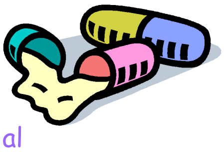
Test d'ipotesi

TEST D'IPOTESI

In medicina una delle più utilizzate tecniche inferenziali è quella nota come *test d'ipotesi.*

Tale procedura è particolarmente utile in situazioni in cui noi siamo interessati a *prendere decisioni tra due o più alternative possibili*, piuttosto che alla stima del valore di uno o più parametri.

Ad esempio



- valutare l'efficacia di un nuovo farmaco rispetto al placebo
- valutare se il trattamento chirurgico di un particolare tumore in una data fase allunga la vita dei pazienti rispetto al trattamento chemioterapico
- valutare se l'esposizione a una determinata sostanza chimica è responsabile di un eccesso di tumori

In tali situazioni la valutazione dell'alternativa migliore è finalizzata a decidere quale intervento operare sulla realtà
(scelta del farmaco, tipo di terapia, tipo di intervento preventivo)

TEST D'IPOTESI

- Risultato ottenuto da un'indagine o un esperimento come uno dei possibili risultati di un modello probabilistico

variabilità biologica

variabilità campionaria

errore di misura

- Il test d'ipotesi ci fornisce un criterio per decidere se un campione qualsiasi appartiene alla classe dei 'PROBABILI' o degli 'IMPROBABILI', in base a quel modello probabilistico

probabilità P calcolata tramite un TEST STATISTICO

TEST D'IPOTESI

H_0: IPOTESI NULLA
Tutte le differenze osservate sono delle semplici fluttuazioni casuali

H_1: IPOTESI ALTERNATIVA
Le differenze riscontrate nelle statistiche campionarie rispecchiano una reale differenza nei parametri delle popolazioni corrispondenti

Esempio:

La glicemia dei diabetici italiani è uguale alla glicemia dei diabetici americani

La glicemia dei diabetici italiani è diversa dalla glicemia dei diabetici americani

Dati del campione

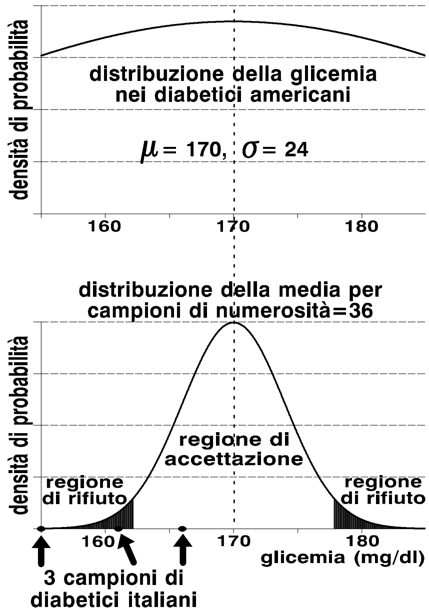
Test statistico

P>0,05 = la probabilità che le differenze osservate siano dovute al caso è superiore al 5%

<u>Non</u> rifiuto l'ipotesi nulla (H_0)
tutte le differenze osservate tra i campioni possono essere attribuite al caso

P<0,05 = la probabilità che le differenze osservate siano dovute al caso è inferiore al 5%

Rifiuto l'ipotesi nulla (H_0)
le differenze osservate tra i campioni rispecchiano delle differenze reali tra le popolazioni



1. Formulazione dell'ipotesi da verificare

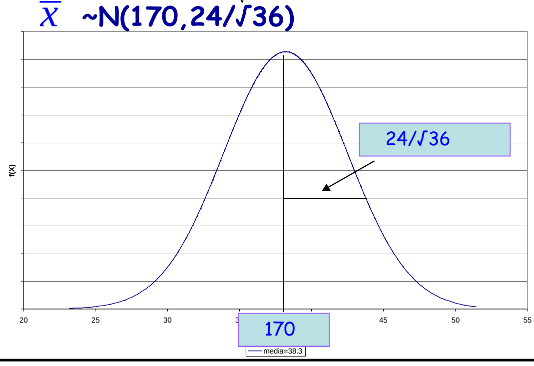
A.(H_0 ipotesi nulla) — **Spiega le differenze osservate come dovute al caso**

