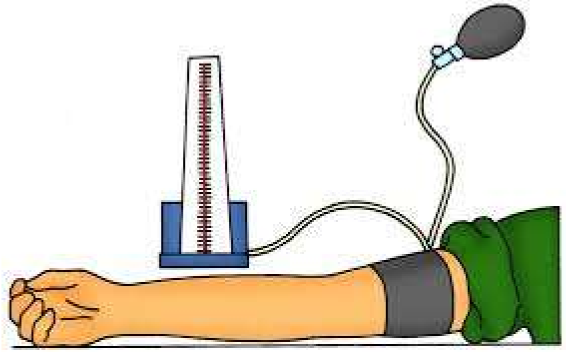
TEST D'IPOTESI

INTRODUZIONE

ESEMPIO 1



La pressione arteriosa sistolica della popolazione maschile si distribuisce pressoché normalmente con una deviazione standard $\sigma = 16$ e media $\theta = 130$.

Si vuole verificare se la pressione dei pazienti di un dato reparto ospedaliero sia diversa dalla pressione della popolazione generale

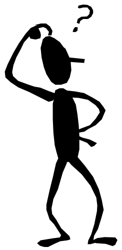
… allora

- Si prende un campione di pazienti di questo reparto (**campione casuale** dell' universo dei pazienti afferenti al reparto): n=64

- La media delle pressioni risulta 133 mmHg

- Se la pressione dei pazienti del reparto si distribuisse come la pressione della popolazione generale, la media μ dell'universo dei pazienti del reparto dovrebbe coincidere con il valore vero θ della pressione della popolazione generale

Si può affermare che i pazienti del reparto hanno un livello pressorio pari/diverso a quello della popolazione generale???



Steps in Hypothesis Testing

1. Express the question and make statements on the phenomenon (population).
2. Collect evidence (sample data) to test the corresponding statement.
3. Analyze the data to assess the plausibility of the statement.

IL TEST D'IPOTESI

Il test per verificare un'ipotesi è una regola che, basandosi su dati sperimentali, porta alla DECISIONE DI RIFIUTARE oppure NO l'ipotesi in studio.

<u>L'IDEA È:</u>

1) Stabilire $H_0 : \mu = \theta$ (ipotesi nulla)

$H_1 : \mu \neq \theta$ (ipotesi alternativa)

μ indica la media delle misurazioni registrate.

θ indica il valore vero

The statement(s)

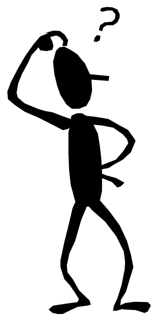
The **null hypothesis**, denoted H_0, is the statement to be tested.
The null hypothesis is a statement of no difference (or no change, or no effect) and it is assumed true until evidence indicates otherwise.

The **alternative hypothesis**, denoted H_1, is a statement that we are trying to find evidence to support.

2) Costruire una regola che consenta di rifiutare H_0 se i dati campionari non sono “consistenti” con $\mathbf{H_0}$

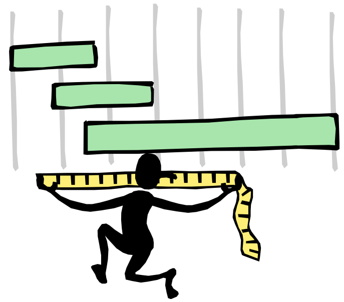
In un campione si osserva la media $\bar{x}$

Si rifiuta $\mathbf{H_0}$ se $\bar{x}$ è molto più piccola o molto più grande di θ

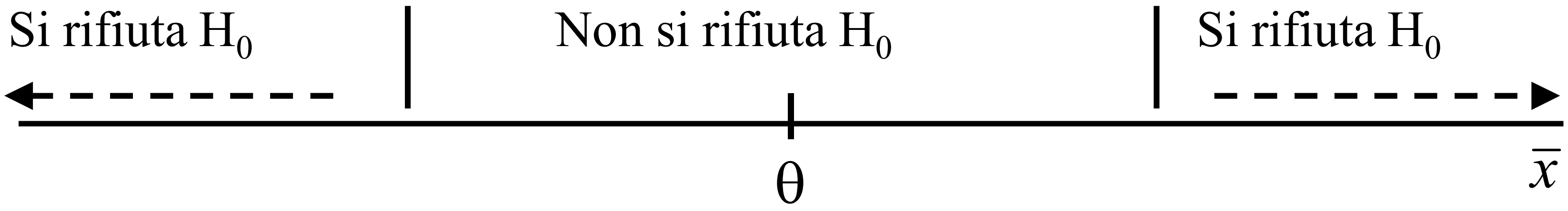


Ma quanto più grande o più piccola?

(dipende anche dalla variabilità del fenomeno)



SI DEVE SCEGLIERE UNA REGIONE CRITICA



Il problema statistico

- Anche se la pressione dei pazienti del reparto si distribuisce come la pressione della popolazione generale, è possibile che per caso si osservi una media campionaria che non è pari a θ

- Anche se la pressione dei pazienti del reparto si distribuisce in modo diverso dalla pressione della popolazione generale, è comunque possibile che per caso si osservi una media campionaria molto vicina a quella della popolazione generale (o anche molto più lontana da θ di quanto sia in realtà)

Il problema statistico

Nel definire la **zona di rifiuto**

c'è bisogno di controllare cosa può succedere grazie al "caso"

ovvero di controllare la

probabilità di sbagliare nel prendere la decisione

e questo lo si può fare!

La distribuzione campionaria della media

Per fortuna… possiamo sapere come “agisce il caso”!

Conosciamo la distribuzione teorica dei valori di $\bar{x}$ che possono essere ottenuti “sotto H_0”, cioè se fosse vera H_0 grazie al teorema del limite centrale:

$$\bar{x} \sim N(\mu = \theta, Var = \sigma^2 / n)$$

(se la numerosità campionaria è abbastanza elevata)

Possiamo allora definire la **regione critica di rifiuto** in modo da stabilire a priori la **probabilità di sbagliare** quando rifiutiamo H_0.

