

## **“R” COME RIDUZIONE: IL DIMENSIONAMENTO DELLO STUDIO È MENO DIFFICILE (E PIÙ UTILE) DI QUANTO SI PENSI**

Laura Ricceri, Flavia Chiarotti

Centro di Riferimento Scienze Comportamentali e Salute Mentale, Istituto Superiore di Sanità, Roma

Molti ricercatori italiani per svolgere attività sperimentale *in vivo* devono presentare un progetto di ricerca al Ministero della Salute in cui è necessario fornire dettagli relativi al numero di animali che si intende utilizzare, specificando *come è stato determinato il numero di animali necessari per lo studio*. È un errore considerare con fastidio la richiesta dello studio di dimensionamento del campione, sia perché proprio essa può rivelarsi un’opportunità di progettazione di un’attività sperimentale che miri alla *massima riduzione del numero di animali utilizzati compatibile con gli obiettivi del progetto*, sia perché si tratta di una richiesta in linea con documenti di indirizzo internazionali (1, 2) relativi alla sperimentazione animale mirati sia alla pubblicazione di lavori scientifici sia alla presentazione di progetti per il finanziamento della ricerca. Il corretto dimensionamento dello studio è stato di recente uno degli argomenti trattati in un ampio e più esteso dibattito sulla riproducibilità degli studi preclinici e sul loro effettivo valore traslazionale; da più parti si è sottolineata l’importanza di un corretto dimensionamento per la riproducibilità degli studi preclinici (3, 4).

### **Formulazione statistica del problema**

Da un punto di vista statistico, per ottenere la “R” di Riduzione si debbono disegnare gli esperimenti, fare studi di dimensionamento e analizzare i risultati in modo appropriato. Nel Riquadro 1 abbiamo riportato un glossario ragionato degli elementi del disegno sperimentale; i termini sono poi utilizzati nel resto del testo. Gli elementi di statistica di base che è necessario conoscere per effettuare un corretto dimensionamento di uno studio sperimentale sono riportati nel Riquadro 2. In questo contesto è importante mettere a fuoco che la dimensione campionaria prescelta ha un’influenza diretta sulla potenza dell’analisi statistica, in ultima analisi sull’affidabilità dei risultati dello studio che viene condotto, ed è quindi necessario che sia stimata appropriatamente. Da un punto di vista pratico, per effettuare *a priori* la stima della dimensione campionaria in uno studio sperimentale è necessario conoscere 4 elementi (Riquadro 3 e Figura 1). Due di questi hanno usualmente valori prestabiliti: per  $\alpha$  si utilizza solitamente il valore di 0,05 (più raramente 0,01); per la potenza del test ( $1-\beta$ ) si utilizza solitamente il valore di 0,80. Viene scelto questo valore di potenza perché risulta sufficiente ad evidenziare come statisticamente significativi effetti grandi (in caso di variabili quantitative, effetti corrispondenti a differenze tra medie dei gruppi  $\geq$  a una deviazione standard, *Standard Deviation*, SD). È interessante notare che valori di potenza maggiore lasciano sostanzialmente invariata la capacità discriminativa del test nel caso di effetti grandi, mentre aumentano la capacità discriminativa del test solo nel caso di effetti medi e piccoli (al di sotto di 1 SD), solitamente meno rilevanti da un punto di vista biologico, e questo al costo di un notevole aumento della dimensione campionaria (5). Valori di potenza  $>0,80$  devono quindi essere presi in considerazione solo in casi particolari e con grande cautela in una prospettiva di Riduzione.

**RIQUADRO 1****Disegno sperimentale: glossario ragionato degli elementi****Disegno sperimentale**

Per disegno sperimentale di uno studio si intende la struttura secondo cui lo studio viene condotto, ovvero selezione delle unità statistiche, dei fattori di cui si vuole valutare l'effetto, del metodo di attribuzione delle unità statistiche ai diversi livelli di tali fattori, delle variabili di interesse e le loro caratteristiche. Tali elementi influiscono sul metodo di rilevazione delle osservazioni sulle unità statistiche, sul tipo di analisi statistica dei dati appropriata e sulla presentazione dei risultati.

**Unità statistica**

È l'unità più piccola a cui possa essere assegnato un trattamento. In generale, la scelta dell'unità statistica dipende dall'ipotesi che si vuole verificare con lo studio. Gli studi *in vivo* vengono solitamente condotti su soggetti umani o animali; gli studi *ex vivo* su campioni biologici, quali ad esempio sangue, urina, saliva; gli studi *in vitro* su colture cellulari. Le unità statistiche possono essere fra loro indipendenti o dipendenti. Due unità statistiche si dicono indipendenti quando l'osservazione effettuata su una di esse non fornisca alcun elemento utile a prevedere quale possa essere l'osservazione sull'altra unità. Due unità statistiche si dicono dipendenti quando l'osservazione effettuata su una di esse permetta di prevedere quale sarà l'osservazione sull'altra unità. Ad esempio, nel caso della sperimentazione animale due unità statistiche sono da considerarsi dipendenti quando i due animali siano fratelli, cioè provengano dalla stessa nidiata. Un altro caso, meno evidente, di dipendenza si ha quando gli animali appartengano allo stesso gruppo sociale (nel quale si siano stabiliti e consolidati rapporti di gerarchia) e vengano ad esempio testati per comportamenti di interazione sociale.

