Calcolo della numerosità campionaria

Prof. Giuseppe Verlato
Sezione di Epidemiologia e Statistica Medica,
Università di Verona

Concetto di potenza statistica

Ipotesi Nulla (H_0)

	vera	falsa
Accetto H_0	Va bene	Errore del II tipo
Rifiuto H_0	Errore del I tipo	Va bene

P(errore del I tipo) = α (alfa)

P(errore del II tipo) = β (beta)

In genere, nel test d'ipotesi la probabilità di errore del I tipo viene fissata al 5% (0,05). Pertanto in un caso su 20 si rifiuterà H_0 (ovvero il test risulterà significativo) per semplice effetto del caso, anche quando H_0 è vera. In termini statistici si sceglie un livello di significatività del 5%.

Ad esempio, se in un test d'ipotesi P<0,01, vuol dire che posso rifiutare H_0 con una probabilità di errore del I tipo inferiore all'1%; in altre parole la probabilità che le differenze osservate siano dovute al caso è inferiore all'1%.

Null hypothesis (H_0)

	true	false
H_0 accepted	O.K.	Type II error
H_0 rejected	Type I error	O.K.

P(type I error) = α (alpha)

P(type II error) = β (beta)

In hypothesis testing probability of type I error is usually set at 5% (0.05). Hence H_0 will be rejected (and the statistical test will turn out to be significant) at RANDOM in one in twenty tests, even if H_0 is true. In statistical terms significance level is set at 5%.

For instance, if in hypothesis testing *P*<0.01, H_0 can be rejected with a probability of type I error lower than 1%; in other words, the probability that observed differences are due to chance is lower than 1%.

POTENZA di un test = 1- beta = 1 - P(errore del II tipo)

E' la probabilità che un test statistico ha di falsificare 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.

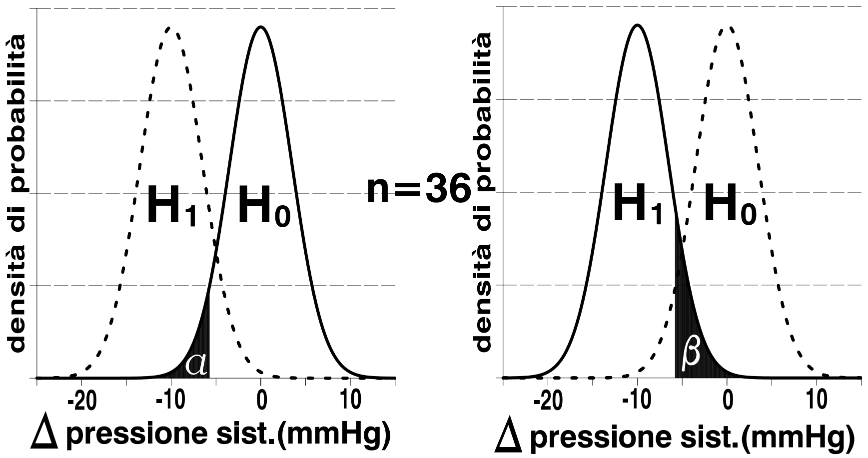
POWER of a statistical test=1-beta=1-P(type II error)

It is the probability to falsify the null hypothesis when null hypothesis is indeed false.