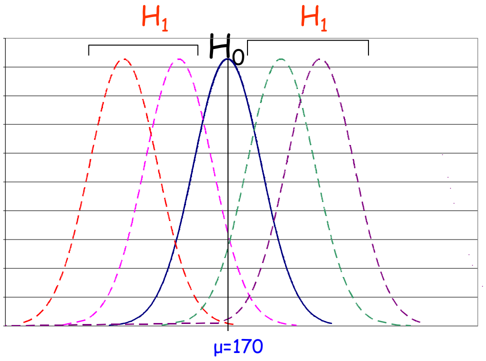
H_0: $\mu=\mu_0=170$ mg/dl

La media della popolazione italiana da cui proviene il campione (μ) è identica alla media della popolazione americana (μ_0)

L'ipotesi è <u>anche</u> una congettura sulla distribuzione campionaria, infatti

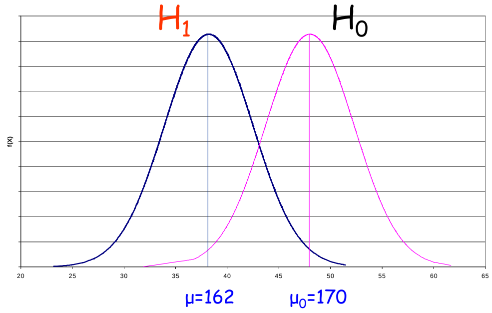
$\overline{X}$ **~N(170,24/√36)**





Ipotesi alternativa

$H_1: \mu \neq \mu_0$

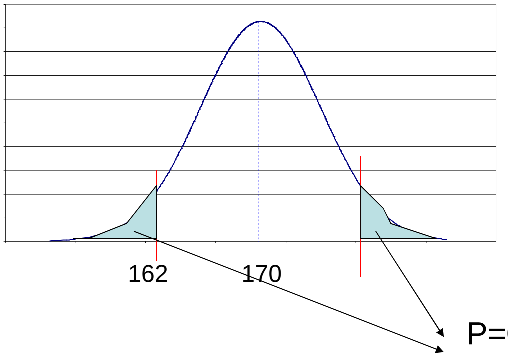


$\mu=162$

TEST STATISTICO

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

$$z = \frac{162 - 170}{24 / \sqrt{36}} = -2$$



Rifiuto H_0, la media della glicemia nella popolazione italiana è diversa da quella della pop. americana