**Fattore/i in studio**

Il fattore in studio può essere uno (disegno uni fattoriale) o più di uno (disegno multifattoriale): ciascuno dei fattori può includere da due a più livelli. Nel caso più semplice in cui il fattore includa solo due livelli, solitamente uno è il livello di riferimento o controllo, e l'altro è il livello di interesse per lo sperimentatore.

**Attribuzione delle unità statistiche ai livelli del fattore in studio**

Negli studi sperimentali, l'attribuzione è sotto il controllo dello sperimentatore, che deve preferibilmente adottare una procedura di assegnazione casuale della singola unità al singolo livello (nel caso di disegno uni fattoriale) o alla combinazione di livelli (nel caso di disegno multifattoriale). La procedura di assegnazione casuale è detta randomizzazione.

**Variabile di interesse**

La variabile di interesse può essere descritta in termini di scala di misurazione, contenuto numerico o non numerico, e delle relazioni che è possibile stabilire tra due osservazioni:

Scala	Contenuto	Relazioni
Nominale	Qualitativo, non ordinabile	Uguaglianza/differenza
Ordinale	Qualitativo, ordinabile	Uguaglianza/differenza Maggioranza/minoranza
	Quantitativo, numerico (origine dell'asse convenzionale = zero non assoluto)	Uguaglianza/differenza Maggioranza/minoranza Somma/differenza
Di rapporti	Quantitativo, numerico (origine dell'asse assoluta = zero assoluto o "fisico")	Uguaglianza/differenza Maggioranza/minoranza Somma/differenza Prodotto/rapporto

Dal tipo di variabile dipendono i metodi di analisi statistica che possono essere utilizzati per descrivere i dati e per valutare le ipotesi in studio.

**RIQUADRO 2****Elementi di statistica inferenziale****Test statistico di ipotesi (inferenza classica)**

Serve a verificare ipotesi relative a fenomeni in una popolazione, utilizzando dati sperimentali ottenuti su un campione. Perché i risultati campionari possano essere estesi alla popolazione occorre che gli elementi che compongono il campione siano un sottogruppo casuale degli elementi della popolazione.

**Ipotesi**

Il test di ipotesi non consente di confermare l'ipotesi oggetto del test (detta, ma solo eventualmente di rifiutarla. Per questo, lo sperimentatore dovrà porre alla base del test l'ipotesi complementare rispetto a quella di suo interesse in modo che, rifiutandola, confermi implicitamente l'ipotesi di suo interesse. Le due ipotesi vengono chiamate ipotesi nulla (indicata con  $H_0$ ), che viene valutata con il test, e ipotesi alternativa (indicata con  $H_1$ ), di interesse per lo sperimentatore:

- *Ipotesi nulla  $H_0$*   
Solitamente,  $H_0$  ipotizza che il(i) campione(i) provenga(n) da una data popolazione, cioè che non vi siano differenze, correlazioni, effetti nel gruppo (o nei gruppi) a confronto.
- *Ipotesi alternativa  $H_1$*   
 $H_1$  può essere uni- o bi-direzionale:  $H_1$  unidirezionale implica che lo sperimentatore sia interessato a dimostrare l'esistenza di un effetto in una sola specifica direzione (es. il nuovo farmaco un'efficacia maggiore rispetto al farmaco di controllo);  $H_1$  bidirezionale implica che lo sperimentatore sia interessato a dimostrare l'esistenza di un effetto in una qualunque direzione (es. il nuovo farmaco ha un'efficacia diversa rispetto al farmaco di controllo).

**Risultato del test: probabilità di Errore di I e II tipo e Potenza del test**

Qualunque sia la realtà ( $H_0$  vera o  $H_0$  falsa) il test potrà portare o a rifiutare o a non rifiutare  $H_0$ . Relativamente alla realtà, il test potrà quindi portare a una decisione corretta o a una decisione sbagliata. In particolare, si potranno commettere due tipi di errore:

- Errore di I tipo = in base al risultato del test si rifiuta  $H_0$ , quando in realtà  $H_0$  è vera
- Errore di II tipo = in base al risultato del test non si rifiuta  $H_0$ , quando in realtà  $H_0$  è falsa.