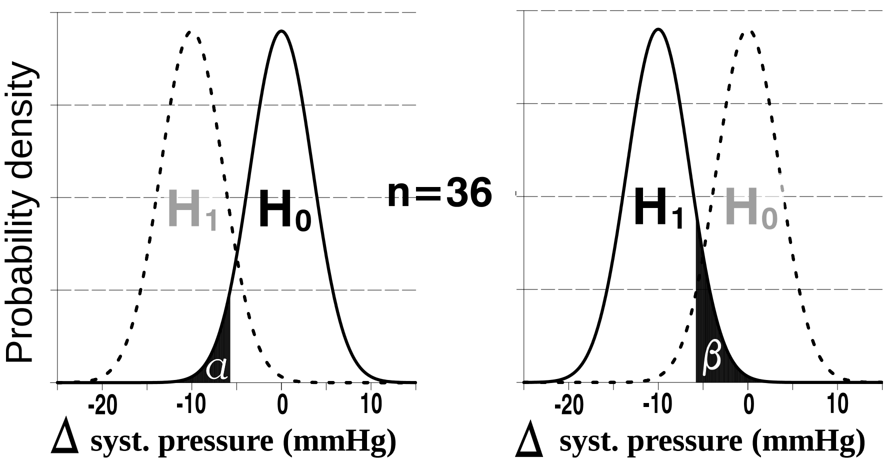
In other words, the power of a statistical test reflects its ability to detect differences, when these differences do exist.

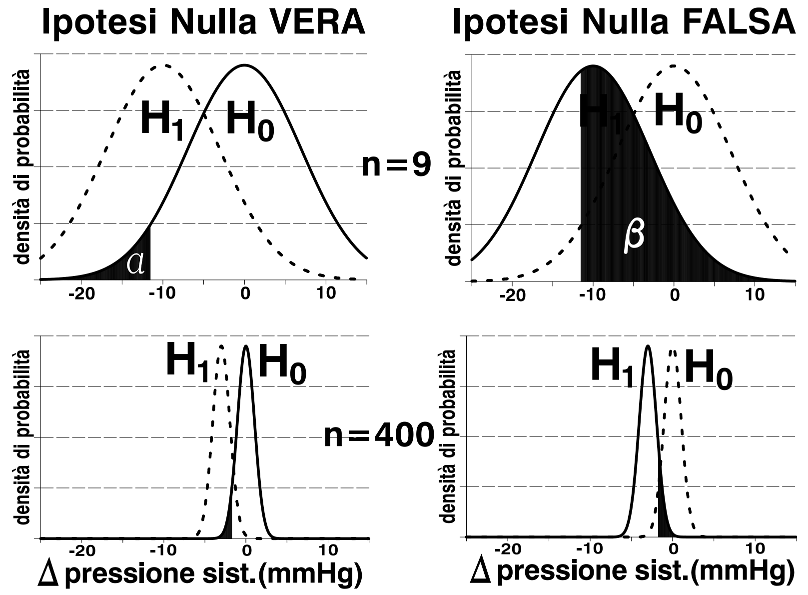
The statistical test has been developed in order to keep constant the level of significance, irrespective of sample size. But this result has been achieved at the expenses of statistical power, which is largely affected by sample size.

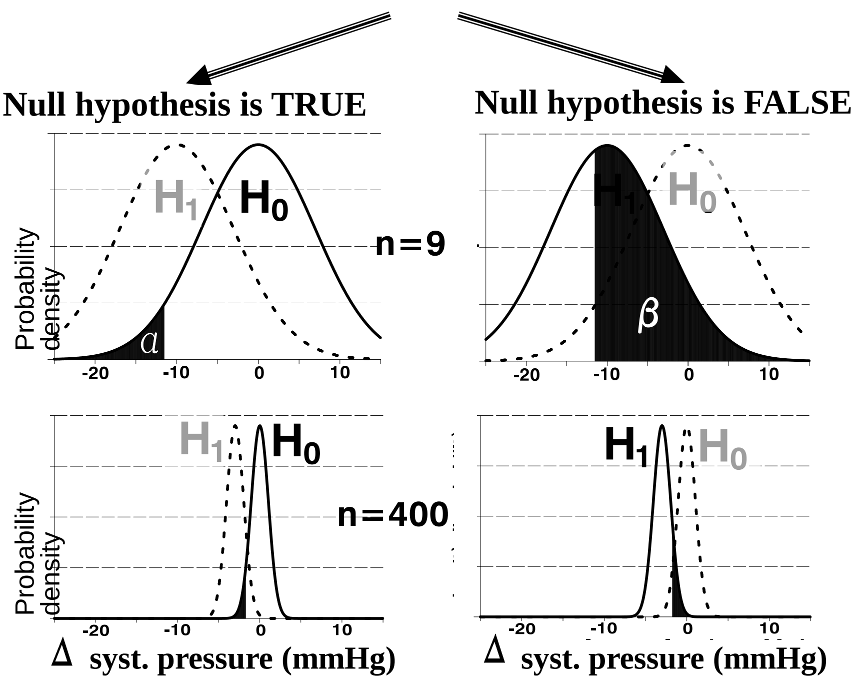
Ipotesi Nulla VERA | Ipotesi Nulla FALSA



