



高雄醫學大學

KAOHSIUNG MEDICAL UNIVERSITY

未來高醫 引領未來

Integrity(誠信)、Internationalization(國際化)、Innovation(創新)
Intelligence(智慧)、Integration(整合)、Impact(影響力)

動物實驗規劃- 實驗動物與統計方法

Instructor: Ping-Ho Chen, Ph.D. (陳丙何)

Affiliation: School of Dentistry; Center for Medical Informatics and Statistics.
KMU

Tel: 07-312-1101 ext 2156-61

E-mail: phchen@kmu.edu.tw

2025/10/13

未來高醫 引領未來



高雄醫學大學

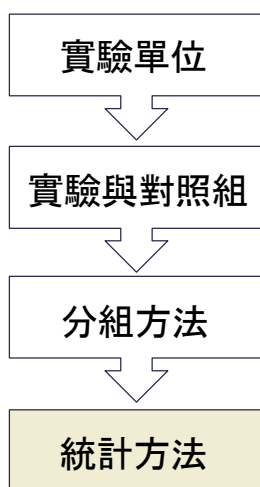
KAOHSIUNG MEDICAL UNIVERSITY

3Rs觀念之「減量Reduction」

- 動物分組方法、組數、每組使用動物數量等請依動物3Rs觀念之「減量Reduction」為考量重點，**說明選擇使用之動物數量的充分理由，包括估計達到顯著統計意義所需之最少數量時的依據。**
 - 引用參考文獻上的平均值(mean)與變異數(variance, σ^2)等數據，假設檢定之檢力(power of the test)，檢定水準(α level)
- 減量原則：在可得最大科學價值下，**採用最少傷害的方法與最少的動物隻數。**



實驗規劃- 實驗樣本與統計方法



- 前導試驗- 依過往經驗，探索實驗方法及數據態樣
 - Pilot Study 預試驗
 - Exploratory Study 探索型研究
- 正式試驗- 驗證假說，需要達到顯著差異
 - Confirmatory Study
 - 運用前導試驗結果來進行樣本數預估
- 重要考量因子
 - Type of comparison
 - Primary variables: continuous/ categorical
 - Errors: type I/ type II

Copyright © Genie Chin, NLAC NARLabs

<https://www.youtube.com/watch?v=A-QS0p9NCo4>

4

Factorial Experimental Design

- 快速篩選多個因子對實驗的影響 ($2^k, k=\text{factors}$)
- 可協助判斷那一個因子影響結果、判斷最適量、決定 sample size
- 一個資料多用途，資料最大化，減少動物使用(share N)

	Control	Drug
Male	 a	 c
Female	 b	 d

EU= mice, N=2

比較 - treatment 及 性別二項變因 (2^2)

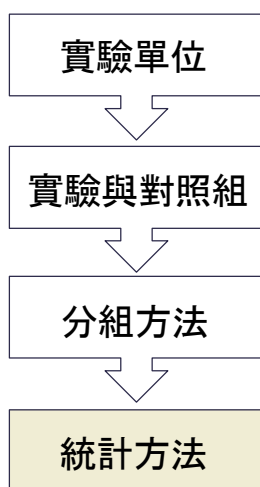
因此

1. 共有4個組別，有4個mean (a, b, c, d)
2. Treatment 差異: $(c+d)/2 - (a+b)/2$
3. 性別差異: $(a+c)/2 - (b+d)/2$
4. 性別對Treatment的影響: $((a-c)-(b-d))/2$

quantitative variable vs categorical variable
multi-way ANOVA vs ?

<https://www.youtube.com/watch?v=A-QS0p9NCo4>

實驗設計-實驗樣本與統計方法



樣本數計算

- 樣本數(N) 在整個實驗期間及報告撰寫時，都應該一致(若要排除應符合排除原則，並說明)
- 運用前導試驗結果來進行樣本數預估 (用那一項指標為基準做樣本數預估?)
- 減量原則：在可得最大科學價值下，採用最少傷害的方法與最少的動物隻數
(數量太多浪費多餘的動物生命，數量太少浪費全部的動物生命)
- 不適合的做法
 - 採用文獻上未經驗證的樣本數
 - 不解釋或硬套公式

<https://www.youtube.com/watch?v=A-QS0p9NCo4>

7



Sample size and power estimation

Instructor: Ping-Ho Chen, Ph.D. (陳丙何)
 Affiliation: School of Dentistry; Center for Medical Informatics and Statistics. KMU
 Tel: 07-312-1101 ext 2156-61
 E-mail: phchen@kmu.edu.tw