Risultato del test	Realtà	
	$H_0$ vera	$H_0$ falsa
Non rifiuto $H_0$	$1-\alpha$	$\beta$
Rifiuto $H_0$	$\alpha$	$1-\beta$

Il test permette di quantificare le probabilità di errore di I e II tipo e, conseguentemente, la potenza del test:

- $\alpha$  Prob(rifiutare  $H_0|H_0$  vera) = probabilità di errore di I tipo;
- $\beta$  Prob(non rifiutare  $H_0|H_0$  falsa) = Prob(rifiutare  $H_1|H_1$  vera) = probabilità di errore di secondo tipo;
- $1-\beta$  Prob(rifiutare  $H_0|H_0$  falsa) = Prob(rifiutare  $H_0|H_1$  vera) = potenza del test.

La probabilità di errore di I tipo è tenuta sotto controllo direttamente, stabilendo il valore di soglia per la significatività statistica ( $p$ ; usualmente, si pone  $p \leq 0,05$ ).

La probabilità di errore di II tipo (e, conseguentemente, la potenza del test) viene controllata indirettamente, intervenendo sulla variabilità della misura ( $SD \sigma$ ), sulla dimensione dell'effetto che si vuole essere in grado di individuare ( $\delta$ ), sulla dimensione campionaria (dimensionamento del campione).

**RIQUADRO 3****Dimensionamento del campione**

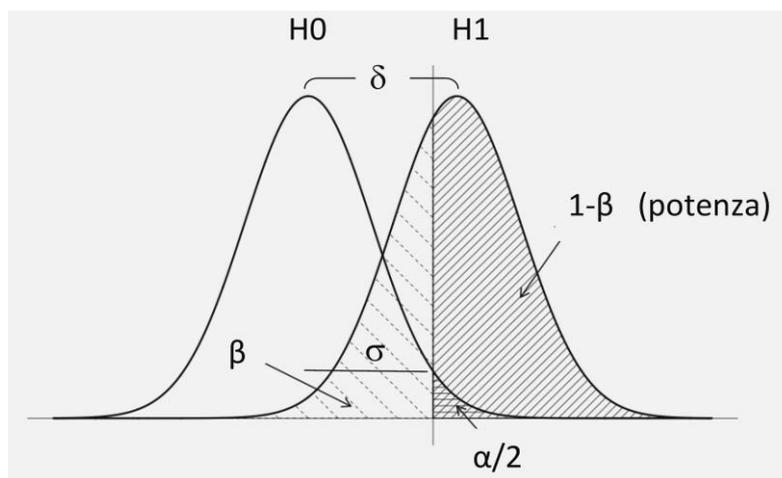
Per dimensionamento del campione si intende il calcolo della minima dimensione campionaria necessaria a valutare le ipotesi in studio ( $H_0$  e  $H_1$ ) con un livello di significatività  $\alpha$  e con una potenza ( $1-\beta$ ) prestabiliti, e con riferimento al test per la valutazione dell'ipotesi in studio appropriato al disegno sperimentale (uno, due o più gruppi a confronto; unità statistiche indipendenti o appaiate; rispetto o violazione degli assunti di normalità e omogeneità della varianza per la variabile risposta prescelta).

**Elementi per il dimensionamento del campione**

Per poter procedere occorre stabilire:

Elemento	Significato	Valore da inserire nel calcolo
$\alpha$	Livello di significatività al quale si vuole condurre il test (valore massimo accettabile)	<u>0,05</u> , 0,01
$H_1$	Tipologia di $H_1$ (bidirezionale / unidirezionale), che si riflette sul tipo di test (2 code / 1 coda)	<u>Test a due code</u> Test a una coda
$1-\beta$	Potenza del test (valore minimo accettabile)	<u>0,80</u> , 0,90, 0,95
$\sigma$	SD della variabile da esaminare nel o nei gruppi in studio	Si ricava da studi precedenti (propri o presi dalla letteratura), o realizzando un piccolo studio pilota
$\delta$	Dimensione dell'effetto che si vuole essere in grado di evidenziare, perché rilevante ai fini dello studio (da un punto di vista clinico, biologico, ecc.)	A scelta, ragionata e motivata, dello sperimentatore

Sono sottolineati i valori standard da utilizzare per il dimensionamento. Gli altri valori riportati sono comunque accettabili ma, poiché comportano o un aumento anche consistente del numero di soggetti sperimentali, o un vincolo sull'interpretazione dei risultati del test, la loro scelta deve essere adeguatamente motivata.



**Figura 1. Elementi per il dimensionamento del campione  $\alpha$ ,  $\beta$  e  $(1-\beta)$ ,  $\sigma$ ,  $\delta$ . Sull'asse delle  $x$  è riportato il valore della statistica del test, sull'asse delle  $y$  il livello di probabilità**

Il terzo elemento che è necessario conoscere, sigma  $\sigma$ , corrisponde alla SD della variabile considerata; poiché si sta operando un'analisi *a priori* è evidente che i valori della SD della variabile non sono ancora noti allo sperimentatore, che deve quindi farne una stima sulla base di studi precedenti che hanno coinvolto quella variabile. Nel caso in cui si può ipotizzare che il trattamento e/o le condizioni sperimentali aumentino la variabilità si può aumentare il valore presunto del sigma per rendere la stima più accurata.

