

# Test statistico



SAPIENZA  
UNIVERSITÀ DI ROMA

[annarita.vestri@uniroma1.it](mailto:annarita.vestri@uniroma1.it)

# Hypothesis

The ideas in hypothesis testing are based on deductive reasoning. We assume that some probability model is true and then ask “What are the chances that these observations came from that probability model?”.

*IN DEDUCTIVE REASONING, WE REASON FROM A HYPOTHESIS TO A CONCLUSION: “IF LORD FASTBACK COMMITTED MURDER, THEN HE WOULD WIPE THE FINGER-PRINTS OFF THE GUN.”*

*INDUCTIVE REASONING, BY CONTRAST, ARGUES BACKWARD FROM A SET OF OBSERVATIONS TO A REASONABLE HYPOTHESIS:*



# 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

# Risk for Congenital Malformation With H1N1 Influenza Vaccine

## A Cohort Study With Sibling Analysis

Jonas F. Ludvigsson, MD, PhD; Peter Ström, MSc; Cecilia Lundholm, MSc; Sven Cnattingius, MD, PhD; Anders Ekblom, MD, PhD; Åke Örtqvist, MD, PhD; Nils Feltelius, MD, PhD; Fredrik Granath, PhD; and Olof Stephansson, MD, PhD

**Background:** Earlier studies reporting varying risk estimates for congenital malformation in offspring of mothers undergoing vaccination against H1N1 influenza during pregnancy did not consider the potential role of confounding by familial (genetic and shared environmental) factors.

**Objective:** To evaluate an association between maternal H1N1 vaccination during pregnancy and offspring malformation, with familial factors taken into account.

**Design:** Population-based prospective study.

**Setting:** Sweden.

**Participants:** Liveborn offspring born between 1 October 2009 and 1 October 2011 to mothers receiving monovalent AS03-adjuvanted H1N1 influenza vaccine (Pandemrix [GlaxoSmith-Kline]) during pregnancy. A total of 40 983 offspring were prenatally exposed to the vaccine, 14 385 were exposed within the first trimester (14 weeks), and 7502 were exposed during the first 8 weeks of pregnancy. Exposed offspring were compared with 197 588 unexposed offspring. Corresponding risks in exposed versus unexposed siblings were also estimated.

**Measurements:** Congenital malformation, with subanalyses for congenital heart disease, oral cleft, and limb deficiency.

**Results:** Congenital malformation was observed in 2037 (4.97%) exposed offspring and 9443 (4.78%) unexposed off-

spring. Adjusted risk for congenital malformation was 4.98% in exposed offspring versus 4.96% in unexposed offspring (risk difference, 0.02% [95% CI, -0.26% to 0.30%]). The corresponding risk differences were 0.16% (CI, -0.23% to 0.56%) for vaccination during the first trimester and 0.10% (CI, -0.41% to 0.62%) for vaccination in the first 8 weeks. Using siblings as comparators yielded no statistically significant risk differences.

**Limitations:** The study was based on live births, and the possibility that data on miscarriage or induced abortion could have influenced the findings cannot be ruled out. Study power was limited in analyses of specific malformations.

**Conclusion:** When intrafamilial factors were taken into consideration, H1N1 vaccination during pregnancy did not seem to be linked to overall congenital malformation in offspring, although risk increases for specific malformations could not be ruled out completely.

**Primary Funding Source:** Swedish Research Council and Swedish Council for Working Life and Social Research.

## BACKGROUND

The addition of azithromycin to standard regimens for antibiotic prophylaxis before cesarean delivery may further reduce the rate of postoperative infection. We evaluated the benefits and safety of azithromycin-based extended-spectrum prophylaxis in women undergoing nonelective cesarean section.

## METHODS

In this trial conducted at 14 centers in the United States, we studied 2013 women who had a singleton pregnancy with a gestation of 24 weeks or more and who were undergoing cesarean delivery during labor or after membrane rupture. We randomly assigned 1019 to receive 500 mg of intravenous azithromycin and 994 to receive placebo. All the women were also scheduled to receive standard antibiotic prophylaxis. The primary outcome was a composite of endometritis, wound infection, or other infection occurring within 6 weeks.

## RESULTS

The primary outcome occurred in 62 women (6.1%) who received azithromycin and in 119 (12.0%) who received placebo (relative risk, 0.51; 95% confidence interval [CI], 0.38 to 0.68;  $P < 0.001$ ). There were significant differences between the azithromycin group and the placebo group in rates of endometritis (3.8% vs. 6.1%,  $P = 0.02$ ), wound infection (2.4% vs. 6.6%,  $P < 0.001$ ), and serious maternal adverse events (1.5% vs. 2.9%,  $P = 0.03$ ). There was no significant between-group difference in a secondary neonatal composite outcome that included neonatal death and serious neonatal complications (14.3% vs. 13.6%,  $P = 0.63$ ).

## CONCLUSIONS

Among women undergoing nonelective cesarean delivery who were all receiving standard antibiotic prophylaxis, extended-spectrum prophylaxis with adjunctive azithromycin was more effective than placebo in reducing the risk of postoperative infection. (Funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development; C/SOAP ClinicalTrials.gov number, NCT01235546.)

## Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China

**IMPORTANCE** Coronavirus disease 2019 (COVID-19) is an emerging infectious disease that was first reported in Wuhan, China, and has subsequently spread worldwide. Risk factors for the clinical outcomes of COVID-19 pneumonia have not yet been well delineated.

**OBJECTIVE** To describe the clinical characteristics and outcomes in patients with COVID-19 pneumonia who developed acute respiratory distress syndrome (ARDS) or died.

**DESIGN, SETTING, AND PARTICIPANTS** Retrospective cohort study of 201 patients with confirmed COVID-19 pneumonia admitted to Wuhan Jinyintan Hospital in China between December 25, 2019, and January 26, 2020. The final date of follow-up was February 13, 2020.

**EXPOSURES** Confirmed COVID-19 pneumonia.

**MAIN OUTCOMES AND MEASURES** The development of ARDS and death. Epidemiological, demographic, clinical, laboratory, management, treatment, and outcome data were also collected and analyzed.

**RESULTS** Of 201 patients, the median age was 51 years (interquartile range, 43-60 years), and 128 (63.7%) patients were men. Eighty-four patients (41.8%) developed ARDS, and of those 84 patients, 44 (52.4%) died. In those who developed ARDS, compared with those who did not, more patients presented with dyspnea (50 of 84 [59.5%] patients and 30 of 117 [25.6%] patients, respectively [difference, 33.9%; 95% CI, 19.7%-48.1%]) and had comorbidities such as hypertension (23 of 84 [27.4%] patients and 16 of 117 [13.7%] patients, respectively [difference, 13.7%; 95% CI, 1.3%-26.1%]) and diabetes (16 of 84 [19.0%] patients and 6 of 117 [5.1%] patients, respectively [difference, 13.9%; 95% CI, 3.6%-24.2%]). In bivariate Cox regression analysis, risk factors associated with the development of ARDS and progression from ARDS to death included older age (hazard ratio [HR], 3.26; 95% CI 2.08-5.11; and HR, 6.17; 95% CI, 3.26-11.67, respectively), neutrophilia (HR, 1.14; 95% CI, 1.09-1.19; and HR, 1.08; 95% CI, 1.01-1.17, respectively), and organ and coagulation dysfunction (eg, higher lactate dehydrogenase [HR, 1.61; 95% CI, 1.44-1.79; and HR, 1.30; 95% CI, 1.11-1.52, respectively] and D-dimer [HR, 1.03; 95% CI, 1.01-1.04; and HR, 1.02; 95% CI, 1.01-1.04, respectively]). High fever ( $\geq 39$  °C) was associated with higher likelihood of ARDS development (HR, 1.77; 95% CI, 1.11-2.84) and lower likelihood of death (HR, 0.41; 95% CI, 0.21-0.82). Among patients with ARDS, treatment with methylprednisolone decreased the risk of death (HR, 0.38; 95% CI, 0.20-0.72).



**Passive exposure to agricultural pesticides and risk of childhood leukemia in an Italian community.**

[Malagoli C](#)<sup>1</sup>, [Costanzini S](#)<sup>2</sup>, [Heck JE](#)<sup>3</sup>, [Malavolti M](#)<sup>1</sup>, [De Girolamo G](#)<sup>4</sup>, [Oleari P](#)<sup>5</sup>, [Palazzi G](#)<sup>6</sup>, [Teggi S](#)<sup>2</sup>, [Vinceti](#)

**Abstract**

**BACKGROUND:**

Exposure to pesticides has been suggested as a risk factor for childhood leukemia, but definitive evidence on this relation and the specific pesticides involved is still not clear.

**OBJECTIVE:**

We carried out a population-based case-control study in a Northern Italy community to assess the possible relation between passive exposure to agricultural pesticides and risk of acute childhood leukemia.

**METHODS:**

We assessed passive pesticide exposure of 111 childhood leukemia cases and 444 matched controls by determining density and type of agricultural land use within a 100-m radius buffer around children's homes. We focused on four common crop types, arable, orchard, vineyard and vegetable, characterized by the use of specific pesticides that are potentially involved in childhood induced leukemia. The use of these pesticides was validated within the present study. We computed the odds ratios (OR) of the disease and their 95% confidence intervals (CI) according to type and density of crops around the children's homes, also taking into account traffic pollution and high-voltage power line magnetic field exposure.

**RESULTS:**

Childhood leukemia risk did not increase in relation with any of the crop types with the exception of arable crops, characterized by the use of 2,4-D, MCPA, glyphosate, dicamba, triazine and cypermethrin. The very few children (n=11) residing close to arable crops had an **OR for childhood leukemia of 2.04 (95% CI 0.50-8.35), and such excess risk was further enhanced among children aged <5 years.**

**CONCLUSIONS:**

Despite the null association with most crop types and the statistical imprecision of the estimates, the increased leukemia risk among children residing close to arable crops indicates the need to further investigate the involvement in disease etiology of passive exposure to herbicides and pyrethroids, though such exposure is unlikely to play a role in the vast majority of cases.