Questa probabilità si chiama **livello di significatività α.**

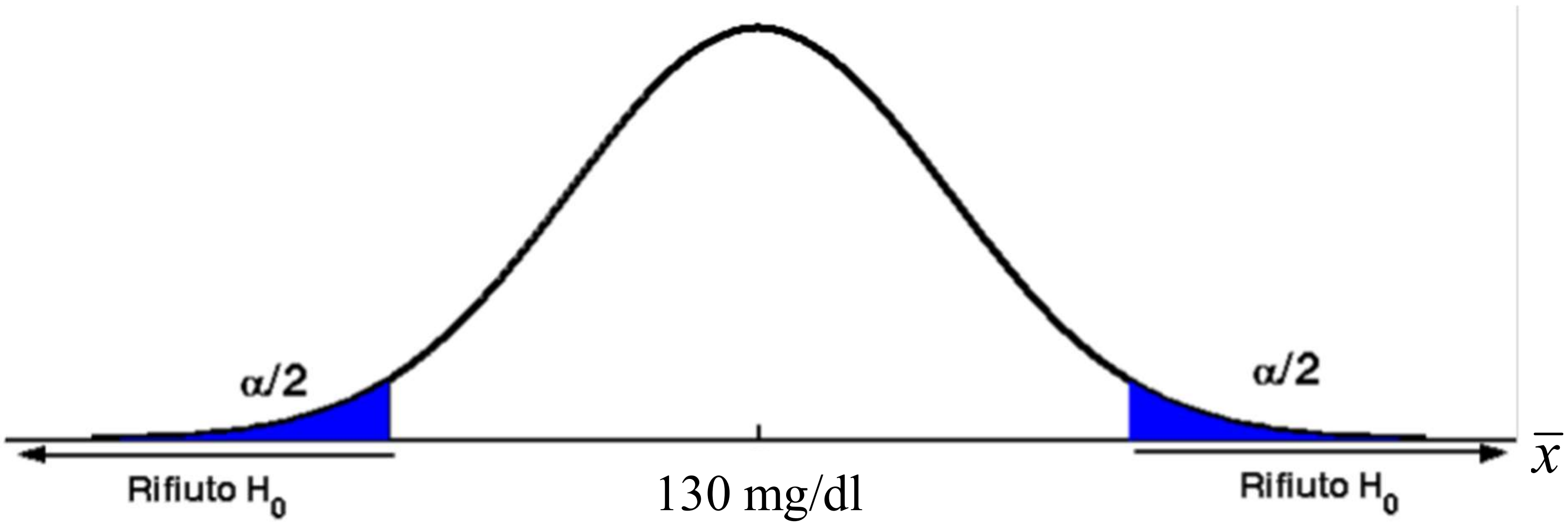
Nel nostro ***esempio*** …

$$X \sim N(\mu = \theta, \sigma^2) \Rightarrow X \sim N(130mmHg, 16^2 mmHg^2)$$

La distribuzione delle medie di campioni di numerosità n = 64 tratti dalla nostra variabile casuale X è:

$$X \sim N(130mmHg, \frac{16^2}{64} mmHg^2)$$

Distribuzione della media campionaria



α è il livello di significatività in base al quale viene definita la regione critica di rifiuto nelle due code

… ma è meglio riferirsi alla trasformata Z, ovvero **standardizzare**

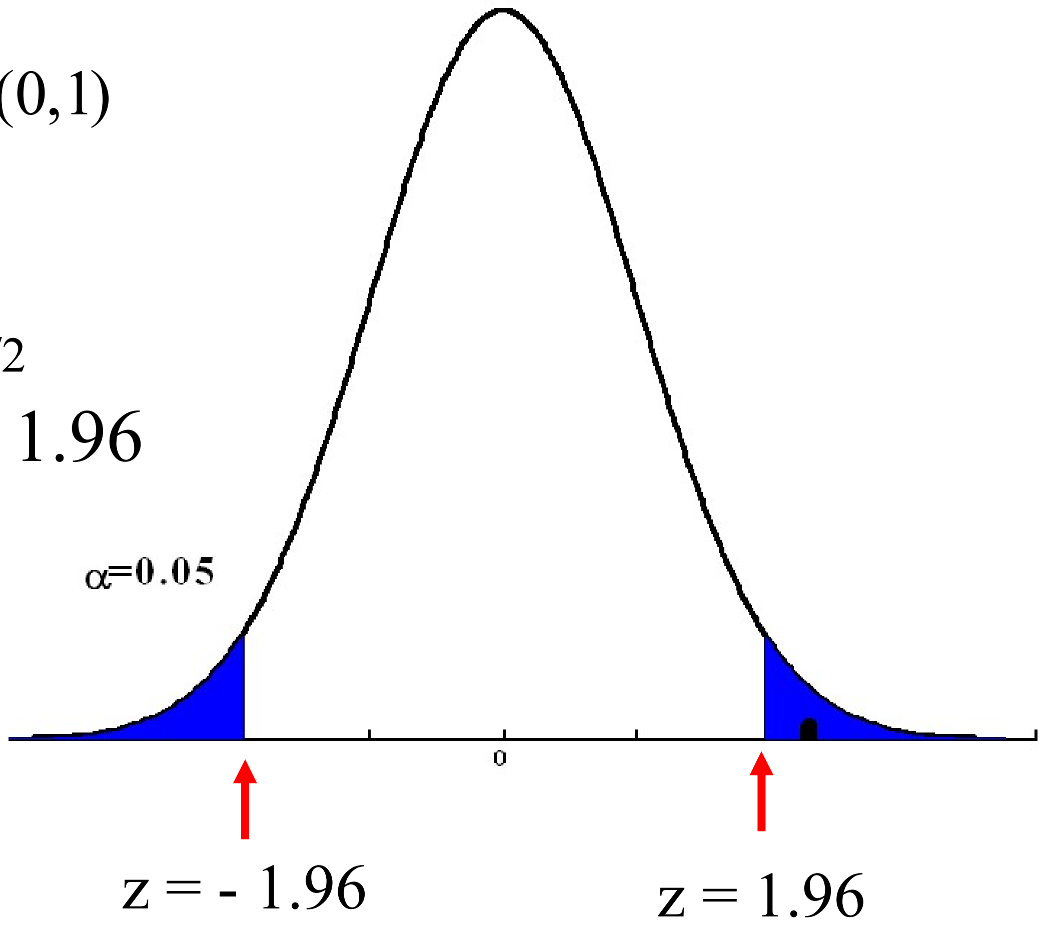
$$Z = \frac{\bar{x} - \theta}{es(\bar{x})} \qquad Z \sim N(0,1)$$

