

Dimensione campionaria e piano di analisi statistica

Importanza della dimensione campionaria

- Campione troppo piccolo →
 - Rischio falsi negativi
 - Stima poco precisa
 - Sovrastima dell'effetto
 - No conclusioni
- Campione troppo grande →
 - Pazienti coinvolti inutilmente
 - Pazienti trattati con placebo
 - Effetti collaterali
 - Spreco di risorse

Nature

NATURE REVIEWS | NEUROSCIENCE VOLUME 14
| MAY 2013 | 365

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button^{1,2}, John P. A. Ioannidis³, Claire Mokrysz⁴, Brian A. Nosek⁵, Jonathan Flint⁶, Emma S. J. Robinson⁶ and Marcus R. Munafò⁷

Abstract | A study with low statistical power has a reduced chance of detecting a true effect, but it is less well appreciated that low power also reduces the likelihood that a statistically significant result reflects a true effect. Here, we show that the average statistical power of studies in the neurosciences is very low. The consequences of this include overestimates of effect size and low reproducibility of results. There are also ethical dimensions to this problem, as unreliable research is inefficient and wasteful. Improving reproducibility in neuroscience is a key priority and requires attention to well-established but often ignored methodological principles.

Domande

- QUANTI SOGGETTI MI SERVONO?
- Altre domande!
 - Qual è l'obiettivo primario?
 - Qual è l'esito/outcome (misurabile)?
 - Che entità avrà l'esito?
 - Più obiettivi?
 - Che analisi si vogliono fare?
 - Quanti soggetti si possono realisticamente arruolare?

Come si calcola

- Il metodo per il calcolo della dimensione campionaria dipende dall'OBIETTIVO PRIMARIO dello studio e dal tipo di ESITO/outcome misurato
- I casi più frequenti sono:
 - **Confronto tra gruppi** (due o più):
 - **Quantità** (es. medie, mediane)
 - **Presenza/assenza** di caratteristiche (es. proporzioni)
 - **Rapporto tra rischi** (es. RR, HR)
 - **Descrizione della popolazione**
 - **Quantità** (es. media, mediana)
 - **Presenza** di caratteristiche (es. prevalenza, %)
- **Casi particolari:**
 - Studi sperimentali di fase I e II (metodi specifici)

Come si calcola - presupposti

- è necessario avere una idea a priori sul problema da studiare
 - Revisione della letteratura, valutazione di studi simili
 - Estrapolazione di dati di studi pilota
 - Stima dell'effetto atteso basata su parere di un esperto
 - Differenza clinicamente rilevante
- **Se non ho alcun dato**
 - Programmo uno studio pilota
 - Calcolo approssimativo basato su ipotesi sull'"Effect size" (stima dell'effetto che tiene conto della variabilità)
 - piccolo, medio, grande

Parametri - Confronto tra gruppi

Livello alpha	<ul style="list-style-type: none"> • Più è basso, meglio è (meno falsi positivi) • Tipicamente 5%
Potere (1-beta)	<ul style="list-style-type: none"> • Più è alto, meglio è (meno falsi negativi) • Tipicamente 80%-90% (→beta 10%-20%)
Test a una o due code	<ul style="list-style-type: none"> • A due code, a meno di non avere ottimi motivi
Effetto atteso	<ul style="list-style-type: none"> • Differenza tra i gruppi/frequenza dell'evento • Clinicamente/biologicamente rilevante • Variabilità
Drop-out	<ul style="list-style-type: none"> • Stima della % di soggetti che non concluderanno lo studio (persi al follow-up)

Esempio - QoL

1)

$\alpha=0.05$ $\beta=0.2$	Media	Dev standard 1	Dimensione campione*
Gruppo 1	70	10	17
Gruppo 2	80	10	17
Differenza	10		

2)

$\alpha=0.05$ $\beta=0.2$	Media	Dev standard 2	Dimensione campione*
Gruppo 1	70	15	37
Gruppo 2	80	15	37
Differenza	10		