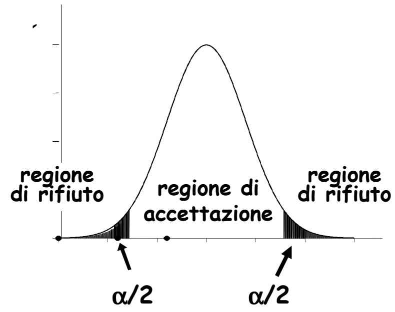
P=0,046 → P <0,05

ALTRE FORMULAZIONI DELL'IPOTESI NULLA

Test bidirezionale

$H_0: \mu = \mu_0$

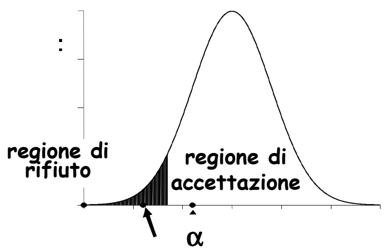
$H_1: \mu \neq \mu_0$



Test unidirezionale

$H_0: \mu = \mu_0$

$H_1: \mu < \mu_0$



Effettuazione di un TEST D'IPOTESI

1. Formulazione H_0 e H_1
2. Scelta del test statistico
3. Calcolo del test statistico

$$\text{test} = \frac{\hat{\theta} - \theta_0}{ES[\hat{\theta}]}$$

dove

$\hat{\theta}$ = stima del parametro di interesse calcolata sui dati campionari

θ_0 = valore del parametro di interesse sotto l'ipotesi nulla H_0

$ES[\hat{\theta}]$ = errore standard dello stimatore calcolato sotto H_0

Effettuazione di un TEST D'IPOTESI-*continua*

4. Test P probabilità di ottenere un risultato come quello osservato o più estremo per motivi casuali

5. Rifiuto (se P è bassa,<0,05) o non rifiuto H_0 (se P non è bassa, >0,05)

TEST D'IPOTESI- confronto media campionaria con media popolazione – campione grande

$H_0: \mu_0 = \mu$

$H_1: \mu_0 \neq \mu$

Test statistico:
$$z = \frac{\bar{x} - \mu_0}{e.s.(\bar{x})} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

(vedi esempio diabete)

TEST D'IPOTESI- confronto media campionaria con media popolazione – campione piccolo e σ ignota

H_0:μ_0=μ

H_1:$\mu_0 \neq \mu$

Test statistico:

$$t = \frac{\bar{x} - \mu_0}{e.s.(\bar{x})} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

t con (n-1) gradi di libertà

Assunzioni:

1- Il campione è stato selezionato casualmente dalla popolazione

2- La variabile è distribuita normalmente nella popolazione

TEST D'IPOTESI- per dati appaiati

H_0:δ=0

H_1:$\delta \neq 0$

Test statistico:

$$t = \frac{\bar{d} - 0}{e.s.(d)} = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

con (n-1) gradi di libertà

Assunzioni:

1- I campioni appaiati sono selezionati casualmente dalla popolazione

2- La popolazione delle differenze è distribuita normalmente

TESTING FOR EQUALITY OF MEANS: PAIRED DATA

Table 7.3. Age at First Word in Months for 10 Children with Cyanotic Heart Disease and Their 10 Siblings

Pair Number	Cyanotic x_1	Sibling x_1	Difference d
1	11.8	9.8	2.0
2	20.8	16.5	4.3
3	14.5	14.5	0.0
4	9.5	15.2	- 5.7
5	13.5	11.8	1.7
6	22.6	12.2	10.4
7	11.1	15.2	- 4.1
8	14.9	15.6	- 0.7
9	16.5	17.2	- 0.7
10	16.5	10.5	6.0

$$\bar{d} = 1.32$$

$$s_d^2 = 22.488$$

Es. test t per dati appaiati

H_0: $\delta=0$

H_1: $\delta\neq0$

Test statistico: $$t = \frac{\bar{d}-0}{\frac{s_d}{\sqrt{n}}} = \frac{1{,}32-0}{\frac{4{,}74}{\sqrt{10}}} = 0{,}880$$

t con (n-1)=9 gradi di libertà

$t_{(\nu,\alpha/2)}=t_{(9,2.5\%)}=2{,}262$ soglia critica

$0{,}880<2{,}262$ non rifiuto H_0

L'età alla prima parola per i bambini cianotici sembrerebbe essere la stessa dei loro fratelli

TEST D'IPOTESI- confronto tra medie – campioni grandi

$H_0: \mu_1 - \mu_2 = 0$

$H1: \mu_1 - \mu_2 \neq 0$

Test statistico:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{e.s.(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Assunzioni:

1- I campioni sono sufficientemente grandi: $n_1 \geq 30$ e $n_2 \geq 30$

2- I due campioni sono selezionati casualmente e indipendentemente dalla popolazione

Esercizio 1:

In uno studio sull'età al menarca, condotto tra le donne statunitensi, si ottennero le seguenti informazioni per le donne di età compresa tra i 21 e i 30 anni e fra i 31 e i 40 anni:

	Donne di età tra i 31 e i 40	Donne di età tra i 21 e i 30
	1	2
n	66	78
x	13.88	12.42
s^2	1.924	1.156

Si può affermare che l'età al menarca è inferiore nelle donne più giovani?