Null hypothesis is TRUE | Null hypothesis is FALSE




Ipotesi Nulla VERA
Ipotesi Nulla FALSA
densità di probabilità
H1
H0
n=9
n=400
α
β
-20
-10
0
10
Δ pressione sist.(mmHg)


Null hypothesis is TRUE
Null hypothesis is FALSE
Probability density
H1
H0
n=9
n=400
α
β
-20
-10
0
10
Δ syst. pressure (mmHg)

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.

STATISTICAL SIGNIFICANCE *versus*
CLINICAL IMPORTANCE

An epidemiological survey, performed on a large number of people, highlighted that smokers sleep less than non-smokers.

The difference was highly significant ($p<0.001$), i.e. it was unlikely to be due to chance.

The difference consisted in 3 minute decrease in sleep duration in smokers as compared to non-smokers.

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.

A test POWER depends on:

1) sample size

2) variability of the characteristic under study

3) minimal difference to be highlighted

4) significance level adopted.

The most important way to achieve an adequate power is to plan an adequate sample size in the study protocol.

Calcolo della numerosità campionaria necessaria per conseguire una potenza adeguata per …

1) confrontare due medie campionarie (test t)

2) valutare se una variazione media è significativa (test t per dati appaiati)

3) confrontare due proporzioni campionarie (test del chi-quadrato)

Calcolo della numerosità campionaria necessaria per conseguire una precisione adeguata per …

1) ottenere un intervallo di confidenza della media di ampiezza non superiore ad un dato valore

2) ottenere un intervallo di confidenza della prevalenza di ampiezza non superiore ad un dato valore

Variabile di risposta quantitativa: Confronto fra 2 medie

$$n > 2 \left[\frac{(z_{\alpha/2} + z_{\beta})\, \sigma}{\delta} \right]^2$$

dove n = numerosità di ciascuno dei due gruppi

$z_{\alpha/2}$ = 1.96 per alfa = 5%

z_{β} = 0.842, 1.282, 1.645 per potenza = 80, 90, 95%

σ = deviazione standard, desunta da studi pilota o dalla letteratura

$\delta = \mu_1 - \mu_2$ = differenza minima clinicamente rilevante

Esempio: Il farmaco di riferimento riduce la pressione sistolica di 25 mmHg, il nuovo farmaco per essere competitivo dovrebbe ridurre la pressione sistolica di almeno 30 mmHg (ovvero 5 mmHg in più). La deviazione standard della riduzione della pressione viene stimata in 10 mmHg da studi precedenti. Si adotta un alfa del 5% e una potenza del 90%, pertanto $z_{\alpha/2} = 1.96$ e $z_{\beta} = 1.282$.

$$n > 2 \left[\frac{(z_{\alpha/2} + z_{\beta})\, \sigma}{\delta} \right]^2$$

$$n > 2 \left[\frac{(1.96 + 1.28)\, 10}{5} \right]^2$$

$$n > 84.06 \qquad n \geq 85$$

Occorrono almeno 85 soggetti per gruppo.