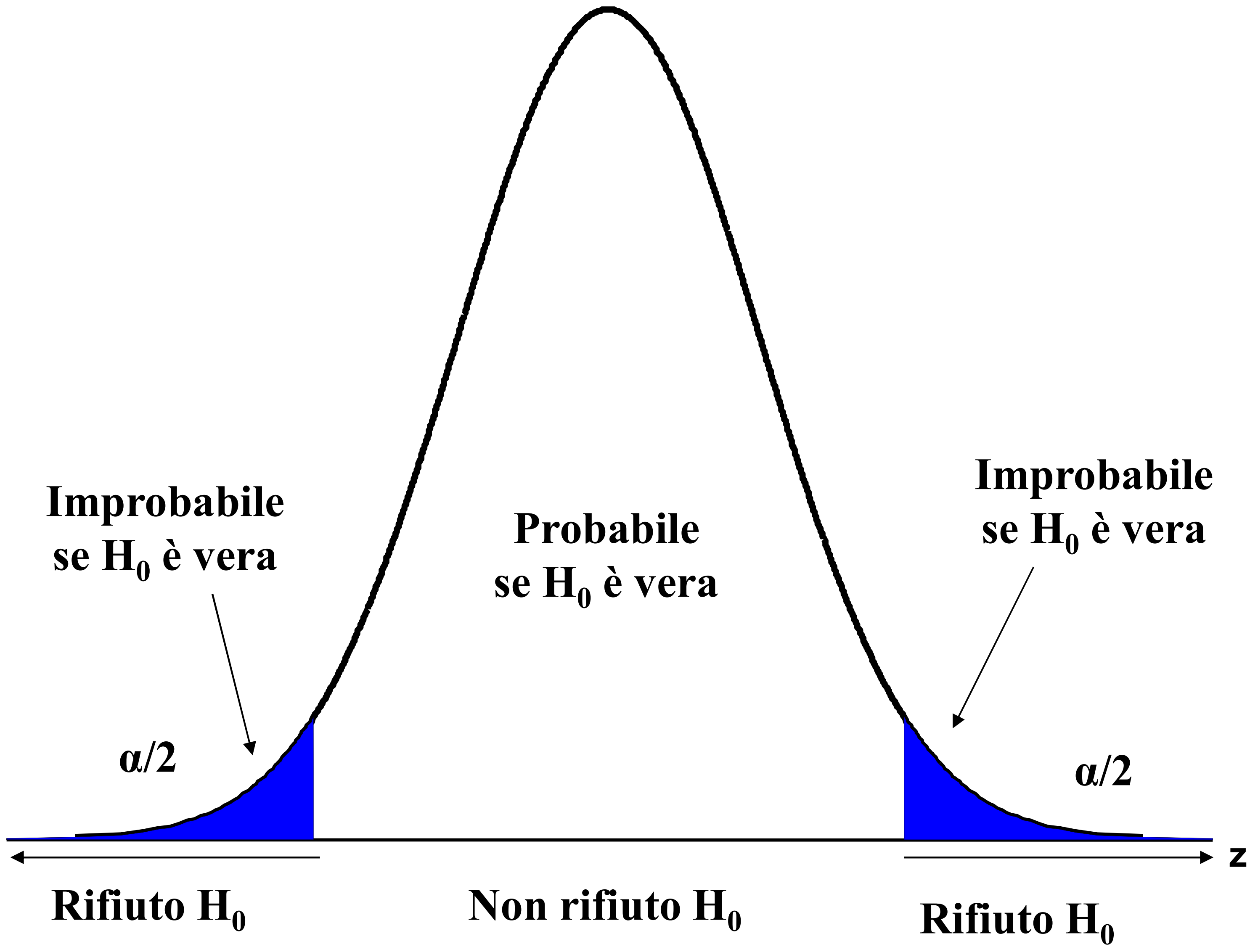
Stabilito α, la soglia è $z_{\alpha/2}$

Es.: α=0.05 ➡ $z_{\alpha/2} = 1.96$

quindi rifiuto H_0 se:

$$\left| \frac{\bar{x} - \theta}{es(\bar{x})} \right| > z_{\alpha/2}$$





NB: quando rifiuto H_0 potrebbe essere che H_0 sia vera ed è successo l'improbabile!

Il procedimento

Logico Matematico

Ipotesi

↓

Argomentazione logico-deduttiva

↓

Contraddizione (reductio ad absurdum)

↓

Conclusione (tesi)
L'ipotesi era falsa!

Logico Statistico

Ipotesi nulla H_0 sul parametro

↓

Statistica test e distribuzione campionaria sotto H_0

↓

"Distanza" tra risultato campionario e teorico atteso

↓

Conclusione
H_0 è vera e la "distanza" è casuale
H_0 è errata e la "distanza" è legata al trattamento

Il procedimento

Logico Statistico

Ipotesi nulla H_0 sul parametro

↓

Statistica test e distribuzione campionaria sotto H_0

↓

"Distanza" tra risultato campionario e teorico atteso

↓

Conclusione
H_0 è vera e la "distanza" è casuale
H_0 è errata e la "distanza" è legata al trattamento

Operativo

Formulare H_0

↓

Calcolare la statistica test sui dati

↓

Calcolare la plausibilità di H_0 visti i dati

↓

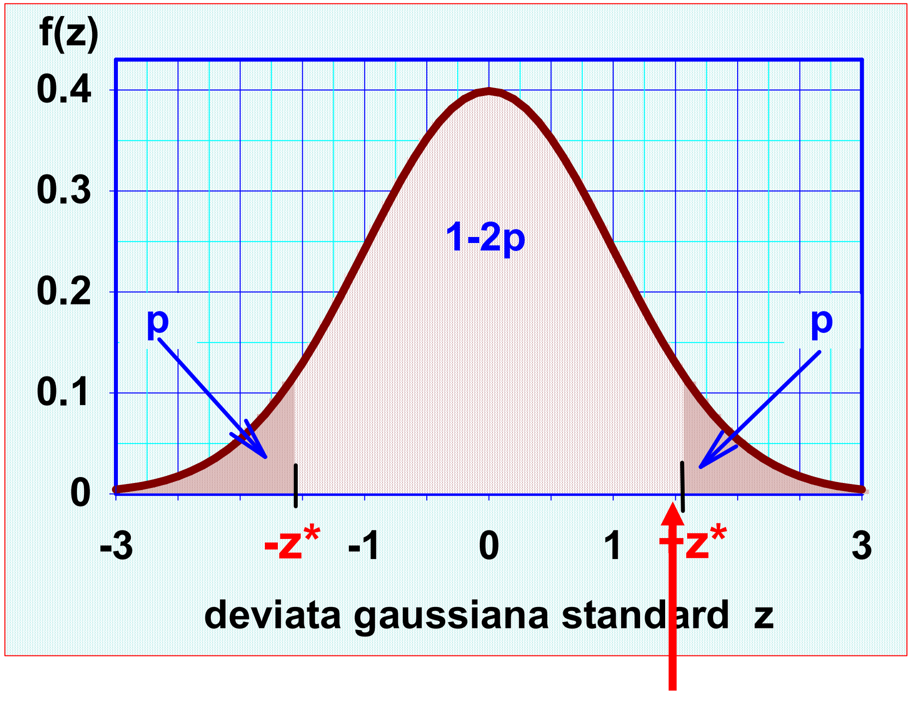
Conclusione
Non rifiuto H_0
Rifiuto H_0

- La logica del test è basata sulla confutazione di un'ipotesi specifica, H_0

- Rispetto ad un'ipotesi specifica posso trovare una specifica distribuzione di campionamento

- L'ipotesi alternativa contiene, invece, un'infinità di valori $\mu \neq \theta$ e quindi di relative distribuzioni di campionamento

Se la pressione media è la stessa mi attendo che i valori **z** siano prossimi a 0 (valore atteso di z), mentre i valori molto discosti da 0 sono improbabili sotto H_0.



Nel nostro ***esempio***:

$$z = \frac{133-130}{16/\sqrt{64}} =$$

$$= 3/2 = 1.5$$

Da cui: $pr\{|z| > 1.50\} = 0.06681 \cdot 2 = 0.13362$

p = 0.134 **NON rifiuto H_0**

Attenzione: il valore di p <u>non indica</u> la probabilità che l'ipotesi nulla sia vera …

… ma la probabilità di osservare un risultato come quello ottenuto (o più estremo) se fosse vera l'ipotesi nulla.