*test a due code, $(Z_{\alpha/2} + Z_{\beta})^2 \cdot 2 \cdot ds^2 / diff^2$

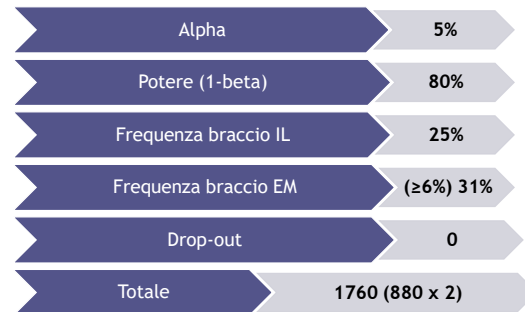
Esempio

STUDY PROTOCOL
GINEXMAL RCT: Induction of labour versus expectant management in gestational diabetes pregnancies

Esito primario: Incidenza (%) di parti cesarei nei due gruppi
Attese: % di parti cesarei inferiore nel gruppo sperimentale, rispetto al gruppo di controllo)

A total sample of 1760 patients (N = 880 per intervention group) will be recruited. The sample size has been estimated to be sufficiently large to achieve significant results with concerns to the main outcome of the study. In particular, it will be able to demonstrate a difference between the two arms $\geq 6\%$ (31% of C-section in the expectant group and 25% in the induction group; relative difference between the 2 groups equal to 20% in favour of induction [13]), considering α error equal to 5% and 80% power.

Esempio - cfr tra gruppi (2 percentuali)



Parametri - studio descrittivo

Intervallo di Confidenza (IC)	<ul style="list-style-type: none"> • Due valori (minimo e massimo) tra i quali, con una certa probabilità, si trova il vero valore nella popolazione di riferimento • Di solito IC al 95%
Valore atteso	<ul style="list-style-type: none"> • Frequenza o valore che mi attendo di trovare nel campione
Ampiezza dell'IC	<ul style="list-style-type: none"> • Precisione che vorrei avere nella stima del valore nel campione
Popolazione di riferimento	<ul style="list-style-type: none"> • Popolazione target da cui viene estratto il campione • Si può usare una popolazione infinita

Esempio - studio descrittivo

OPEN ACCESS Freely available online

PLOS ONE

The Submerged Dyslexia Iceberg: How Many School Children Are Not Diagnosed? Results from an Italian Study

Chiara Barbiero^{1,2}, Isabella Lonciari², Marcella Montico¹, Lorenzo Monasta¹, Roberta Penge³, Claudio Vio⁴, Patrizio Emanuele Tressoldi⁵, Valentina Ferluga¹, Anna Bigoni¹, Alessia Tullio¹, Marco Carrozzi², Luca Ronfani^{1*}, for the CENDI (National Committee on the Epidemiology of Dyslexia) working group and for the Epidemiology of Dyslexia of Friuli Venezia Giulia working group (FVGwg)¹

To determine the sample size, the total population of children attending grade four in the Region (9687 children) was considered. A sample size of 1500 children was estimated, hypothesizing a prevalence of dyslexia of 4%, ranging from 3% to 5%, with a precision of 5% and a power of 80%. An extra 15% were enrolled to compensate possible drop outs. Based on the average number of children per class, we decided to randomly extract clusters of grade four classes rather than children. 94 classes were thus randomly extracted, comprising 1774 children.

Esempio - studio descrittivo

Popolazione di riferimento	9687
Frequenza del disturbo	4%
Intervallo di confidenza (IC)	al 95%
Ampiezza dell'IC al 95%	3% - 5% (4% ± 1%)
Drop-out	15%
Totale	1480

www.openepi.com

Dimensione del Campione per Frequenza % in una Popolazione (Campione Random)		
Dimensione della popolazione	10000	Se ampia, lascia un milione
Frequenza(p) % Anticipata	4	Tra 0 & 99,99. Se sconosciuta, usa 50%
Limiti di confidenza +/- percento di 100	1	Precisione % assoluta
Effetto del progetto (per complesse indagini campione-DEFF)	1	1.0 per campione random

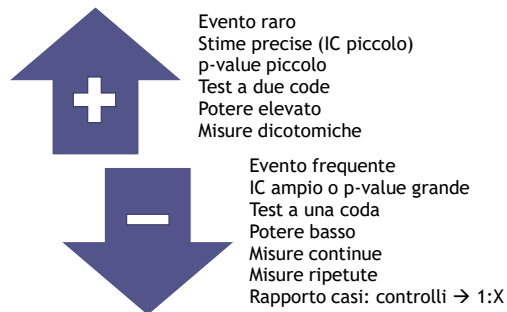
Dimensione del Campione per Frequenza in una Popolazione