# 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

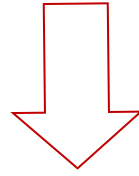
$H_0$ : La glicemia dei diabetici  
italiani è uguale alla glicemia  
dei diabetici americani



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

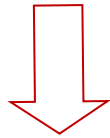
$H_1$ : La glicemia dei diabetici italiani  
è diversa dalla glicemia dei  
diabetici americani

# Dati del campione



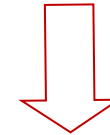
## Test statistico

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

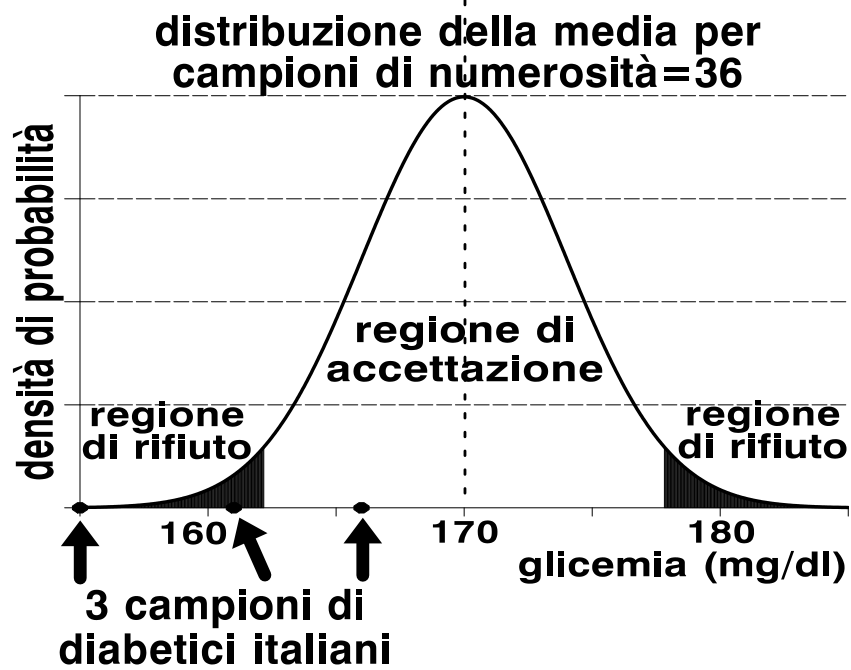
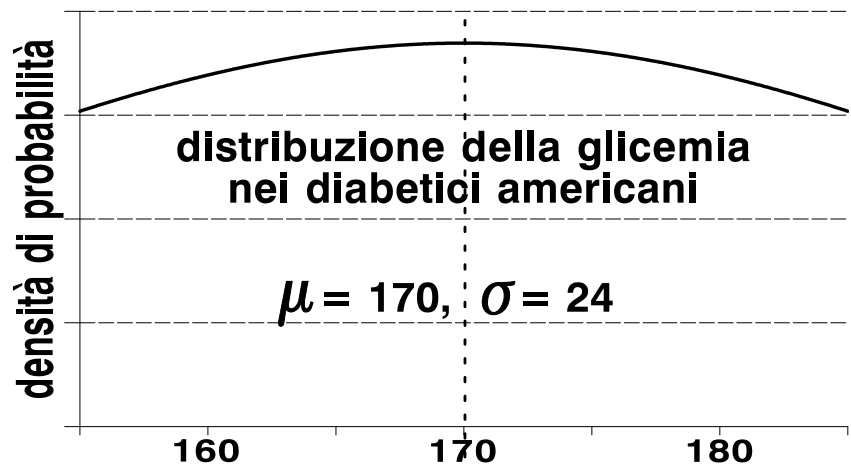


**Accetto l'ipotesi nulla ( $H_0$ )**  
tutte le differenze osservate tra i campioni possono essere attribuite al caso

$P < 0,05$  = 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



Un campione estratto fornisce una media della glicemia pari a 162

# 1. Formulazione dell'ipotesi da verificare

A. ( $H_0$  ipotesi nulla )

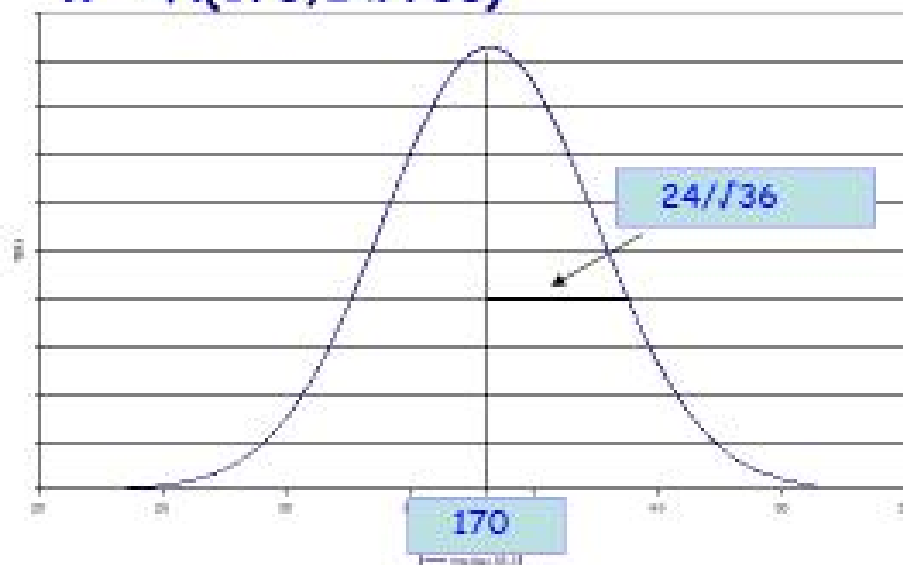
Spiega le differenze osservate come dovute al caso