Il quarto elemento è quello forse più "ostico" e che mette maggiormente in difficoltà il ricercatore: il valore  $\delta$  infatti dovrebbe dipendere completamente da quale sia l'effetto minimo che verrebbe considerato rilevante nell'esperimento che deve ancora essere condotto, e che può divergere (anche molto) da quanto è stato fatto in precedenza da altri o dallo stesso ricercatore. Nel caso di studi di valutazione degli effetti farmacologici di qualche nuova molecola, la presenza di altri farmaci già disponibili sul mercato con effetto noto può fornire valide informazioni sulle dimensioni dell'effetto che possa essere considerato rilevante (es. abbassamento pressione sistolica del 20% del valore pre-trattamento, a cui corrisponde, partendo da un valore di pressione pari a 180 mg/Hg, una riduzione  $\delta=36$  mg/Hg, che porterebbe la pressione a 144 mg/Hg, un valore sufficiente a ridurre consistentemente l'incidenza di eventi cardiovascolari pressione-correlati). Nel caso invece di un farmaco destinato a una malattia assai invalidante per la quale non siano disponibili altre terapie può essere considerato rilevante anche un effetto di minori dimensioni.

Più difficile può risultare valutare la dimensione dell'effetto in ambiti meno vicini ad aspetti biomedici/farmacologici. Nel caso in cui non si abbia alcuna idea della dimensione minima da considerare, si può scegliere  $\delta$  in funzione di  $\sigma$ . Per esempio, per la differenza fra medie Cohen (6) propone dei valori di riferimento per l'indice  $d = \delta/\sigma$ : effetto piccolo:  $d$  da 0,2 a 0,4 (da cui  $\delta$  da  $0,2 \cdot \sigma$  a  $0,4 \cdot \sigma$ ); effetto medio:  $d$  da 0,5 a 0,7 ( $\delta$  da  $0,5 \cdot \sigma$  a  $0,7 \cdot \sigma$ ); effetto grande:  $d$  da 0,8 a 1,0 ( $\delta$  da  $0,8 \cdot \sigma$  a  $1,0 \cdot \sigma$ ). Ovviamente, minore è la dimensione dell'effetto che si vuole essere in grado di individuare, maggiore sarà la dimensione campionaria necessaria.

La dimensione dell'effetto (*Standardised Effect Size*, SES) è un parametro del proprio studio che può essere utile calcolare non solo ai fini del dimensionamento. Infatti, è l'indice che viene utilizzato negli studi di meta-analisi, permettendo di confrontare studi caratterizzati da diversi contesti e disegni sperimentali, e da diverse variabilità. Recentemente, SES viene anche impiegato per interpretare l'eventuale discrepanza tra ordine di grandezza degli effetti e significatività/non significatività statistica (valore  $p$ ). Studi sovradimensionati possono portare a risultati significativi anche quando la dimensione dell'effetto sia minima, tale da risultare irrilevante da un punto di vista biologico/clinico; al contrario, nel caso di studi inevitabilmente sottodimensionati come quelli che riguardano malattie rare e farmaci orfani, effetti anche consistenti da un punto di vista biologico/clinico potrebbero non raggiungere la significatività statistica. SES è in tutti questi casi molto più informativo del solo valore  $p$  (7).

## Un esempio di dimensionamento del campione

A titolo di esempio, consideriamo un esperimento volto a verificare l'effetto della delezione di un gene sull'attività locomotoria di roditori. Allo scopo, piccoli di topo di due genotipi diversi, *wild-type* (WT = gruppo 1) ed emizigoti per il gene di interesse (HZ = gruppo 2), vengono sottoposti a un test di *open-field* della durata di 20 minuti, durante i quali viene rilevata la distanza percorsa (cm) utilizzando il sistema computerizzato di *videotracking* del percorso tracciato dall'animale (software Ethovision, Noldus, NL).

Dai risultati di esperimenti precedenti si può ricavare la distanza media attesa percorsa dai topi WT e la relativa SD; supponiamo inoltre di aver verificato, sempre sulla base dei dati di esperimenti precedenti, che la variabile di interesse rispetti gli assunti di normalità e omogeneità delle varianze dei test parametrici.

Supponiamo che il valore della media della variabile risposta esaminata nel gruppo di controllo WT sia:

$$\text{Media gruppo 1} = 1907 \text{ cm.}$$

Supponiamo inoltre che la SD della variabile risposta nel gruppo di controllo WT sia:

$$\text{SD gruppo 1} = 381 \text{ cm.}$$

Supponiamo di poter ipotizzare che:

$$\text{SD gruppo 2} \approx \text{SD gruppo 1} = 381 \text{ cm.}$$

Supponiamo infine di volere essere in grado di evidenziare come significativo un incremento della media della variabile risposta nel gruppo sperimentale HZ pari al 20%, tale per cui sia:

$$\text{Media gruppo 2} \geq 2289 \text{ cm, cioè } \Delta = 2289 - 1907 = 382 \text{ cm.}$$

Tale valore, rapportato alla SD, dà  $\delta$  di Cohen =  $382 \text{ cm} / 381 \text{ cm} = 1.003$ , che secondo Cohen (1988) corrisponde a un effetto di dimensione grande.