Significato di p=0.134 (livello di probabilità sotto H_0)

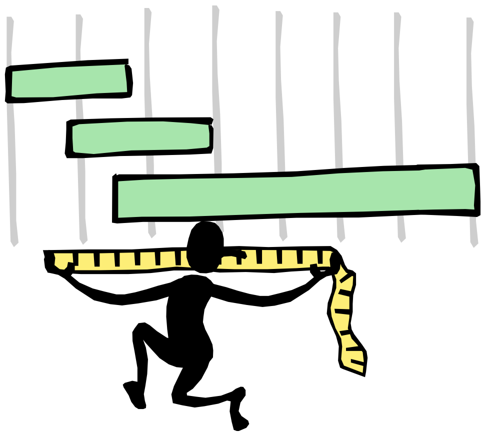
Se la pressione media fosse la stessa, un risultato campionario uguale o più estremo (nella coda della distribuzione) di quello osservato nel campione ($\bar{x}$ =133 mmHg) si verificherebbe 13 volte su 100.

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

(nell'esempio piuttosto debole)

- È possibile che la pressione media del reparto sia diversa da quella della popolazione generale, ma i dati non lo evidenziano (p=0.134)

- È **plausibile** che la pressione del reparto venga dalla distribuzione della pressione della popolazione generale



l'esperimento suggerisce che

la pressione non sia diversa da quella della popolazione generale

Intervallo di confidenza

Calcoliamo l'intervallo di confidenza al 95% della media

$$I.C._{95\%} = \bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$$

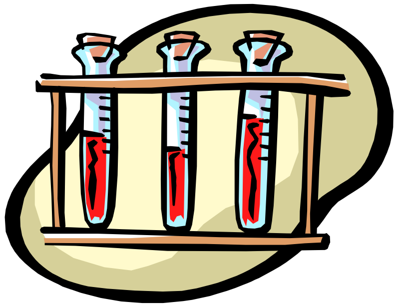
$$I.C._{95\%} = 133 \pm 1.96 \times \frac{16}{\sqrt{64}} =$$

$$= 133 \pm 3.92 = (129.08, 136.92)$$

Conclusione

- Possiamo dire con una buona (95%) confidenza che l'ignoto parametro μ è compreso tra 129.08 e 136.92

- Poiché questo intervallo contiene come valore plausibile per μ il valore 130 mmHg, questo equivale a dire che i nostri dati sono compatibili con l'ipotesi nulla



ESEMPIO 2

Un laboratorio è stato fornito di un nuovo strumento per determinare la quantità di colesterolo nel sangue.

Tale strumento misura con:

- Imprecisione pari a $\sigma = 7.0$ mg/dl
- Le misure (o equivalentemente gli errori di misura) hanno distribuzione gaussiana

Voglio dimostrare se lo strumento è accurato

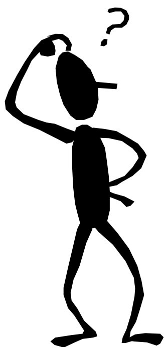
… allora

•Prendo un campione di sangue (standard) di cui conosco la concentrazione di colesterolo ($\theta = 180.0$ mg/dl)

•Lo misuro $n = 25$ volte con il nuovo strumento (**campione casuale** dell' universo delle misure di θ)

•La media delle misure risulta $\overline{x} = 183.5$ mg/dl

•Se il **metodo** è **accurato**, la media μ dell’universo di misure coincide con il valore vero θ



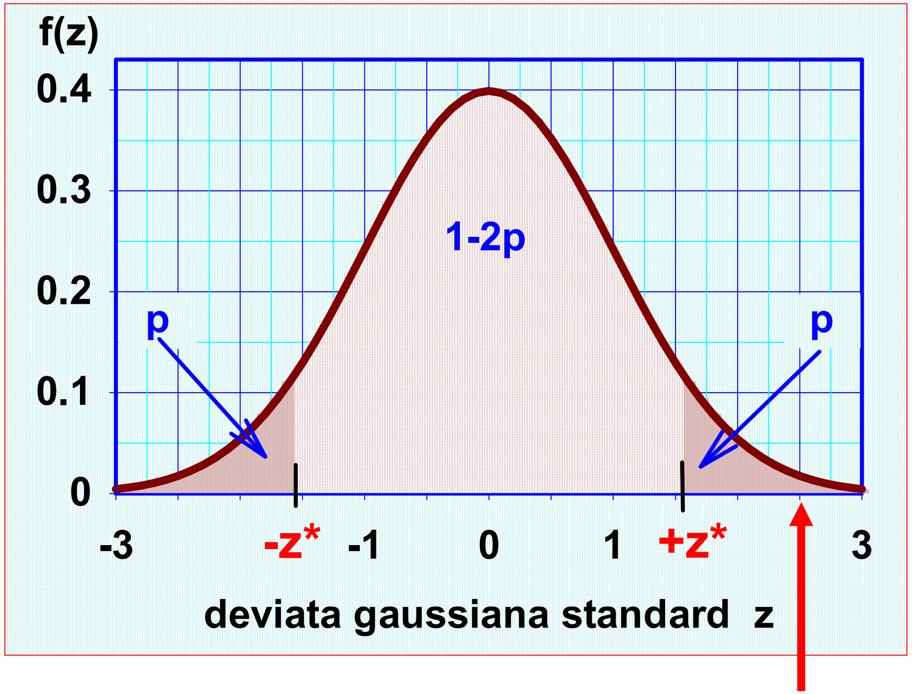
È possibile che un metodo del tutto accurato fornisca tale risultato?

Se il metodo è accurato mi attendo che i valori **z** siano prossimi a 0 (valore atteso di z), mentre i valori molto discosti da 0 sono improbabili sotto H_0.

Nel nostro ***esempio***:

$$z = \frac{183.5 - 180.0}{7.0/\sqrt{25}} = $$

$$= 3.5/1.4 = 2.5$$



z = 2.5

Da cui: $pr\{|z| > 2.50\} = 0.00621 \cdot 2 = 0.01242$

p = 0.012 **rifiuto H_0**

Attenzione: il valore di p <u>non indica</u> la probabilità che l'ipotesi nulla sia vera …

… ma la probabilità di osservare un risultato come quello ottenuto (o più estremo) se fosse vera l'ipotesi nulla.

Significato di p=0.012 (livello di probabilità sotto H_0)

Se in realtà il metodo fosse accurato, un risultato campionario uguale o più estremo (nella coda della distribuzione) di quello osservato nel campione ($\bar{x}$ =183.5 mg/dl) si verificherebbe 12 volte su 1000.

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

Intervallo di confidenza (IC)

Calcoliamo l'intervallo di confidenza al 95% della media

$$I.C._{95\%} = \bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$$

$$I.C._{95\%} = 183.5 \pm 1.96 \times \frac{7.0}{\sqrt{25}} =$$

$$= 183.5 \pm 2.74 = (180.76, 186.24)$$

Cosa ci dice IC?

