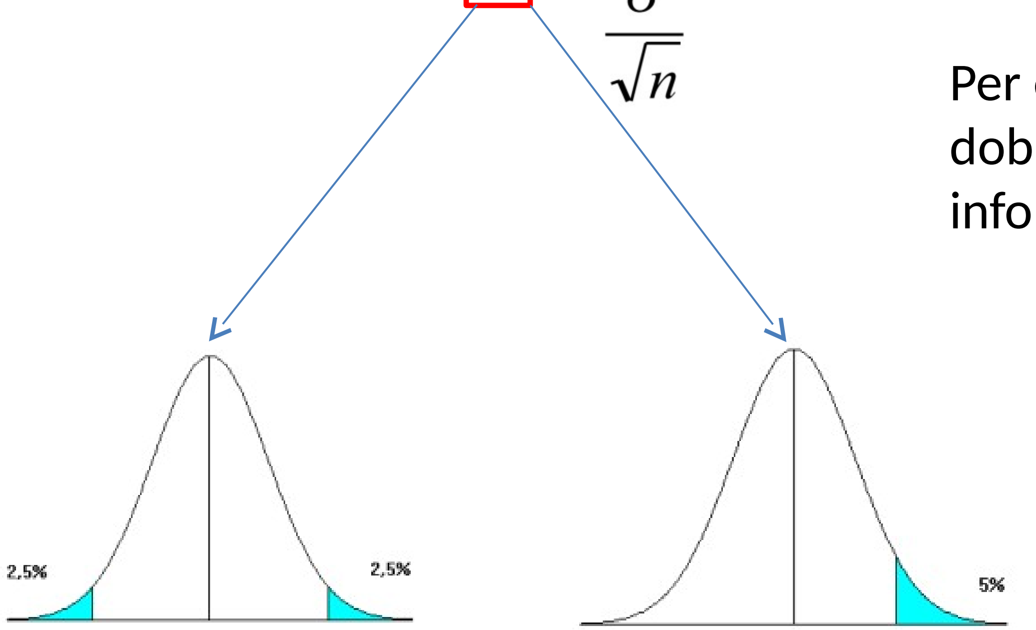


# Potenza di un test (Power)

4. Direzione del test (una o due  code)

Lo stesso valore di Z può essere associato a due probabilità a seconda dell'ipotesi alternativa

$$\boxed{z} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$



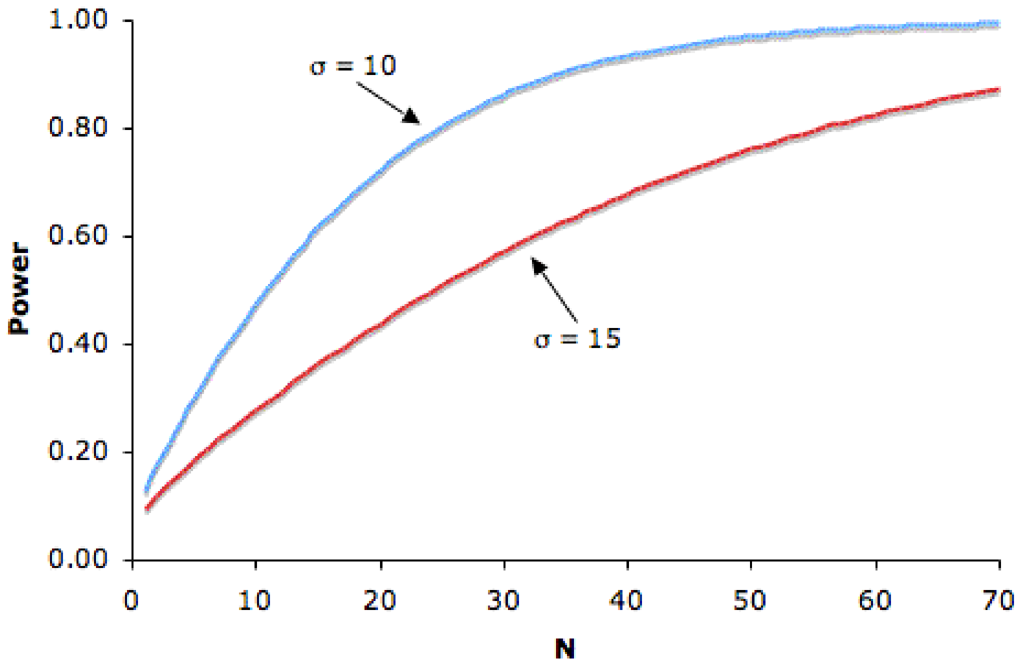

Qual è più potente?

Per eseguire un test a una coda dobbiamo avere delle informazioni a priori

# Potenza di un test (Power)

5. La dimensione del campione (n)

$$z = \frac{\bar{x} - \mu}{\dfrac{\sigma}{\boxed{\sqrt{n}}}}$$



# Potenza di un test (Power)

6. La scelta del test

A partire dagli stessi dati, non tutti i test hanno la stessa capacità di rifiutare l'ipotesi nulla quando è falsa

E' quindi molto importante scegliere il test più adatto!