Sulla base di questi dati, si può quindi procedere con il dimensionamento del campione. In particolare, si dovranno specificare gli elementi come specificato in Tabella 1.

**Tabella 1. Esempio di elementi per il dimensionamento del campione sperimentale e rispettivi valori da inserire nel calcolo**

Elementi	Valori da inserire nel calcolo
$\alpha$ (H1)	0,05 (test a due code)
1- $\beta$	0,80
Tipo di test	parametrico, per il confronto fra due gruppi
Rapporto tra la numerosità (N) dei due gruppi (N1/N2)	1
Test statistico	T test, scelto in base all'assenza o alla presenza di dipendenza tra le unità statistiche tra ed entro gruppi*

I test statistici appropriati per l'analisi dei disegni sperimentali più frequentemente utilizzati, e ai quali ci si deve quindi riferire per il dimensionamento del campione, sono riportati in Tabella 2.

Nel caso in cui si decida di adottare un disegno per gruppi dipendenti (come indicato precedentemente, ad esempio fratelli di una stessa nidiata nel caso di roditori) (8-11), occorrerà anche specificare la correlazione attesa tra le due serie di osservazioni, stimata mediante il coefficiente di correlazione lineare di Pearson. Se possibile, converrà utilizzare stime ottenute da esperimenti precedenti, ma qualora non si abbiano dati utili si potrà porre  $r = 0,5$  (correlazione positiva di moderata entità).

**Tabella 2. Selezione dei più frequenti test statistici per il confronto tra due o più livelli di un fattore in un disegno unifattoriale**

Tipo di variabile	N. livelli	Legame tra le unità statistiche	Assunti	Test	Specifiche in G*Power Test family Statistical test
Nominale	2	Indipendenza ENTRO gruppo Indipendenza TRA gruppi		Test della probabilità esatta di Fisher	Exact Proportions: Inequality, two independent groups (Fisher's exact test)
	≥2	Indipendenza ENTRO gruppo Indipendenza TRA gruppi		Test del Chi quadrato	χ <sup>2</sup> tests Goodness-of-fit tests: Contingency tables
Ordinale o quantitativa	2	Indipendenza ENTRO gruppo Indipendenza TRA gruppi	Continuità della variabile (anche se misurata nel discreto)	Test U di Mann-Whitney	t tests Means: Wilcoxon-Mann-Whitney test (two groups)
		Indipendenza ENTRO gruppo Dipendenza TRA gruppi	Continuità della variabile (anche se misurata nel discreto)	Test di Wilcoxon	t tests Means: Wilcoxon signed-rank test (matched pairs)
Quantitativa	2	Indipendenza ENTRO gruppo Indipendenza TRA gruppi	Distribuzione della variabile Normale σ omogenei tra gruppi	Test t di Student per gruppi indipendenti	t tests Means: Difference between two independent means (two groups)
		Indipendenza ENTRO gruppo Dipendenza TRA gruppi	Distribuzione della variabile Normale	Test t di Student per gruppi appaiati	t tests Means: Difference between two dependent means (matched pairs )
Quantitativa	>2	Indipendenza ENTRO gruppo Indipendenza TRA gruppi	Distribuzione della variabile Normale σ omogenei tra gruppi	ANOVA per gruppi indipendenti	F tests ANOVA: Fixed effects, omnibus, one-way
		Indipendenza ENTRO gruppo Dipendenza TRA gruppi	Distribuzione della variabile Normale	ANOVA per misure ripetute	F tests ANOVA: Repeated measures, within factors

ANOVA: Analysis Of VAriance

A questo punto abbiamo tutti gli elementi per effettuare il dimensionamento e possiamo utilizzare uno specifico software; ne esistono molti disponibili in rete sia gratuiti che a pagamento. Ne consigliamo uno gratuito e costantemente aggiornato, messo a punto da un gruppo di ricercatori dell'università di Düsseldorf, per il quale sono anche disponibili delle pubblicazioni esplicative (12, 13) e un sito con materiale informativo (<http://www.gpower.hhu.de/en.html>). Per segnalare l'importanza del disegno sperimentale riportiamo lo stesso esempio, nella situazione 1 considerando i due gruppi indipendenti, nella situazione 2 considerando i due gruppi dipendenti o appaiati (*paired*).