- Possiamo dire con una buona (95%) confidenza che l'ignoto parametro μ è compreso tra 180,76 e 186,24 mg/dl

- Poiché questo intervallo…..

Caution:

We never “accept” the null hypothesis, because, without having access to the entire population, we don’t know the exact value of the parameter stated in the null.

Rather, we say that we do not reject the null hypothesis.

This is just like the court system: we never declare a defendant “innocent”, but rather say the defendant is “not guilty”.

Relazione tra test d'ipotesi e IC

Nel fare il test, se: $x_S^* < \overline{x} < x_D^*$ NON RIFIUTO H_0

(gli estremi sono su scala originale)

So che la probabilità che ciò si verifichi è (1-α)

Intervallo di probabilità

$$\theta - z_{\alpha/2}\,\sigma/\sqrt{n} < \overline{x} < \theta + z_{\alpha/2}\,\sigma/\sqrt{n}$$

Se lo scrivo in funzione di $\overline{x}$ ho:

$$\overline{x} - z_{\alpha/2}\,\sigma/\sqrt{n} < \theta < \overline{x} + z_{\alpha/2}\,\sigma/\sqrt{n}$$

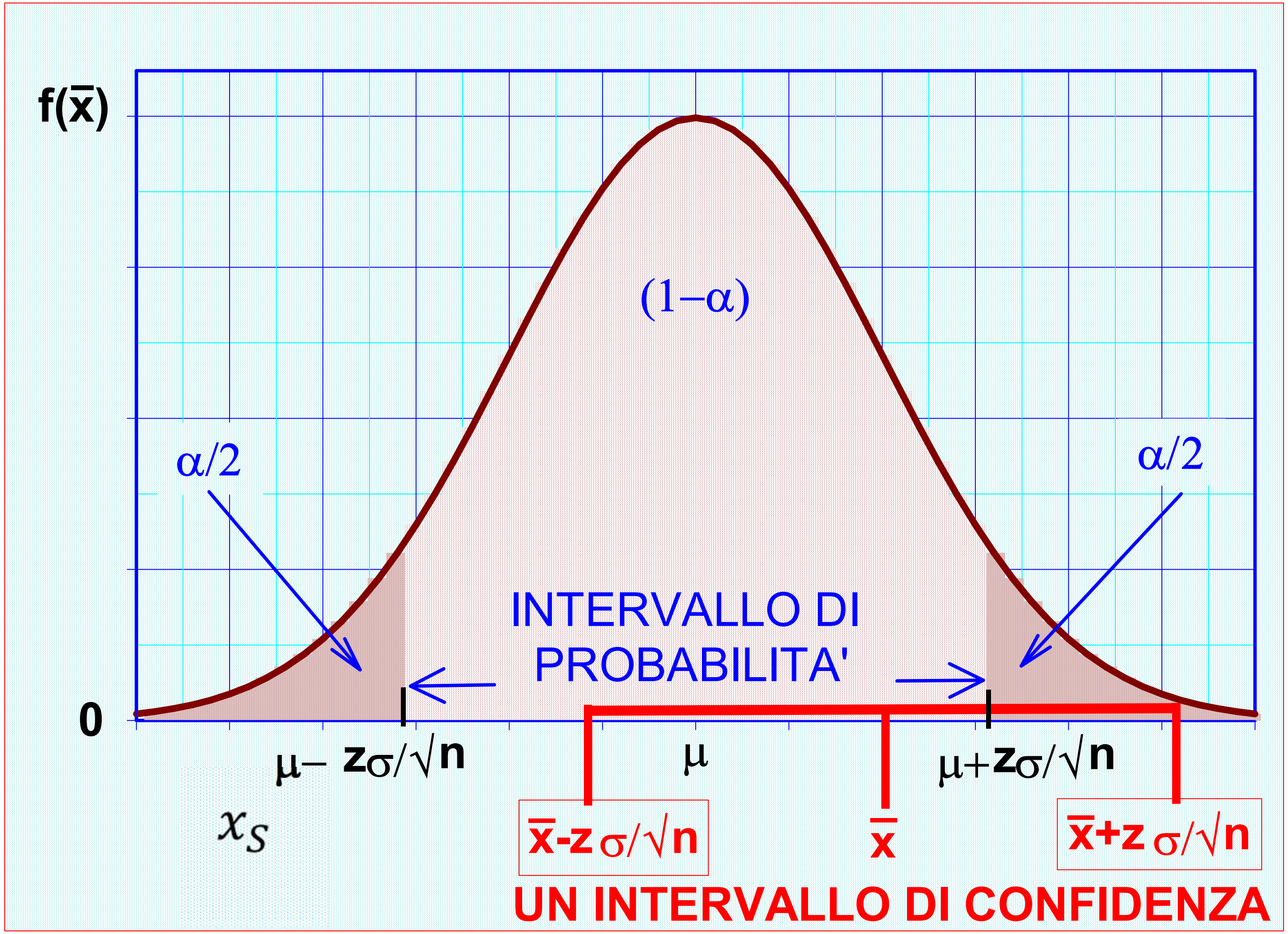
che è l'intervallo di confidenza al 95% per il parametro μ (se α=0.05)

Quindi, se non ho rifiutato H_0: μ=θ, **al livello di significatività α,**

ne deriva che l'IC a (1- α)% **include μ=θ**

Viceversa: se RIFIUTO H_0 , l'IC a (1- α)% **non include** il valore μ=θ

Relazione tra test d'ipotesi e IC



Nel nostro ESEMPIO 1….

NON rifiuto $\mathbf{H_0}$ $\mu = \theta = 130$ a livello di significatività 5% e **p-value = 0.134**

Un campione di 64 soggetti, σ=16 mmHg e $\bar{x} = 133\,mmHg$

IC al 95% della media

$$I.C._{95\%} = \bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$$

$$133 - 1.96 \cdot 16/\sqrt{64} < \mu < 133 + 1.96 \cdot 16/\sqrt{64}$$

$$129.08 < \mu < 136.92$$

Riassumendo:

Test su H_0 e IC: due lati della stessa medaglia!