8



Sample size 計劃書撰寫

Power estimation 投稿論文

9

Sample-Size Estimation When Testing for the Mean of a Normal Distribution (Two-Sided Alternative)

- Suppose we wish to test $H_0: \mu = \mu_0$ vs. $H_1: \mu = \mu_1$ where the data are normally distributed with mean μ and known variance σ^2 . The sample size needed to conduct a one-sided test with significance level α and power $1 - \beta$ is

$$n = \frac{\sigma^2 (z_{1-\beta} + z_{1-\alpha/2})^2}{(\mu_0 - \mu_1)^2}$$

- Sample-Size Estimation Based on CI Width
- Suppose we wish to estimate the mean of a normal distribution with sample variance s^2 and require that the two-sided $100\% \times (1 - \alpha)$ CI for μ be no wider than L . The number of subjects needed is approximately

$$n = 4z_{1-\alpha/2}^2 s^2 / L^2$$

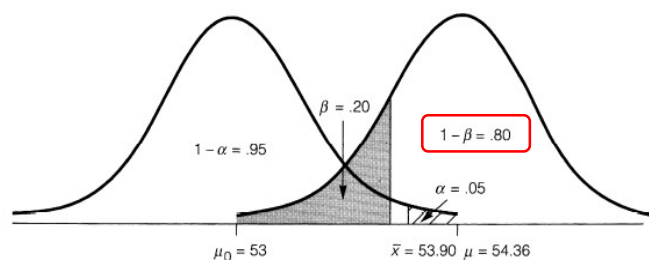
Cited from Bernard Rosner, Fundamentals of Biostatistics, 7th edition, 2011

10

檢定力

- $1 - \beta$: power (power of a test).
 - 檢力 or 檢定力。正確的拒絕虛無假設的機率。
 - $1 - \beta = P(\text{拒絕 } H_0 \mid H_1 \text{ 為真})$

圖 9.7 樣本平均數 $\mu_0 = 53$ 和 $\mu = 54.36$ 之分佈



Jan W.Kuzma 原著(史麗珠、林麗華編譯)，基礎生物統計學(Basic statistics for the health sciences)，學富文化事業有限公司，2008

- 如何設計高檢力(80%或90%)的統計檢定?
(β 很小：如0.2或0.1)
 - 可以放寬顯著水準(α)
 - $\alpha = 0.05 \rightarrow \alpha = 0.1$
 - α 越小，則 β 值越大
 - 要同時減少 α 及 β 錯誤，唯一的方法是減少重疊的部分
 - 增加樣本數

Jan W.Kuzma 原著(史麗珠、林麗華編譯)，基礎生物統計學(Basic statistics for the health sciences)，學富文化事業有限公司，2008



Summary

➤當考量統計檢定力 (Power)

- 樣本數會增加
- 提高實驗準確性
- 降低動物試驗的不準確性
- 避免動物使用浪費

13

Q: α 與 p 的差異?

- α (顯著水準(significant(α) level))
 - 研究者事先所設定的犯第一類型錯誤的顯著水準
 - 通常檢定水準就是允許犯型一誤差的最大機率值，即檢定水準為 α 最大值
 - 允許代表的檢定水準是自己定的(大部份都是訂 $\alpha=0.05$)
- p 值
 - 根據實得統計量去拒絕虛無假設(H_0)時(接受顯著性結果)，實際犯錯的機率
 - 統計考驗的實得顯著水準(observed significance level)
 - 根據樣本資料，所計算出來的值必須得要有樣本資料才能夠計算 p -value。所以不同的樣本資料，就會有不同的 p -value
 - p 值越小表示 H_0 越可能為錯，接受 H_1 的正確率越高。
 - p 值越大表示 H_1 越可能為錯，接受 H_1 的錯誤率越高。



14

單尾與雙尾檢定

Input Parameters		Output Parameters	
Test family	t tests	Statistical test	Means: Difference between two independent means (two groups)
Type of power analysis	Compromise: Compute implied α & power - given β/α ratio, sample size, and effect size		
Tail(s)	One	Noncentrality parameter &	2.5000000
Effect size d	One	Critical t	1.6589948
β/α ratio	4	Df	98
Sample size group 1	50	α err prob	0.0501570
Sample size group 2	50	β err prob	0.2006281
		Power (1 - β err prob)	0.7993719