Computing sample size to achieve enough power to detect possible difference between two means

$$n > 2 \left[\frac{(z_{\alpha} + z_{\beta})\, \sigma}{\delta} \right]^2$$

where

n = number of subjects in each group

$z_{\alpha} = 1.96$ for alpha = 5%

$z_{\beta} = 0.842, 1.282, 1.645$ for power = 80, 90, 95%

σ = standard deviation, derived from pilot studies or current literature

$\delta = \bar{x}_1 - \bar{x}_2$ = minimal clinically-important difference

Example: The reference drug decreases systolic pressure by 25 mmHg. To represent a real improvement, the new drug should reduce systolic pressure by at least 30 mmHg (i.e. 5 mmHg more). Standard deviation of pressure changes is estimated to be 10 mmHg from previous studies. Alpha is set at 5% and power at 90%, hence $z_{\alpha/2} = 1.96$ e $z_\beta = 1.282$.

$$\mathbf{n > 2\left[\frac{(z_\alpha + z_\beta)\,\sigma}{\delta}\right]^2}$$

$$\mathbf{n > 2\left[\frac{(1.96 + 1.28)\,10}{5}\right]^2}$$

$$\mathbf{n > 84.06 \qquad n \geq 85}$$

85 subjects per group are needed to achieve 90% power.

Calcolo della numerosità campionaria necessaria per conseguire una potenza adeguata per …

1) confrontare due medie campionarie (test t di Student)
2) valutare se una variazione media è significativa (test t di Student per dati appaiati)
3) **confrontare due proporzioni campionarie (test del chi-quadrato)**

Calcolo della numerosità campionaria necessaria per conseguire una precisione adeguata per …

1) ottenere un intervallo di confidenza della media di ampiezza non superiore ad un dato valore
2) ottenere un intervallo di confidenza della prevalenza di ampiezza non superiore ad un dato valore

Variabile di risposta qualitativa:
Confronto fra 2 proporzioni

$$n > \left[\frac{z_{\alpha/2}\sqrt{2\pi(1-\pi)} + z_\beta\sqrt{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)}}{\pi_1-\pi_2} \right]^2$$

dove **n = numerosità necessaria per ciascuno dei 2 gruppi**
$z_{\alpha/2}$ = 1.96 **per alfa = 5%**
z_β = 0.842, 1.282, 1.645 **per potenza = 80, 90, 95%**
$\pi = (\pi_1 - \pi_2) / 2$
π_1 = probabilità attesa nella popolazione dei controlli, desunta da studi pilota o dalla letteratura
$\pi_1 - \pi_2$ = differenza minima clinicamente rilevante, che si vuole evidenziare con una potenza 1-β
π_2 = probabilità attesa nel gruppo trattato

Esempio: Nei pazienti affetti da tumore X in stadio Y, la sopravvivenza a 5 anni è del 30% con il trattamento standard. Dati preliminari suggeriscono che nei pazienti sottoposti ad un nuovo trattamento la sopravvivenza salga al 40%. Si adotta un alfa del 5% e una potenza dell' 80%, pertanto $z_{\alpha/2}$ = 1.96 e z_β = 0.842.

$$n > \left[\frac{z_{\alpha/2}\sqrt{2\pi(1-\pi)} + z_\beta\sqrt{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)}}{\pi_1-\pi_2} \right]^2$$

$$n > \left[\frac{1.96\sqrt{2\ 0.35(1-0.35)} + 0.84\sqrt{0.3(1-0.3)+0.4(1-0.4)}}{0.3-0.4} \right]^2$$

$$n > 355.94 \qquad n \geq 356$$

Occorrono almeno 356 soggetti per gruppo.