1) Dato il valore campionario:

$$z = \frac{\bar{x} - \theta}{\sigma / \sqrt{n}} = 1.50 \; e \; \Pr\{|z| > 1.50\} = 0.134$$

"p-value"=0.134 ovvero p>0.05

Se fosse vera H0, un risultato uguale o più estremo di quello ottenuto si verificherebbe più di 5 volte su 100 (13 volte su 100)

2) IC al 95% per μ è (129.08; 136.92)

In base all'osservazione campionaria, il valore vero del parametro μ sarebbe compreso tra 129 e 137, con la probabilità che questa affermazione sia FALSA 5 volte su 100

Viceversa nell'ESEMPIO 2 …

Rifiuto H_0 : μ=θ =180mg/dl a livello di significatività del 5%

e p-value=

Un campione di 25 rilevazioni, con $\overline{x} =$ 183,5 e σ = 7.0 mg/dl

I.C. al 95%

$$= 183.5 \pm 1.96 \times \frac{7.0}{\sqrt{25}} =$$

$$= 183.5 \pm 2.74 = (180.76, 186.24)$$

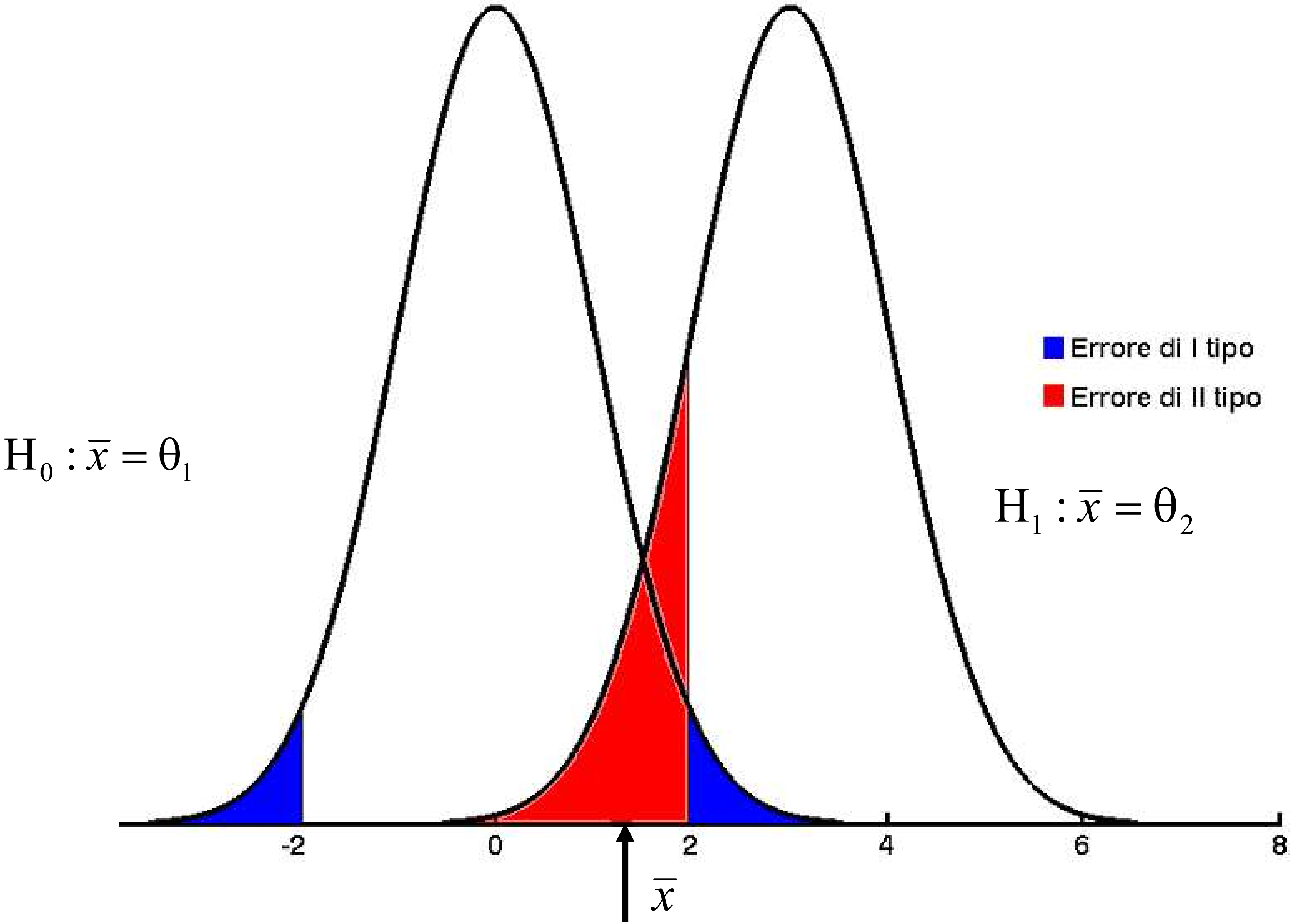
TEST d’IPOTESI

SIGNIFICATIVITA’
e
POTENZA del TEST

Criterio di decisione

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

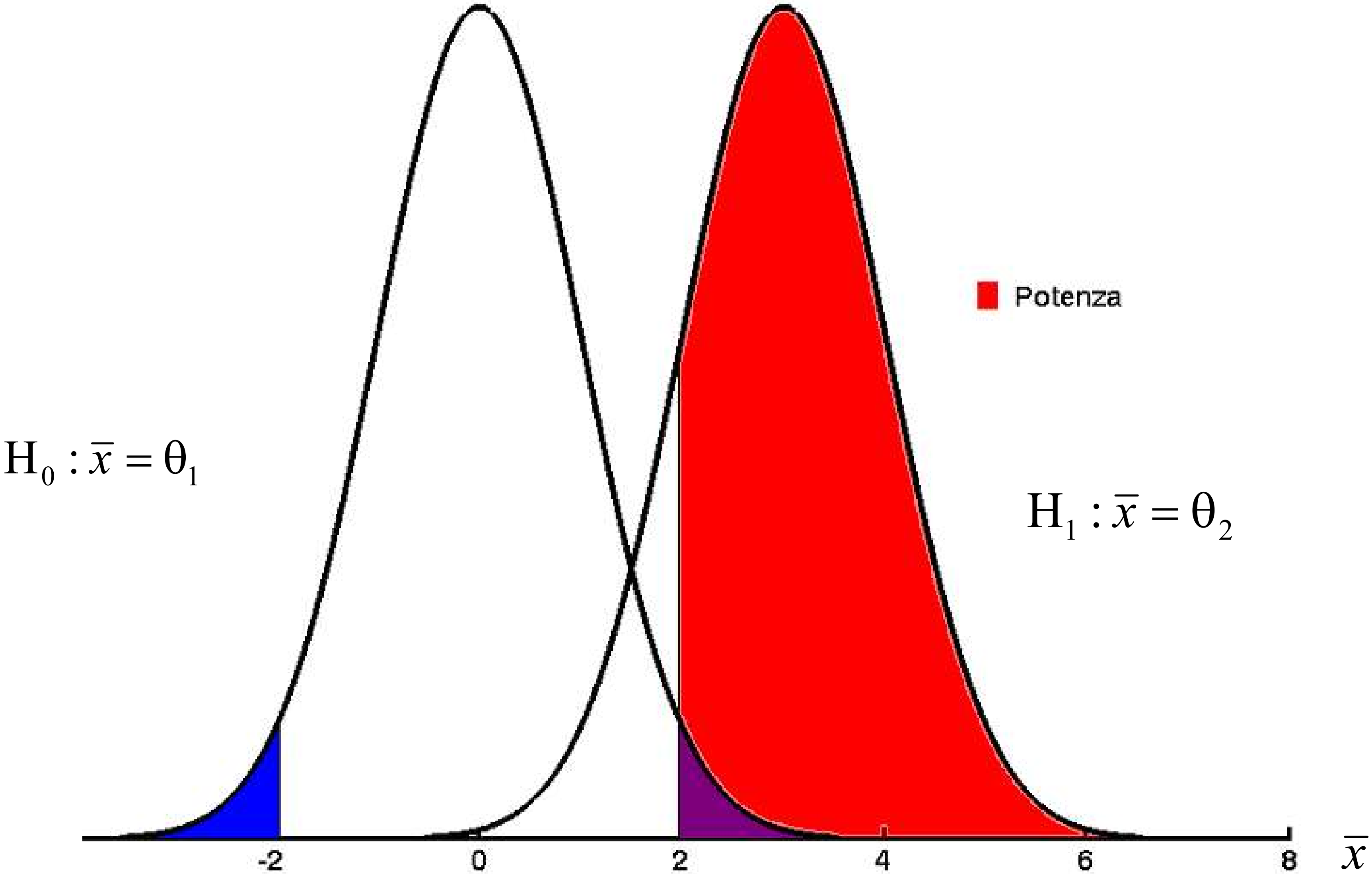
Errore di secondo tipo



Criterio di decisione

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

Potenza del test