## Situazione 1: gruppi indipendenti

Se i piccoli di topo, sia WT sia HZ, provengono tutti da nidiate diverse, le unità statistiche saranno da considerarsi indipendenti tra ed entro gruppi. Poiché dobbiamo confrontare fra loro due soli gruppi relativamente a una variabile quantitativa che per ipotesi rispetta gli assunti alla base dei test parametrici, per il dimensionamento faremo riferimento al test t di Student per gruppi indipendenti, che nel programma *G\*Power* viene individuato selezionando le seguenti opzioni (Figura 2).

Sulla base dei dati sopra riportati la dimensione campionaria ottimale sarà di 17 unità per gruppo sperimentale, per un totale di 34 unità sperimentali. Questo valore è gestibile da un punto di vista sperimentale, ma potrebbe essere inferiore se la variabilità delle misure entro ciascun gruppo fosse minore, a parità di altre condizioni. La variabilità entro gruppo si può ridurre o cercando di essere più rigorosi nelle procedure sperimentali, o agendo sul disegno sperimentale, passando ad esempio da un disegno per gruppi indipendenti a un disegno per gruppi appaiati.

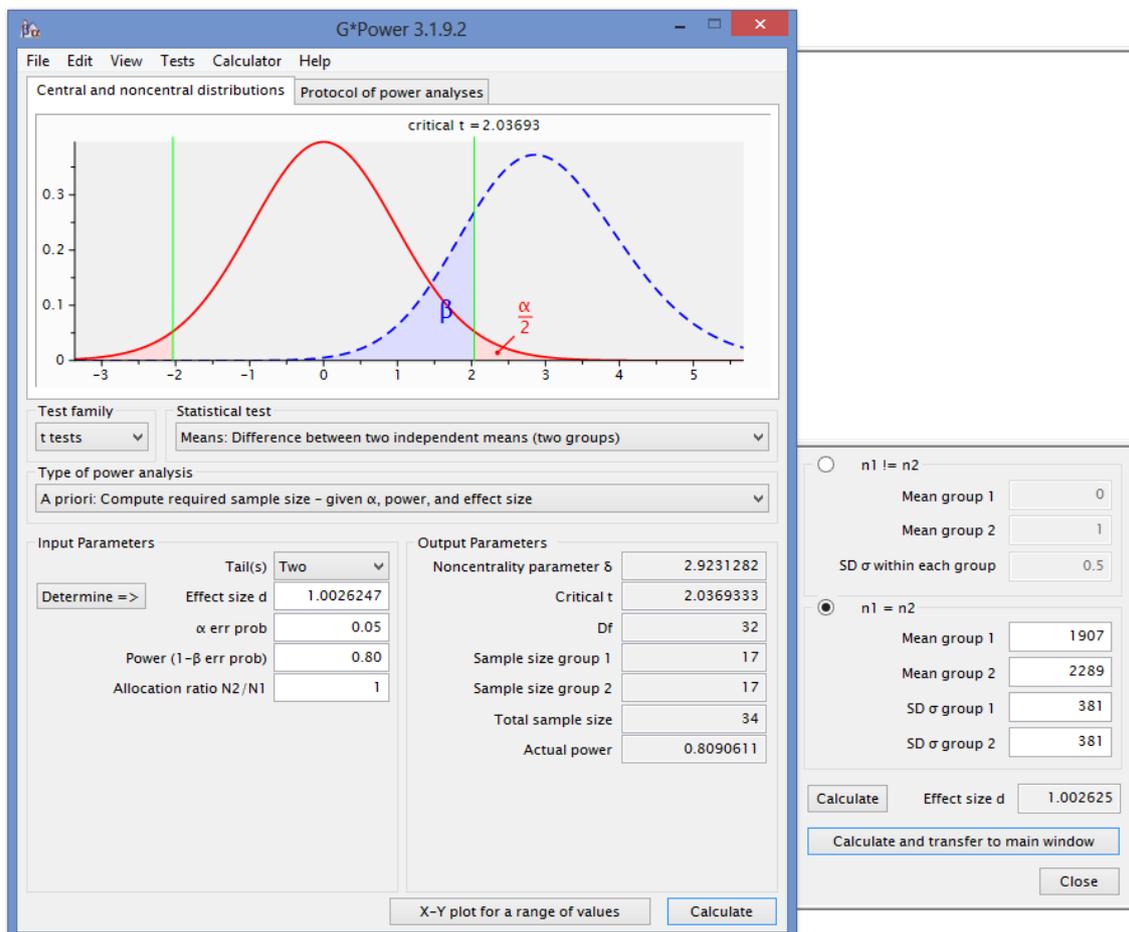


Figura 2. Schermata del programma *G\*power* relativa all'esempio per campioni indipendenti

## Situazione 2: gruppi dipendenti

Se i piccoli di topo provengono a coppie dalle stesse nidiatae (un WT e un HZ per nidiatae), le unità statistiche sono da considerarsi indipendenti entro gruppo ma dipendenti tra gruppi di genotipo. Poiché dobbiamo confrontare fra loro due soli gruppi relativamente a una variabile quantitativa, possiamo far riferimento per il dimensionamento al test t di Student per gruppi dipendenti, che nel programma *G\*Power* viene individuato selezionando le seguenti opzioni (Figura 3). Supponiamo di non avere informazioni riguardo alla correlazione tra coppie di misure, e ipotizziamo quindi che tale correlazione sia  $r = 0,50$  (correlazione positiva moderata).