Si vuole verificare il seguente sistema d'ipotesi:
$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

$ES(\overline{X}_1 - \overline{X}_2) = \sqrt{(1.924/66)+(1.156/78)} = 0.2101$

$$Z = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{ES(\overline{x}_1 - \overline{x}_2)} = \frac{13.88 - 12.42}{0.2101} = 6.95 \Longrightarrow$$

p <0.001 (errore I° tipo nel rifiuto di Ho)

z=6.95 e pertanto c'è un notevole margine di sicurezza per affermare che l'età al menarca è differente nelle donne più giovani che non nelle donne più anziane

L'intervallo di confidenza al 95% della differenza tra le due medie campionarie è pari a:

$$(\overline{x}_1 - \overline{x}_2) \pm z_{\alpha/2} \cdot ES(\overline{x}_1 - \overline{x}_2) \longrightarrow (\overline{x}_1 - \overline{x}_2) \pm 1.96 \cdot ES(\overline{x}_1 - \overline{x}_2)$$

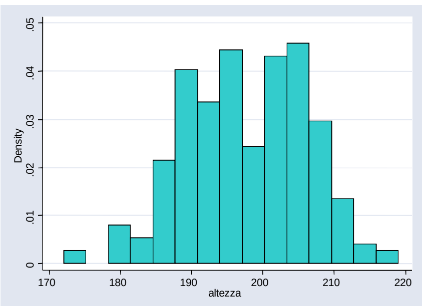
(13.88-12.42)±1.96*0.2101

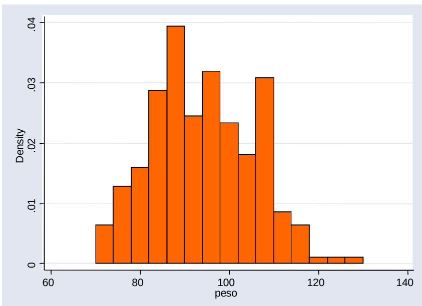
1.46 ±0.41

da 1.05 a 1.87 anni

ESEMPIO 2

Esiste una differenza nell'altezza e nel peso nei giocatori di basket del campionato italiano in base alla nazionalità?



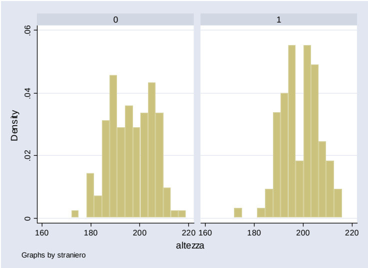


Altezza

	N	media	s
Italiani	133	196.50 cm	9.04 cm
Stranieri	104	199.05 cm	7.75 cm

t = -2.3013 P = 0.0222

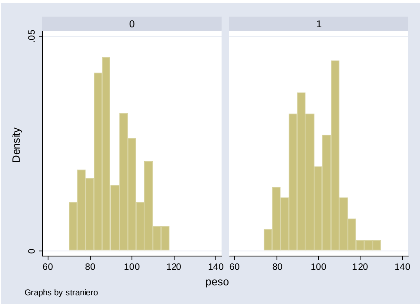
I giocatori stranieri sono più alti in modo statisticamente significativo dei giocatori italiani



Peso

	N	media	s
Italiani	133	91.71 kg	11.1 kg
Stranieri	104	98.74 kg	11.0 kg

t = -4.83 P = 0.0000001



TEST D'IPOTESI- confronto tra medie – piccoli campioni e σ ignota

$H_0: \mu_1 - \mu_2 = 0$

$H1: \mu_1 - \mu_2 \neq 0$

Test statistico: $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - 0}{e.s.(\bar{x}_1 - \bar{x}_2)} = \dfrac{\bar{x}_1 - \bar{x}_2}{s_{pooled}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

$$s^2_{pooled} = \frac{(n_1 - 1) * s_1^2 + (n_2 - 1) * s_2^2}{n_1 + n_2 - 2}$$

t con (n1+n2-2) g.l.