Rischio di errore di tipo I (α):

Probabilità di rifiutare H_0 quando è vera H_0

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

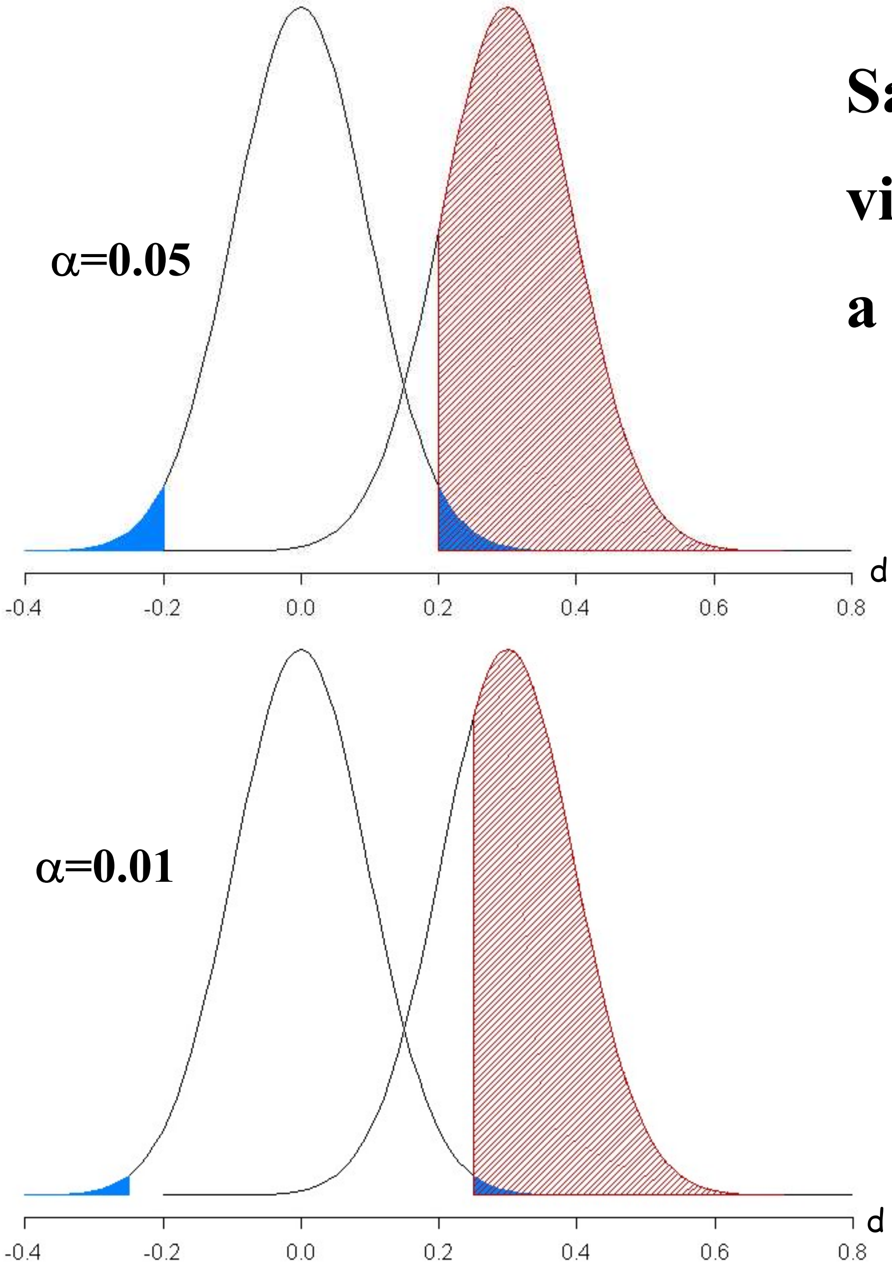
Di solito si fissa $\leq 5\%$

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

Probabilità di rifiutare H_0 quando è vera una specifica H_1

es. si conclude che B differisce da A quando effettivamente B è meglio o peggio di A.

Di solito si fissa $\geq 80\%$



Sarebbe ottimale avere l'errore α vicino a 0 e la potenza $(1-\beta)$ vicino a 1

Ma succede che:

se $\alpha \downarrow$, allora la potenza $\downarrow$

come mostrato in figura

(dall'alto al basso)

Oppure:

se la potenza $\uparrow$, allora $\alpha \uparrow$

(dal basso all'alto)

Correct interpretation of statistical testing results

It requires that we examine:

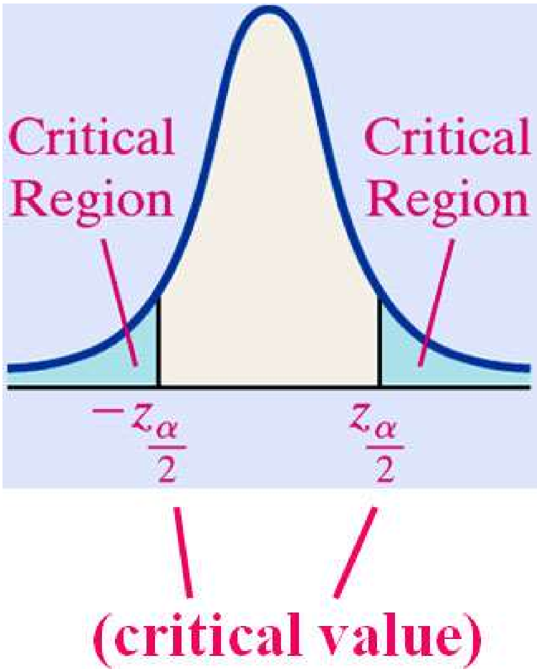
- The effect estimates
- The interval estimates, I.e. the CI
- We calculate the exact p-value
- We critically examine the statistical assumptions
- More broadly, we examine the hidden assumptions about how results were generated and chosen for presentation.