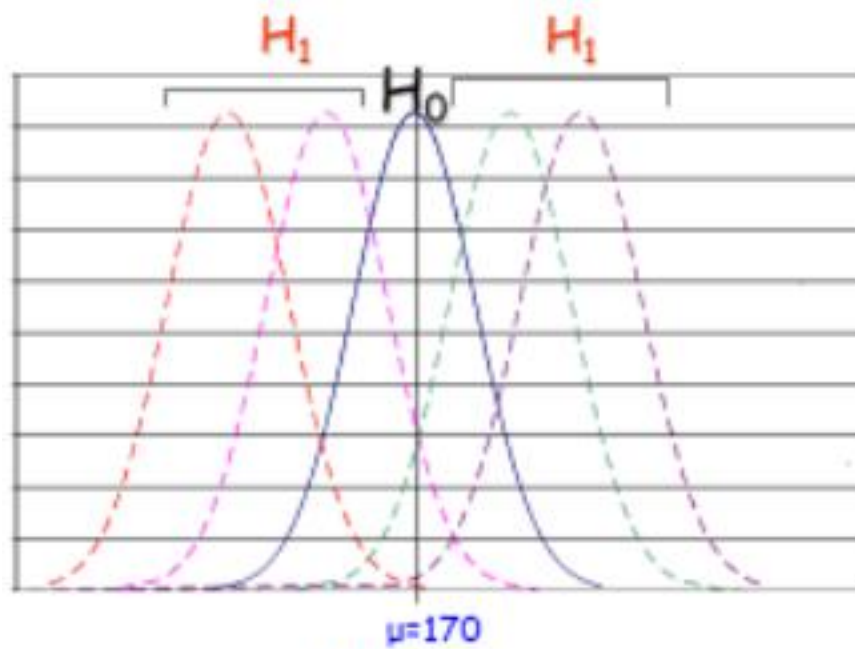
$$H_0: \mu = \mu_0 = 170 \text{ mg/dl}$$

La media della popolazione italiana da cui proviene il campione ( $\mu$ ) è identica alla media della popolazione americana ( $\mu_0$ )