Assunzioni:

1- Le popolazioni da cui sono selezionati i campioni hanno una distribuzione approssimativamente normale della variabile studiata.

2- I campioni sono selezionati casualmente e indipendentemente dalle due popolazioni

3- Le varianze delle due popolazioni sono uguali.

Comparison of birth weights of children born to 15 non-smokers with those of children born to 14 heavy smokers.

Birth weight (Kg)	
Non-smokers	**Heavy smokers**
3.99	3.18
3.79	2.84
3.60	2.90
3.73	3.27
3.21	3.85
3.60	3.52
4.08	3.23
3.61	2.76
3.83	3.60
3.31	3.75
4.13	3.59
3.26	3.63
3.54	2.38
3.51	2.34
2.71	
$\bar{x}$ = 3.5933	$\bar{x}$ =3.2029
s =0.3707	**s** =0.4927
n= 15	**n**= 14

$$s_{pooled} = \sqrt{\left[\frac{14\times 0.3707^2 + 13\times 0.4927^2}{(15+14-2)}\right]} = 0.4337kg$$

$$t = \frac{(3.5933-3.2029)}{0.4337\sqrt{(1/15+1/14)}} = \frac{0.3904}{0.1612} = 2.42, d.f. = 15+14-2 = 27$$

$t_{0.025;27} = 2.05$

2.42>2.05 → Rifiuto H0

i pesi dei bimbi delle forti fumatrici presentano una differenza statisticamente significativa rispetto a quelli delle non fumatrici

esempio2

Altezza del salto (POTENZA MUSCOLARE)	media ± ds
18 sportivi che fanno uso di creatina	52.3 cm ±2.1
18 sportivi che non ne fanno uso	50.9 cm ±1.7

Con il **test statistico** ci chiediamo se la differenza osservata nei due gruppi è abbastanza piccola da essersi verificata per caso, ipotizzando che non esista alcuna differenza nella popolazione

↓

IPOTESI NULLA

50.9 cm vs 52.3 cm differenza −1.4 cm

Calcoliamo la PROBABILITA' di ottenere una differenza come quella osservata se l'H_0 fosse vera

P {(media salto non uso-media salto creatina)|H_0}

↓

test statistico

Test statistico → **Test t di Student**

Variabile quantitativa

↓

t=-2.1984 p= 0.0348

```
. ttesti 18 50.9 1.7 18 52.3 2.1

Two-sample t test with equal variances
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      18        50.9    .4006938         1.7    50.05461    51.74539
       y |      18        52.3    .4949747         2.1    51.25569    53.34431
---------+--------------------------------------------------------------------
combined |      36        51.6    .3353984     2.01239    50.91911    52.28089
---------+--------------------------------------------------------------------
    diff |                -1.4    .6368324               -2.694199   -.1058008
------------------------------------------------------------------------------
Degrees of freedom: 34

                      Ho: mean(x) - mean(y) = diff = 0

     Ha: diff < 0               Ha: diff != 0              Ha: diff > 0
       t =  -2.1984               t =  -2.1984               t =  -2.1984
   P < t =   0.0174         P > |t| =   0.0348         P > t =   0.9826
```


CRITERIO DECISIONALE

se P è piccola (≤0.05) la differenza non è dovuta al caso

RIFIUTIAMO H_0

la differenza è STATISTICAMENTE SIGNIFICATIVA

se P è grande (>0.05)

la differenza è probabilmente dovuta al caso

non abbiamo sufficienti prove contro H_0

la differenza NON è STATISTICAMENTE SIGNIFICATIVA

CONCLUSIONI

P = 0.0348 la differenza non è dovuta al caso

RIFIUTIAMO H_0

la differenza nell'altezza del salto tra i due gruppi è STATISTICAMENTE SIGNIFICATIVA