TEST a DUE CODE

Rejection region, given α

Use the N(0,1) table to determine the critical values.

Two-Tailed

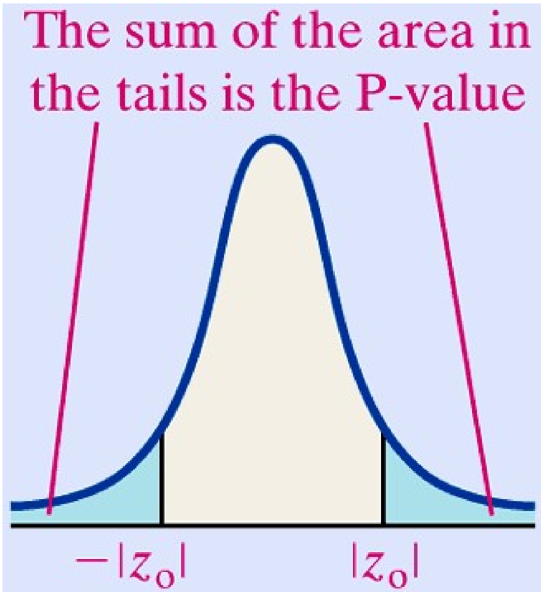


TEST a DUE CODE e p-value

Use N(0,1) Table to estimate the *P*-value.

Z_0 is the test statistics

Two-Tailed



$$P - \text{value} = 2P(z > |z_0|)$$

The system of Hypothesis

According to a study published in March 2006, the mean length of a call on a cellular telephone was 3.25 minutes. A researcher believes that the mean length of a call has increased since then.

The hypothesis deals with a population mean, μ. If the mean call length on a cellular phone is no different than in 2006, it will be 3.25 minutes so the **null hypothesis is H_0: $\mu = 3.25$.**

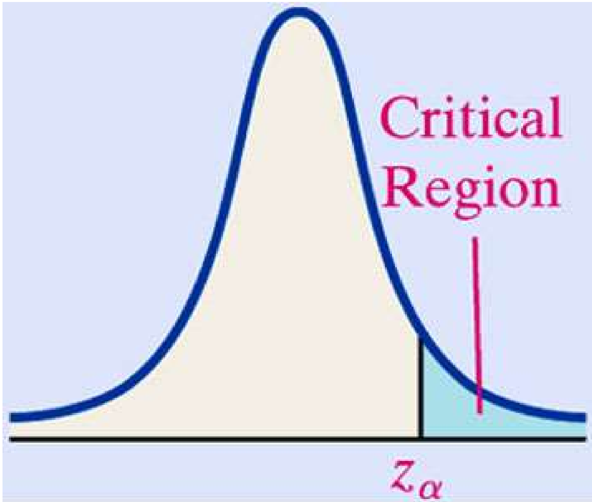
Since the researcher believes that the mean call length has increased, the **alternative hypothesis is:**
H_1: $\mu > 3.25$, a right-tailed test.

Right-tailed test

Rejection region, given α

Use the N(0,1) table to determine the critical values

Right-Tailed



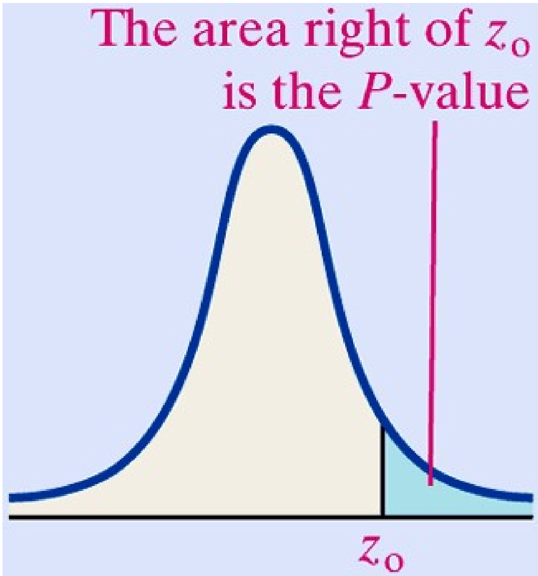
(critical value)

Hypothesis Tests – p-value

Use N(0,1) Table to estimate the *P*-value.

Z_0 is the test statistics

Right-Tailed



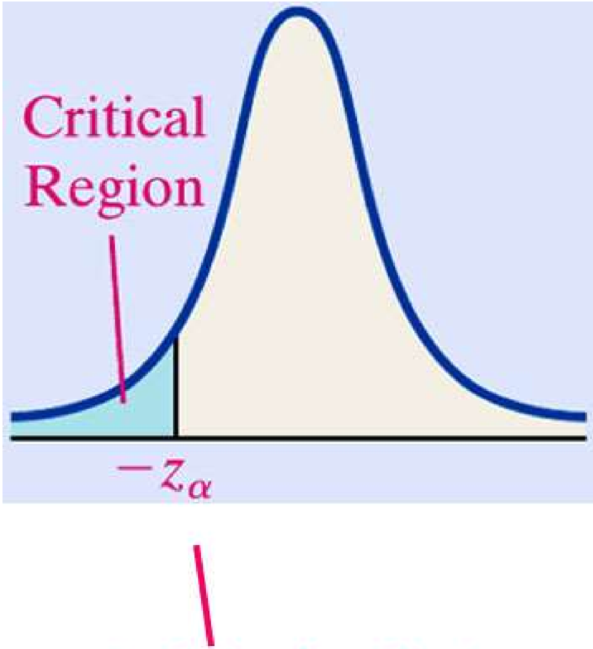
$$P - \text{value} = P(z > z_0)$$

Left-tailed test

Rejection region, given α

Use the N(0,1) table to determine the critical values

Left-Tailed



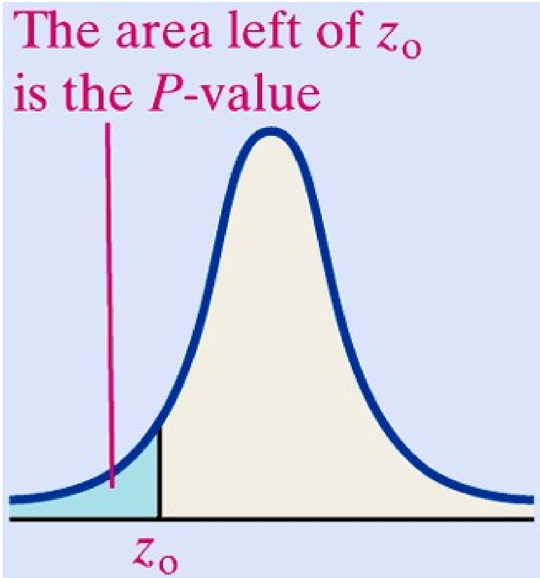
(critical value)

Left-tailed test – P-value

Use N(0,1) Table to estimate the *P*-value.

Z_0 is the test statistics

Left-Tailed



$$P - \text{value} = P(z < z_0)$$