Calcolo della numerosità campionaria necessaria per conseguire una potenza adeguata per …

1) confrontare due medie campionarie (test t di Student)
2) valutare se una variazione media è significativa (test t di Student per dati appaiati)
3) confrontare due proporzioni campionarie (test del chi-quadrato)

Calcolo della numerosità campionaria necessaria per conseguire una precisione adeguata per …

1) ottenere un intervallo di confidenza della media di ampiezza non superiore ad un dato valore
2) **ottenere un intervallo di confidenza della prevalenza di ampiezza non superiore ad un dato valore**

Variabile di qualitativa:

calcolo dell'intervallo di confidenza di una prevalenza

$$n > \pi\,(1-\pi)\left[\frac{2\,z_{\alpha/2}}{\delta}\right]^2$$

dove n = numerosità di ciascuno dei due gruppi

$z_{\alpha/2}$ = 1.96 per alfa = 5%

π = prevalenza attesa sulla base di uno studio pilota o della letteratura scientifica

δ = ampiezza massima accettabile per l'intervallo di confidenza

DA DOVE DERIVA QUESTA FORMULA?

limite superiore IC   limite inferiore IC

$$(p + z_{\alpha/2} \cdot\sqrt{p(1-p)/n}) - (p - z_{\alpha/2} \cdot\sqrt{p(1-p)/n}) \leq \delta$$

$$2\, z_{\alpha/2} \cdot\sqrt{p(1-p)/n} \leq \delta$$

divido il I e il II membro per 2 $z_{\alpha/2}$

$$\sqrt{p(1-p)/n} \leq \delta /(2\, z_{\alpha/2})$$

elevo il I e il II membro al quadrato

$$p(1-p)/n \leq \delta^2 /(2\, z_{\alpha/2})^2$$

moltiplico per n e divido per il II membro

$$p(1-p)\,(2\, z_{\alpha/2})^2/ \delta^2 \leq n$$

Esempio: Si vuole stimare la prevalenza (probabilità) di asma in una popolazione. Dati preliminari provenienti dalla letteratura suggeriscono che la prevalenza di asma si aggiri intorno al 5%. Qual è la numerosità campionaria necessaria per ottenere un intervallo di confidenza al 95% di ampiezza inferiore o uguale al 2%?

Dati: $\pi = 0.05$ $1-\alpha = 95\%$ $z_{\alpha/2} = 1{,}96$
ampiezza IC $\leq 2\%$ $n = ?$

$n \geq 0{,}05*0{,}95 * (2 * 1{,}96)^2/ 0{,}02^2$

$n \geq 0{,}0475 * (3{,}92)^2/ 0{,}0004$

$n \geq 0{,}0475 * 15{,}36 / 0{,}0004$

$n \geq 1824{,}76$ **$n \geq 1825$**

4th problem: Using the Confidence Interval to plan sample size.

A study aims at computing the prevalence (probability) of asthma in a population. Preliminary data from the current literature suggest that asthma prevalence could be about 6%. Which is the sample size necessary to estimate asthma prevalence with a width of the 95% confidence interval not greater than 2% ?

Data: $p = 0.06$ $1-\alpha = 95\%$ $z_{\alpha/2} = 1.96$ CI width $\leq 2\%$ $n = ?$

$(p + z_{\alpha/2} \cdot\sqrt{p(1-p)/n}) - (p - z_{\alpha/2} \cdot\sqrt{p(1-p)/n}) \leq \delta$

$2\, z_{\alpha/2} \cdot\sqrt{p(1-p)/n} \leq \delta$ dividing by $2\, z_{\alpha/2}$

$\sqrt{p(1-p)/n} \leq \delta / (2\, z_{\alpha/2})$ squaring the equation

$p(1-p)/n \leq \delta^2 / (2\, z_{\alpha/2})^2$ multiplying by n and dividing by the 2nd member

$p(1-p)\, (2\, z_{\alpha/2})^2 / \delta^2 \leq n$

$n \geq 0.06 * 0.94 * (2 * 1.96)^2 / 0.02^2$ $n \geq 0.0564 * (3.92)^2 / 0.0004$

$n \geq 0.0564 * 15.36 / 0.0004$ **$n \geq 2166.58$** **$n \geq 2167$**