L'ipotesi è anche una congettura sulla distribuzione campionaria, infatti

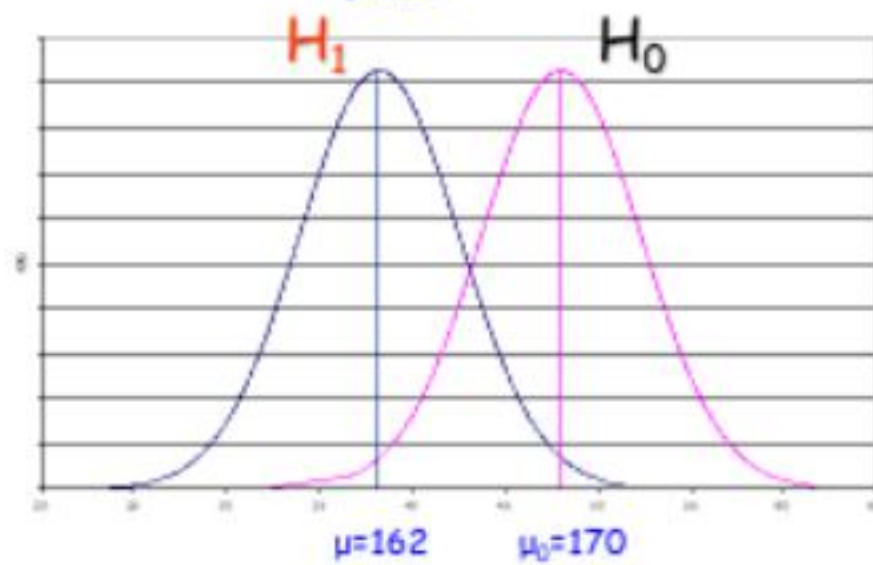
$$\bar{x} \sim N(170, 24/\sqrt{36})$$





Ipotesi alternativa

$$H_1: \mu \neq \mu_0$$

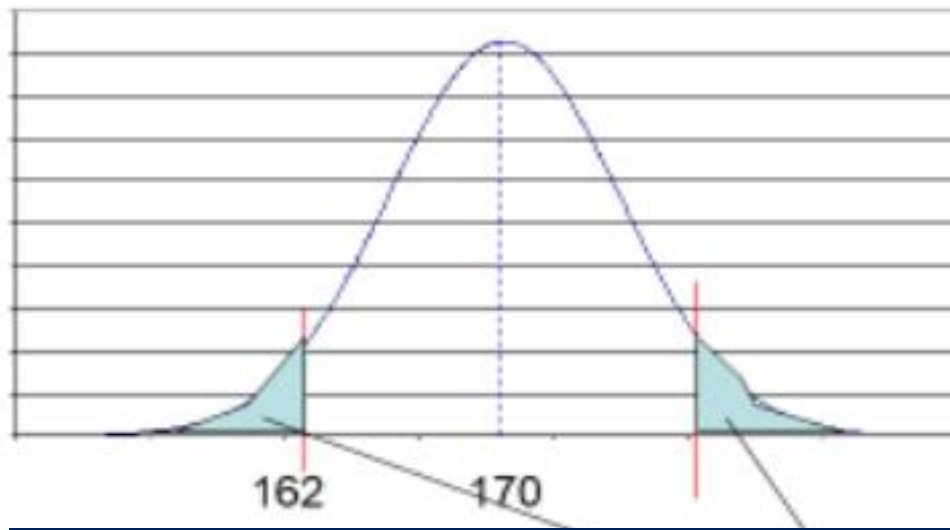


$$\mu=162$$

TEST  
STATISTICO

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

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



-2

2

$P=0,046$



$P < 0,05$

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



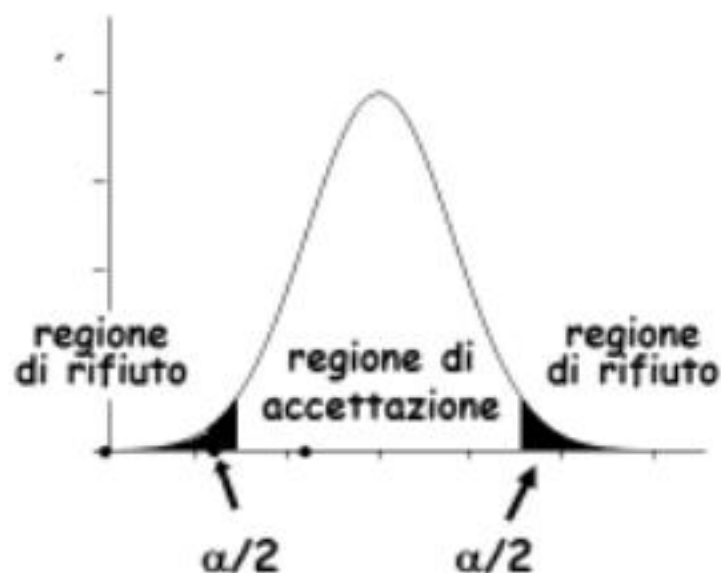


# ALTRE FORMULAZIONI DELL'IPOTESI NULLA

## Test bidirezionale

$$H_0: \mu = \mu_0$$

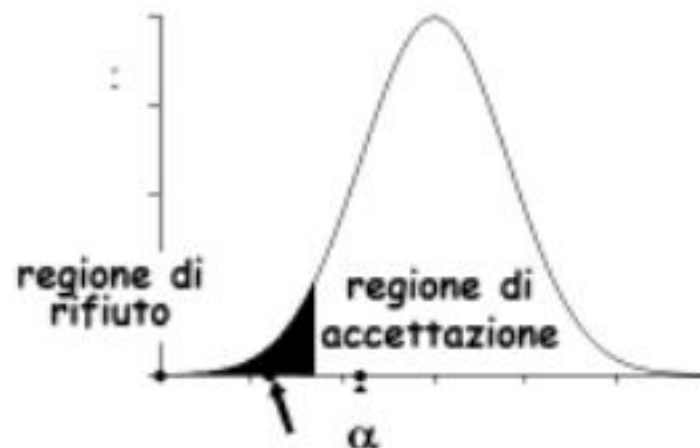
$$H_1: \mu \neq \mu_0$$



## Test unidirezionale

$$H_0: \mu = \mu_0$$

$$H_1: \mu < \mu_0$$



In qualsiasi studio clinico è prevista la raccolta di dati, che vengono utilizzati per descrivere la popolazione in studio e per rispondere a uno o più quesiti di ricerca.

Di solito, l'obiettivo di uno studio è valutare se esista un'associazione fra alcune caratteristiche o se intervenendo, ad esempio con un trattamento, sia possibile modificare specifici parametri o modificare la storia naturale di una patologia.

Ad esempio, si potrebbero raccogliere dati per indagare se esiste una associazione fra livello di obesità e livello di controllo metabolico in una popolazione di soggetti affetti da diabete tipo 2, o per valutare se un intervento educativo rivolto a promuovere stili di vita più salutari sia in grado di ridurre i livelli di emoglobina glicosilata in soggetti con diabete tipo 1.

Per rispondere a questi quesiti, utilizziamo i test statistici, che ci permetteranno di accettare o di rifiutare (confutare) una ipotesi.