Dimensione della popolazione (fattore di correzione della popolazione finita o fpc)(N): 10000
 Frequenza % ipotizzata del fattore di outcome nella popolazione (p): 4% +/- 1
 Limiti di confidenza come % di 100(assoluto +/- %)(d): 1%
 Effetto del progetto (per indagini di grappoli)-DEFF: 1

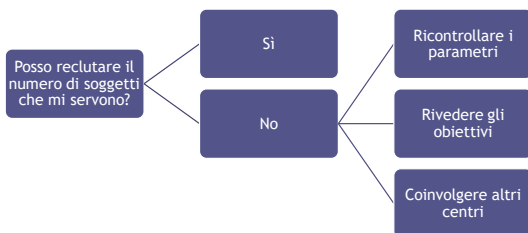
ConfidenzaLivello(%)	Dimensione del Campione
95%	1286
80%	594
90%	942
97%	1532
99%	2031
99.9%	2937
99.99%	3677

Equazione
 Dimensione del campione $n = [DEFF * Np(1-p)] / [(d^2 / Z_{1-\alpha/2}^2 * (N-1) + p * (1-p))]$

Da cosa dipende la numerosità



Fattibilità



Analisi statistica

Analisi statistica - indicazioni generali

- Indicare come verranno sintetizzati i dati (coerentemente con il tipo di variabili!):
 - Frequenze assolute e relative (i.e. N e %),
 - Indici di posizione (es. media, mediana, quartili)
 - Indici di variabilità (es. deviazione standard, range, range interquartile)
 - Indici di correlazione (es. r di Pearson)
 - Misure di rischio (es. tasso di incidenza, sopravvivenza)
 - Rapporti tra rischi (es. odds ratio, risk ratio, hazard ratio)
 - ...
- Se opportuno riportare anche l'intervallo di confidenza
 - (es. IC al 95% per OR, percentuali ecc...)

Analisi statistica - indicazioni generali

- Le analisi vanno definite per ogni outcome (primario e secondari)
 - Indicare
 - Tipo di test (tenere conto del tipo di variabile)
 - Se test a una o due code
 - Livello di significatività (p-value)
 - Definire gruppi e sottogruppi su cui verranno fatte le analisi
 - Es: analisi stratificata per sesso, sottogruppi di patologie
 - Importante nei trial randomizzati

Analisi statistica - indicazioni generali

- Le analisi devono essere coerenti con l'obiettivo principale e con quanto scritto nella definizione della dimensione campionaria
- Se viene fatta anche un'analisi multivariata (→ stime 'aggiustate' per i fattori confondenti), indicare la tecnica usata
 - es. **Regressione logistica, di Poisson, di Cox**

NB: non può essere fatta se i soggetti sono troppo pochi!

Esempio

- "I dati saranno sintetizzati utilizzando indici di posizione (medie, mediane) e di variabilità (deviazioni standard, range interquartile) se relativi a variabili quantitative e mediante frequenze assolute e relative per le variabili qualitative o rese categoriche"
- "Le curve di sopravvivenza grezze saranno stimate con il metodo di Kaplan-Meier. Le mediane di sopravvivenza saranno riportate con i rispettivi intervalli di confidenza (IC) al 95%."

Esempio

- "Le differenze in termini di sopravvivenza tra gruppi di interesse saranno testate utilizzando il log-rank test, considerando come statisticamente significativo un p-value inferiore a 0.05"
- "I rischi di morte nei gruppi di interesse, aggiustati per i potenziali fattori confondenti (es. età, sesso), saranno confrontati utilizzando gli Hazard Ratio (HR), con i rispettivi IC al 95%. Gli HR aggiustati verranno stimati attraverso modelli di regressione a rischi proporzionali di COX (previa verifica della validità dell'assunto di proporzionalità dei rischi!!)

Esempio

- "Per valutare l'obiettivo primario dello studio, la differenza tra i due gruppi verrà analizzata con il test del chi-quadrato"
- "L'effetto dei possibili confondenti verrà controllato utilizzando una regressione logistica multivariata"
- "In caso di variabili continue, le differenze tra i due gruppi verranno valutate con il T-Test o con il test di Mann-Whitney se non sarà verificato l'assunto di normalità della distribuzione"