↓

la creatina aumenta l'altezza del salto rispetto al non uso

SIGNIFICATIVO

non facilmente spiegato dal caso, cioè da errore di campionamento

P rappresenta la probabilità che il risultato sia dovuto al caso

MA

non dice niente sull'ENTITA' della DIFFERENZA

Con una variabile di tipo quantitativo, qual è il test statistico da effettuare?

Confronto fra soggetti diversi		Misure ripetute sugli stessi soggetti		Confronto fra variabili diverse
↓	↓	↓	↓	↓
2 gruppi	Più di 2 gruppi	2 misurazioni	Più di 2 misurazioni	
↓	↓	↓	↓	
t di Student	ANOVA a 1 criterio	t di Student per dati appaiati	ANOVA per misure ripetute	Regressione e Correlazione

ANOVA = ANalysis Of VAriance (Analisi della varianza)

SIGNIFICATIVITA' STATISTICA e RILEVANZA CLINICA

Un'indagine epidemiologica, condotta su un gran numero di persone, ha messo in luce che i fumatori dormono meno della popolazione generale.

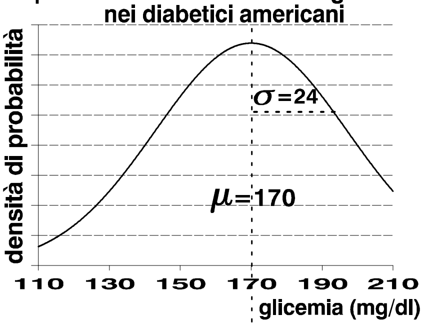
La differenza aveva una significatività elevata ($P<0.001$), ovvero ben difficilmente poteva essere attribuita al caso.

La differenza consisteva in 3 minuti di sonno in meno nei fumatori rispetto ai non-fumatori.

L’intervallo di confidenza come test d’ipotesi

L’intervallo di confidenza è una stima intervallare, ma può anche essere considerato un vero e proprio test d’ipotesi.

Ipotetica distribuzione della glicemia nei diabetici americani



IPOTESI NULLA: Nei diabetici italiani la distribuzione della glicemia è uguale a quella dei diabetici americani. Pertanto la media per campioni di numerosità 36 si distribuisce con:

$\mu = 170$

$\sigma/\sqrt{n} = 24/\sqrt{36} = 4$

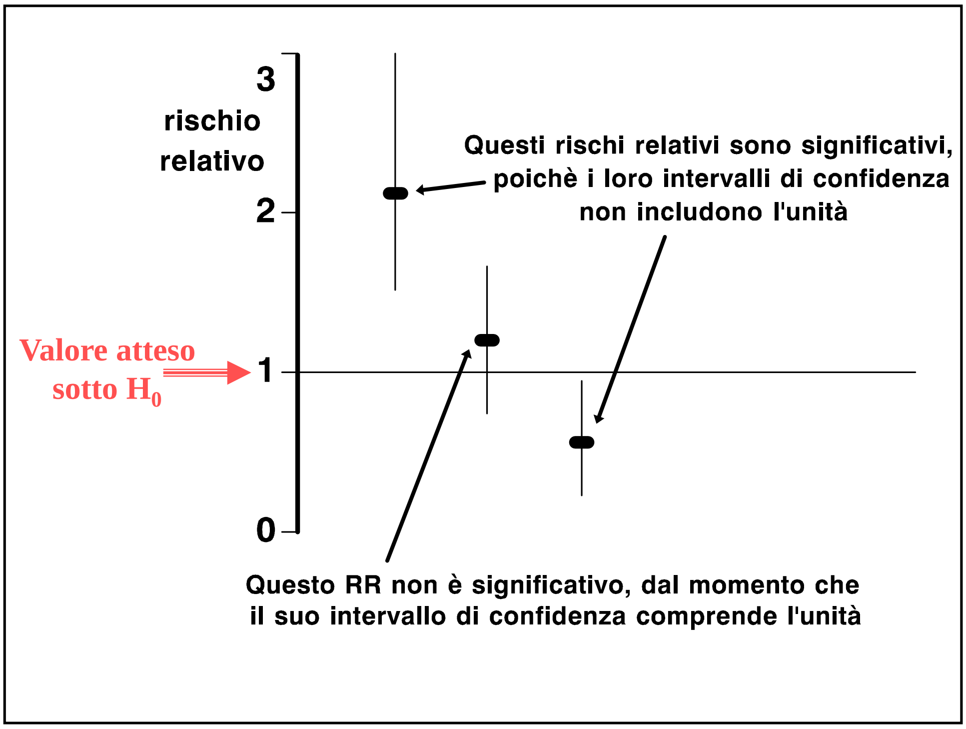
stime intervallari di μ

$155 \pm 1{,}96*4$ 147,2 —— 162,8

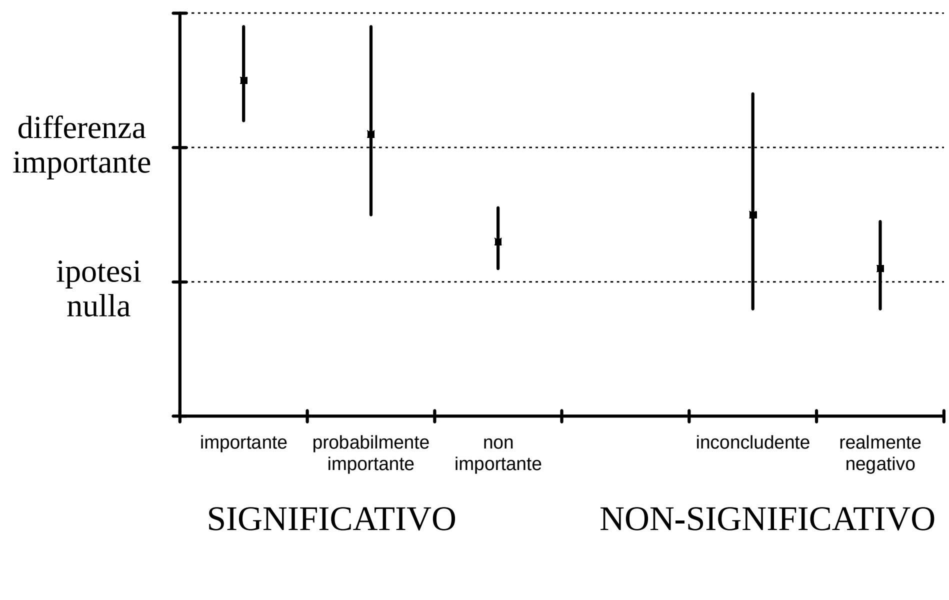
$161 \pm 1{,}96*4$ 153,2 —— 168,8

$166 \pm 1{,}96*4$ 158,2 —— 173,8

I primi due intervalli di confidenza sono significativi perché non contengono 170 mg/dl (valore atteso sotto H0)



Valutazione clinico-statistica di un intervallo di confidenza



“Overemphasis on hypothesis testing - and the use of P values to dichotomise significant or non-significant results - has detracted from more useful approaches to interpreting study results, such as estimation and confidence intervals.

In medical studies investigators are usually interested in determining the size of difference of a measured outcome between groups, rather than a simple indication of whether or not it is statistically significant ...

Confidence intervals, if appropriate to the type of study, should be used for major findings in both the main text of a paper and its abstract.”

Gardner MJ, Altman DG (1986) Confidence intervals rather than P values: estimation rather than hypothesis testing. British Medical Journal, 292: 746-750