**Tutti i test statistici partono dall'ipotesi nulla che non esista una relazione fra le variabili in studio.**

Se ad esempio volessimo saggiare l'ipotesi che, fra i soggetti con diabete, il livello di controllo metabolico dipende dal livello di obesità espresso dall'indice di massa corporea (BMI), l'ipotesi nulla di partenza sarebbe rappresentata da una completa assenza di associazione fra controllo metabolico e BMI.

Scopo del test statistico sarà quello di suggerirci se accettare questa ipotesi nulla, o se invece rifiutarla, ritenendo quindi che ci possa essere un rapporto significativo fra controllo metabolico e livello di obesità.

Tuttavia in qualsiasi studio, per grande che esso sia, il numero di pazienti arruolati costituirà sempre una piccola frazione di tutti gli individui affetti da quella specifica patologia.

Sebbene si assuma che il campione studiato sia rappresentativo della popolazione da cui esso deriva, a causa della variabilità biologica, ciò potrebbe non essere vero e le conclusioni a cui perveniamo analizzando il nostro campione potrebbero quindi essere erranee.

Per tale ragione, i risultati di un test statistico sono sempre espressi in termini probabilistici.

In altre parole, nell'accettare o rifiutare l'ipotesi nulla possiamo commettere un errore; se tuttavia la probabilità di commettere tale errore è molto bassa, allora potremo accettare con sufficiente fiducia le conclusioni a cui siamo arrivati.

Ma in base a quale regola decidiamo di accettare o di rifiutare l'ipotesi nulla?

La decisione viene presa in base al valore di  $p$  associato al nostro test statistico.

Il valore di  $p$  sta a indicare la probabilità di commettere un errore rifiutando l'ipotesi nulla, e cioè la probabilità di sbagliare, affermando ad esempio che c'è una relazione significativa fra controllo metabolico e obesità.

**Per convenzione, rifiutiamo l'ipotesi nulla, tutte le volte che il valore di  $p$  è minore o uguale a 0,05.**

Un valore di  $p < 0,05$  sta a indicare che abbiamo una probabilità di sbagliare minore del 5%, rifiutando l'ipotesi nulla.

**Devziata gaussiana Standard: Aree per z>+z\* (o per z<-z\*)**

z*	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831

**Devziata gaussiana Standard: Aree Per  $z > +z^*$  (o per  $z < -z^*$ )**

$z^*$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
2-4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
3.0	.00135	.00097	.00069	.00048	.00034	.00023	.00016	.00011	.00007	.00005
4.0	.00003	.00002	.00001	.00001	.00001	.00000	.00000	.00000	.00000	.00000



## esempio

Supponiamo che i pazienti con BMI compreso fra 25 e 27 (n=53) abbiano valori medi di HbA1c di  $7,18 \pm 1,6$  (media  $\pm$  deviazione standard), mentre fra i soggetti francamente obesi, con BMI  $> 30$  (n=55) il valore di HbA1c sia di  $7,91 \pm 1,9$ .

La domanda che ci poniamo è la seguente: i livelli di HbA1c sono significativamente più elevati nelle persone obese?

In altre parole, esiste una relazione fra livelli di HbA1c e BMI?

L'ipotesi nulla di partenza è che non ci sia relazione fra le due variabili, e che quindi i livelli medi di HbA1c non differiscano in modo significativo nei due gruppi di pazienti, classificati in base al loro BMI.

L'ipotesi alternativa è invece rappresentata dall'esistenza di una relazione fra le due variabili, e quindi di un controllo metabolico significativamente peggiore nelle persone obese ( $BMI > 30$ ) rispetto alle altre.

Applicando il test statistico appropriato, che in questo caso è rappresentato dal test t di student ( $t=-2.16$ ) per dati non appaiati, otterremo un valore di p pari a 0,033.

Cosa ci suggerisce questo valore di p?

Un valore di p di 0,033 sta a indicare che, rifiutando l'ipotesi nulla e affermando quindi che c'è una relazione significativa fra BMI e livelli di HbA1c, abbiamo una probabilità di sbagliare di circa il 3%.

In altre parole, la probabilità che sia vera l'ipotesi nulla, e che non ci sia quindi relazione fra le due variabili, è inferiore al 5%.

Alla luce di questi risultati, possiamo ragionevolmente affermare che i livelli medi di HbA1c sono significativamente più alti nelle persone con BMI > 30, rispetto a quelle con BMI compreso fra 25 e 27.

# Effettuazione di un TEST D'IPOTESI

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

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

dove

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

$\theta_0$  = valore del parametro di interesse sotto l'ipotesi nulla  $H_0$

$\text{ES}[\hat{\theta}]$  = 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,  $>0,05$ )

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

Confronto fra soggetti diversi



2 gruppi



t di Student



Più di 2 gruppi



ANOVA a 1  
criterio

Misure ripetute sugli stessi soggetti



2 misurazioni



t di Student per  
dati appaiati



Più di 2  
misurazioni



ANOVA per  
misure ripetute

Confronto fra  
variabili diverse



Regressione e  
Correlazione

ANOVA = ANalysis Of VARIance (Analisi della varianza)

# TEST D'IPOTESI ED ERRORI DELLA LOGICA DECISIONALE

	Realtà	
Verdetto	<u>Colpevole</u>	<u>Innocente</u>
<u>Colpevole</u>	Corretto	Non corretto
<u>Innocente</u>	Non corretto	Corretto

In statistica questi errori accadono in conseguenza della variabilità casuale.