A tali condizioni, la dimensione campionaria ottimale sarà di 10 coppie, per un totale di 20 unità sperimentali, con una forte riduzione rispetto al disegno sperimentale per gruppi indipendenti. La dimensione campionaria verrebbe ulteriormente ridotta se la correlazione tra coppie di misure fosse più alta: ad esempio, qualora fosse  $r = 0,70$  (buona correlazione positiva), il numero di coppie richieste a parità di altre condizioni sarebbe pari a 7, per un totale di 14 unità statistiche.

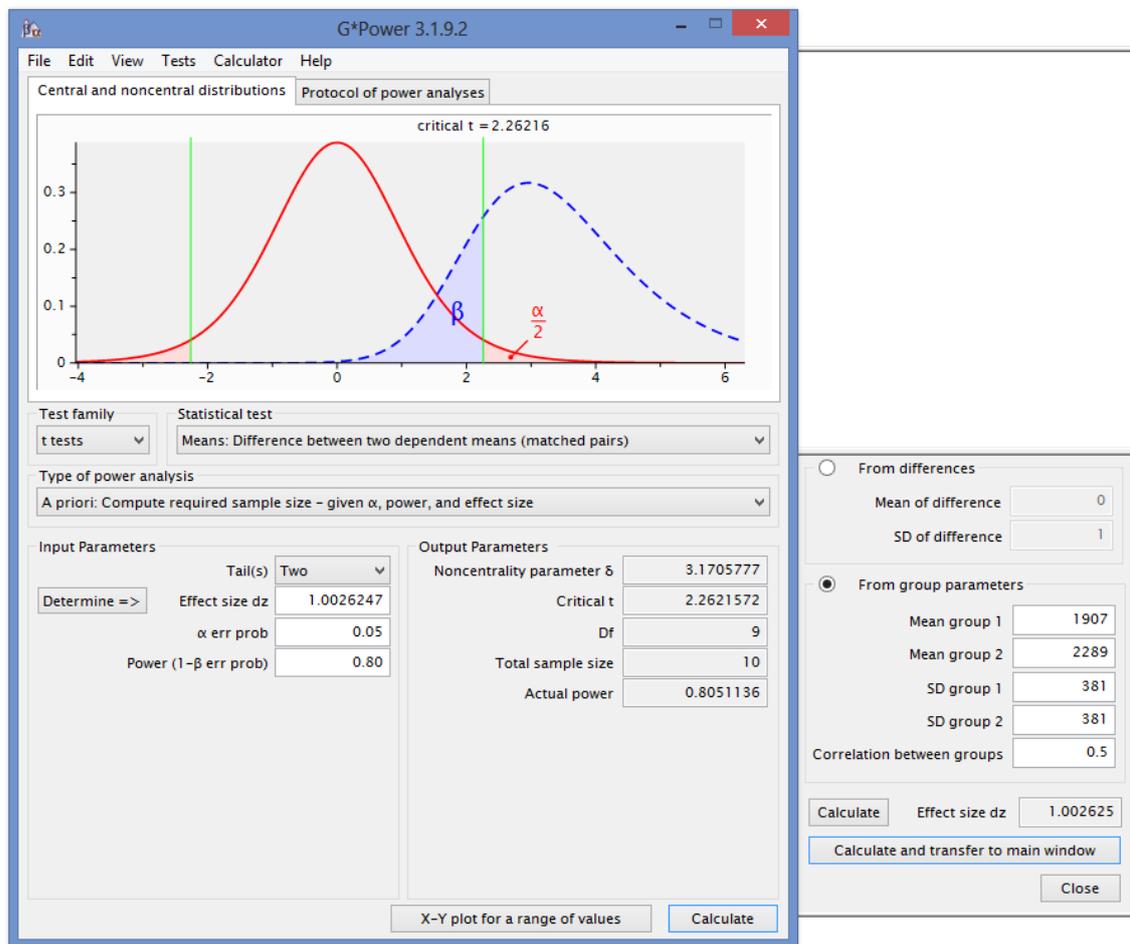


Figura 3. Schermata del programma *G\*power* relativa all'esempio per campioni dipendenti

Prima di procedere con la sperimentazione, la plausibilità biologica dei valori ipotizzati per la differenza tra le medie che si vuole essere in grado di evidenziare, la variabilità entro gruppo

e il coefficiente di correlazione, andrebbe accuratamente verificata sulla base dei risultati di esperimenti precedenti.