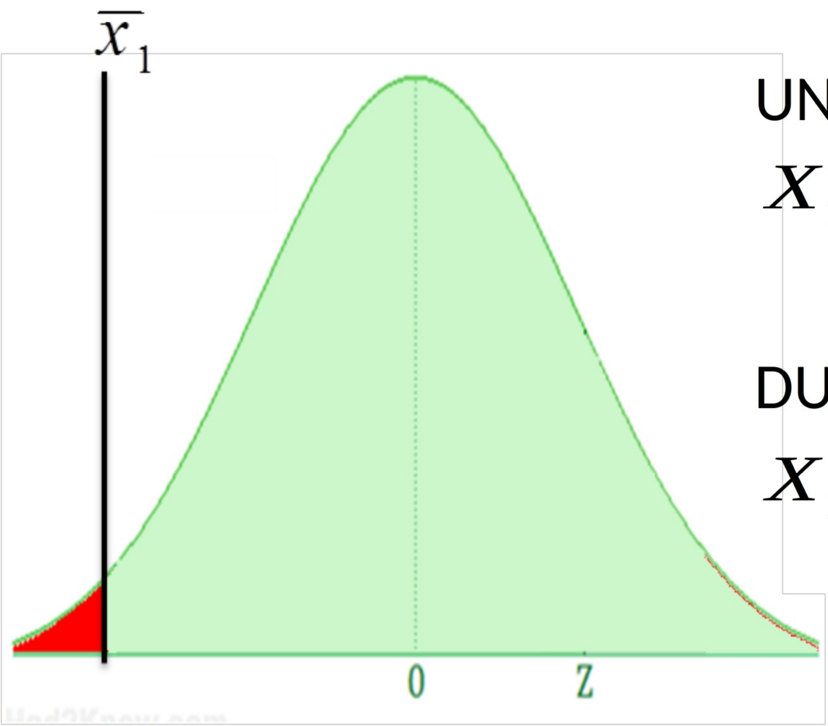
Test diversi hanno condizioni di validità differenti e sono più o meno robusti (sopportano in modo differente l'allontanamento dalle condizioni di validità)

Aspettiamo le prossime lezioni…

# Il valore di P

Il **valore *p*** indica la probabilità di ottenere un risultato pari o più estremo di quello osservato, **supposta vera l'ipotesi nulla**

Talvolta viene anche chiamato **livello di significatività osservato**




UNA CODA

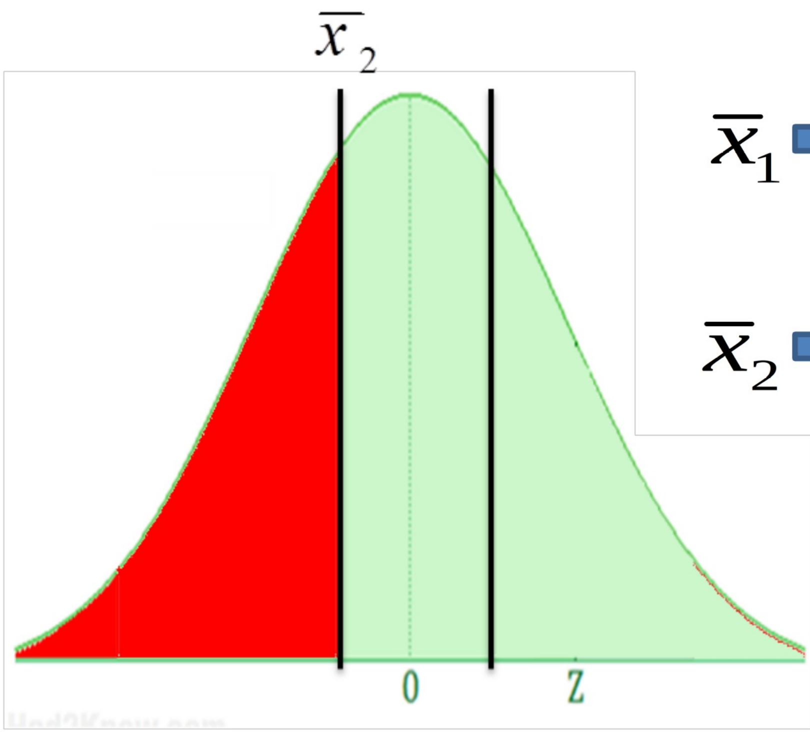
$X_1 \Rightarrow Z_{calcolato} \Rightarrow P = 0.045$

DUE CODE

$X_1 \Rightarrow Z_{calcolato} \Rightarrow P = ?$

# Il valore di P

Il calcolo del P dipende dall'ipotesi alternativa!




$$\overline{x}_1 \Rightarrow Z_1 \Rightarrow P = 0.40$$

$$\overline{x}_2 \Rightarrow Z_2 \Rightarrow P = ?$$