Lo statistico indica il rischio che considera accettabile di fornire una conclusione diversa dalla realtà (ignota).

	Ipotesi nulla	
Statistico	<u>falsa</u>	<u>vera</u>
<u>Rifiuta</u>	Corretto	Non corretto
<u>Non rifiuta</u>	Non corretto	Corretto

	Ipotesi nulla	
Statistico	<u>falsa</u>	<u>vera</u>
<u>Rifiuta</u>	Corretto	Errore $\alpha$
<u>Non rifiuta</u>	Errore $\beta$	Corretto



Nel procedimento di verifica delle ipotesi applichiamo metodi probabilistici e quindi possiamo trovarci a rifiutare l'ipotesi nulla quando essa è vera.

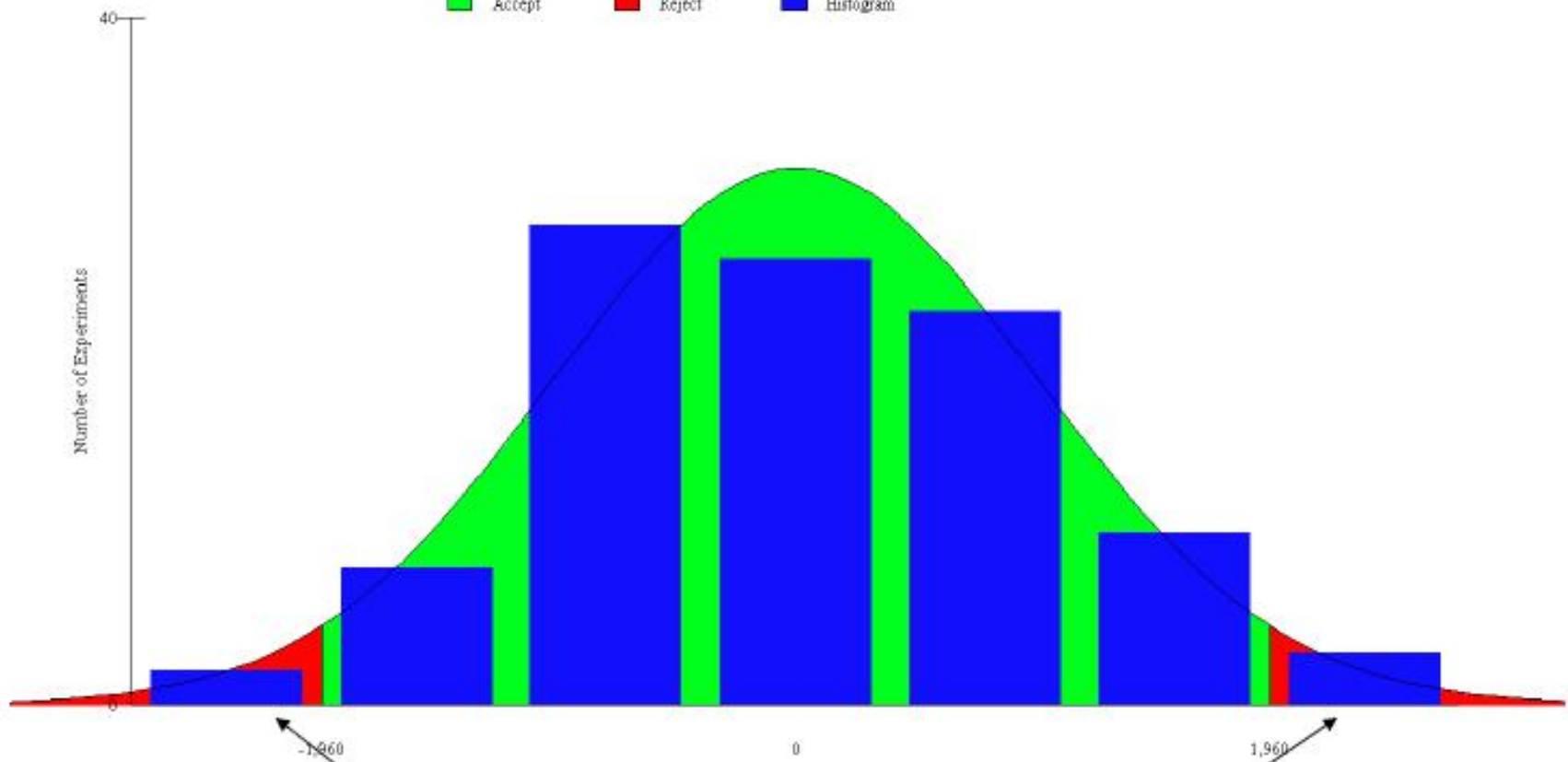
Si parla in questo caso di Errore di 1° tipo o Errore  $\alpha$ .

Di solito l'errore di primo tipo è fissato al 5% ( $\alpha = 0.05$ ). Questo significa che ripetendo molte volte un esperimento, nella situazione in cui  $H_0$  è vera, ci troveremo a rifiutarla nel 5% degli esperimenti.