## Dimensionamento come valore aggiunto

A volte il numero di animali è limitato da esigenze di fattibilità dell'esperimento: in questi casi si raccomanda comunque di fare uno studio di sensibilità, per valutare la dimensione dell'effetto che si è in grado di evidenziare con la dimensione campionaria imposta e per valori di  $\alpha$  e  $1-\beta$  ritenuti accettabili. Qualora la dimensione dell'effetto risultante dal calcolo di sensibilità sia palesemente irrealistica (troppo grande per poter essere raggiunta in quel *setting* sperimentale), il ricercatore è fortemente consigliato di non effettuare lo studio, in quanto le conclusioni di esso sarebbero inaffidabili e gli animali utilizzati, anche se pochi, verrebbero sacrificati inutilmente.

La linea guida ARRIVE (*Animal Research: Reporting of in vivo Experiments*) per la sperimentazione su animali (1) costituisce un interessante tentativo della comunità scientifica del Regno Unito – in particolare dell'attivo e fattivo centro nazionale britannico per le 3R, (*National Centre for the Replacement, Refinement & Reduction of Animals in Research*, NC3Rs) – da sempre assai sensibile e attiva nella protezione del benessere degli animali da laboratorio, per diffondere tra i ricercatori la cultura del dimensionamento. Secondo questo documento – che sempre più gruppi editoriali, società scientifiche e agenzie di finanziamento della ricerca stanno adottando\* – il dimensionamento è obbligatorio e va documentato in dettaglio in ogni studio sperimentale che preveda l'utilizzo di animali. Anche la scelta del disegno sperimentale adottato deve essere esplicitata insieme ai test statistici adottati per l'analisi dei dati. Si tratta di informazioni quasi scontate, ma è opportuno ricordare che indagini condotte sempre nel Regno Unito mostrano una certa reticenza/resistenza da parte dei ricercatori su questi temi nelle loro pubblicazioni (14).

Ci piace sottolineare che in lingua inglese si parla più di *power analysis* che di *sample size estimation*, quasi a dare la precedenza all'analisi della potenza statistica *post hoc* piuttosto che alla determinazione *a priori* della dimensione campionaria, suggerendo quindi una prospettiva un po' ribaltata rispetto a quella fin qui presentata. In questo ribaltamento, su studi effettuati ci si può chiedere quale sia la potenza statistica effettivamente raggiunta con i valori di  $n$ ,  $\alpha$ ,  $\sigma$  e  $\delta$  dati. La maggiore familiarità con questi elementi di statistica e la loro conoscenza fa inoltre acquisire una chiave di lettura importante per ri-esaminare non solo i propri risultati ma anche la letteratura di riferimento, tenendo presente che non sempre, soprattutto sulla statistica, ci si può affidare al vaglio editoriale, anche nel caso di riviste prestigiose.

E' importante saper cogliere il valore aggiunto che il dimensionamento dello studio e l'analisi di potenza portano nel disegnare l'esperimento, in modo che esso possa fornire risultati affidabili e replicabili, riducendo al minimo necessario il numero di animali da utilizzare.

## Bibliografia

1. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biology* 2010;8(6):e1000412.

---

\* Per un elenco aggiornato vedi "ARRIVE endorsers" all'indirizzo: <https://www.nc3rs.org.uk/arrive-animal-research-reporting-vivo-experiments>.

2. Henderson VC, Kimmelman J, Fergusson D, Grimshaw JM, Hackam DG. Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for *in vivo* animal experiments. *PLoS Medicine* 2013;10(7):e1001489.
3. Cressey D. UK funders demand strong statistics for animal studies. *Nature* 2015;520(7547):271-2.
4. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 2013;14(5):365-76.
5. Fitts DA. Ethics and animal numbers: informal analyses, uncertain sample sizes, inefficient replications, and type I errors. *Journal of the American Association for Laboratory Animal Science* 2011;50(4):445-53.
6. Cohen J. *Statistical power analysis for the behavioral sciences*. 1980/10/01 ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
7. Festing MF. Extending the statistical analysis and graphical presentation of toxicity test results using standardized effect sizes. *Toxicologic Pathology* 2014;42(8):1238-49.
8. Chiarotti F. Statistical analysis of behavioral data. *Curr Protoc Toxicol* 2005;Chapter 13:Unit13.8.
9. Chiarotti F, Alleva E, Bignami G. Problems of test choice and data analysis in behavioral teratology: the case of prenatal benzodiazepines. *Neurotoxicology and Teratology* 1987;9(2):179-86.
10. Lazic SE, Essioux L. Improving basic and translational science by accounting for litter-to-litter variation in animal models. *BMC Neuroscience* 2013;14:37.
11. Zorrilla EP. Multiparous species present problems (and possibilities) to developmentalists. *Developmental Psychobiology*. 1997;30(2):141-50.
12. Faul F, Erdfelder E, Lang AG, Buchner A. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 2007;39(2):175-91.
13. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods* 2009;41(4):1149-60.
14. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, *et al.* Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PloS one* 2009;4(11):e7824.