Come può accadere?

Exp #58: HistogramDistribution of Z Statistic

Accept Reject Histogram



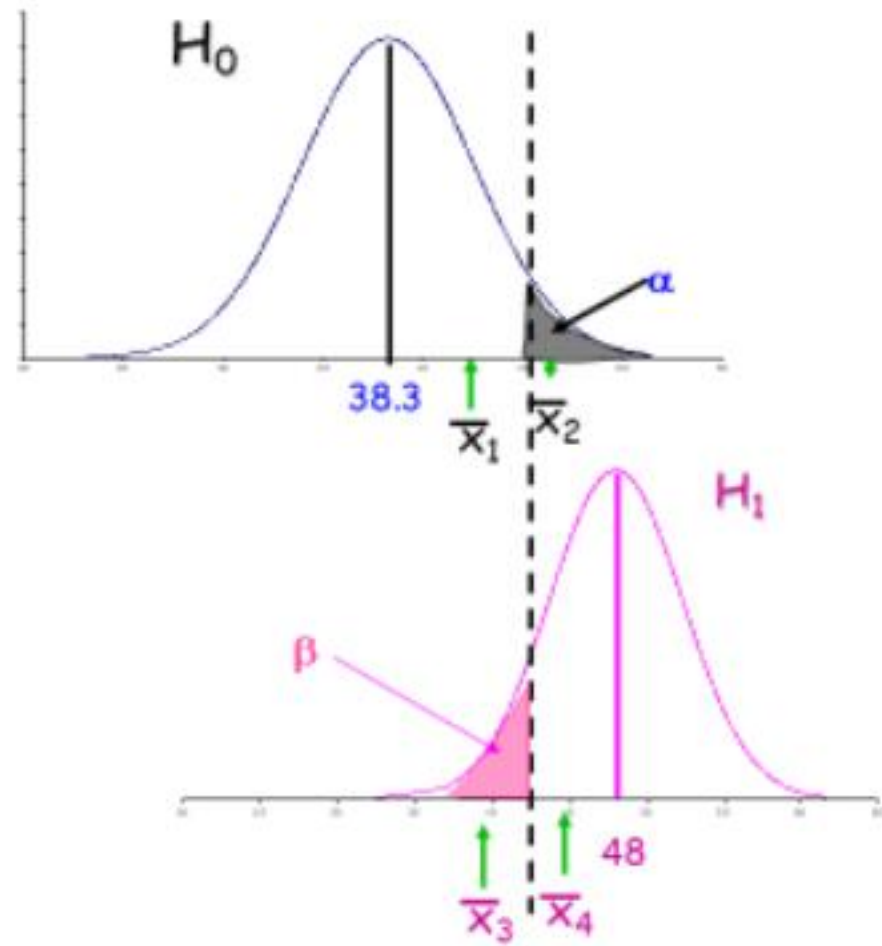
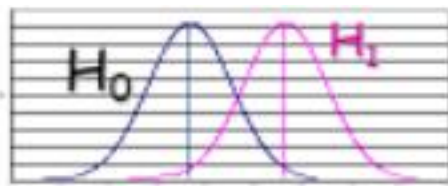
Basta che il nostro campione sia uno di questi!

Quando una decisione è presa sulla base del risultato di un test di ipotesi, si possono commettere due tipi di errore:

- **ERRORE DI I° TIPO ( $\alpha$ )**: rifiutare l'ipotesi nulla  $H_0$  quando essa è vera (falso positivo)
- **ERRORE DI II° TIPO ( $\beta$ )**: non rifiutare l'ipotesi nulla  $H_0$  quando essa è falsa (falso negativo)

Complementare all'errore  $\beta$  abbiamo la potenza dello studio:

$$\text{Potenza} = 1 - (\text{Errore } \beta)$$



# Potenza di un test



La potenza di un test è la probabilità di rifiutare  $H_0$  quando essa è falsa (1-b)



**Esempio: POTENZA DEL TEST=(1-beta )=0.73**

**Vuol dire che accettiamo di non rilevare delle differenze vere nel 27% dei casi (1-0,73)**

**POTENZA di un test =  $(1 - \beta) = 1 - \text{Pr}(\text{errore del II tipo})$**

**E' la probabilità che un test statistico ha di rifiutare l'ipotesi nulla quando l'ipotesi nulla è effettivamente falsa.**

**In altre parole, la Potenza di un test è la sua capacità di cogliere delle differenze, quando queste differenze esistono.**

**Il test statistico è costruito in modo da mantenere costante il livello di significatività, indipendentemente dalla numerosità campionaria. Ma questo risultato viene raggiunto a spese della potenza del test, che aumenta all'aumentare della numerosità campionaria.**

**La POTENZA di un test dipende:**

- 1) dalla numerosità del campione**
- 2) dalla variabilità del fenomeno in studio**
- 3) dalla differenza minima che si vuole mettere in evidenza**
- 4) dal livello di significatività adottato.**

**Il modo principale per raggiungere un'adeguata potenza è pianificare un'adeguata numerosità campionaria nel protocollo dello studio.**

# **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.**