- 無100%的信心去預期所得結果之方向，那您還是進行雙尾考驗
- 犧牲一點統計考驗力，進行雙尾的統計顯著性考驗是保守但客觀的做法。

為什麼要考慮耗損？

動物耗損與樣本數修正

- 在進行動物試驗時，**可能產生動物耗損或實驗失敗**。
- 若未提前考量，將導致**後續分析樣本不足**，影響統計結果。

範例說明：

- 經計算：每組需 **10 隻動物**。
- 已知此類研究 **耗損率約 20%**。
- 需將耗損量納入計算：

$$10 \div (1 - 0.2) = 12.5 \approx 13 \text{ 隻}$$

→ 因此，每組應使用 **13 隻動物**（含預期耗損）。



核心原則：樣本數的準確性來自四大要素

要素	說明
技術操作	熟練且穩定的操作可減少誤差。
飼養環境	良好且恆定的環境可降低外部變異。
實驗設計	優異設計能平衡組間差異與誤差。
實驗動物品質	高品質動物能穩定生理條件與反應。

結論

- 降低變異、穩定變異，是讓樣本估算**發揮應有效益**的關鍵。
- 綜合上述條件，方能產出**具統計意義與生醫價值**的研究成果。

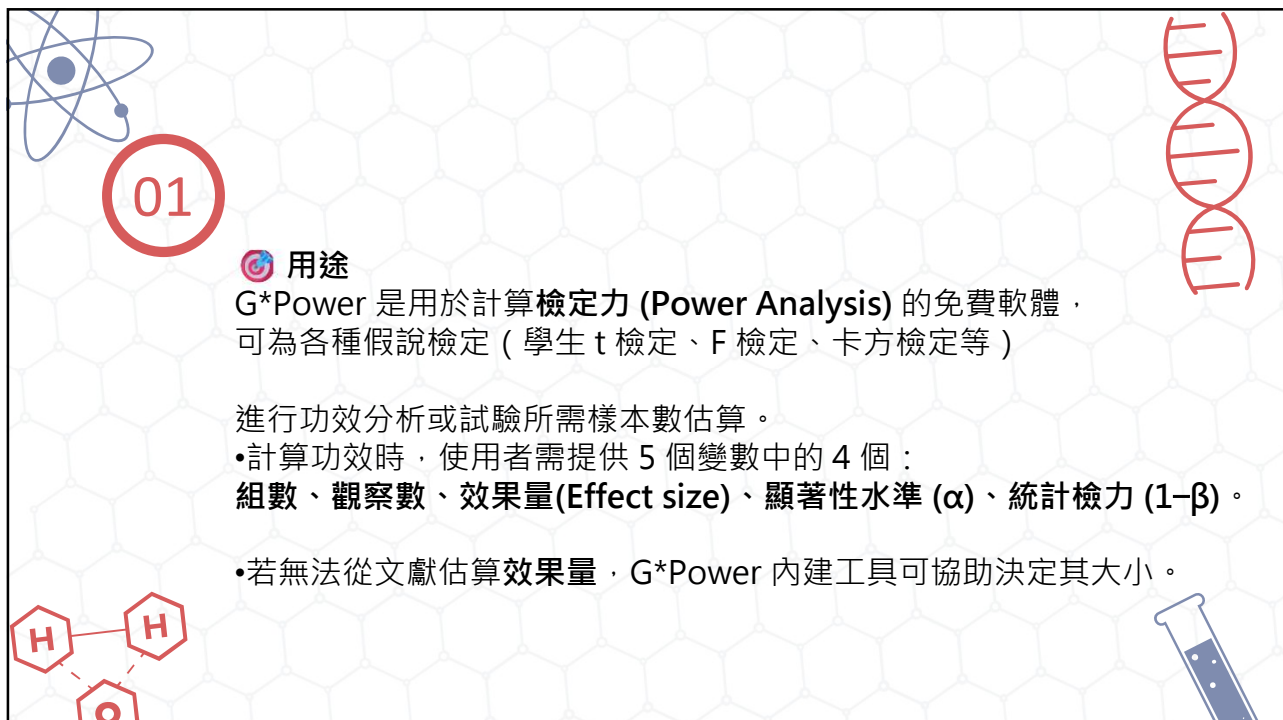




G-power軟體介紹

<https://www.psychologie.hhu.de/arbeitsgruppen/allgemein-e-psychologie-und-arbeitspsychologie/gpower>

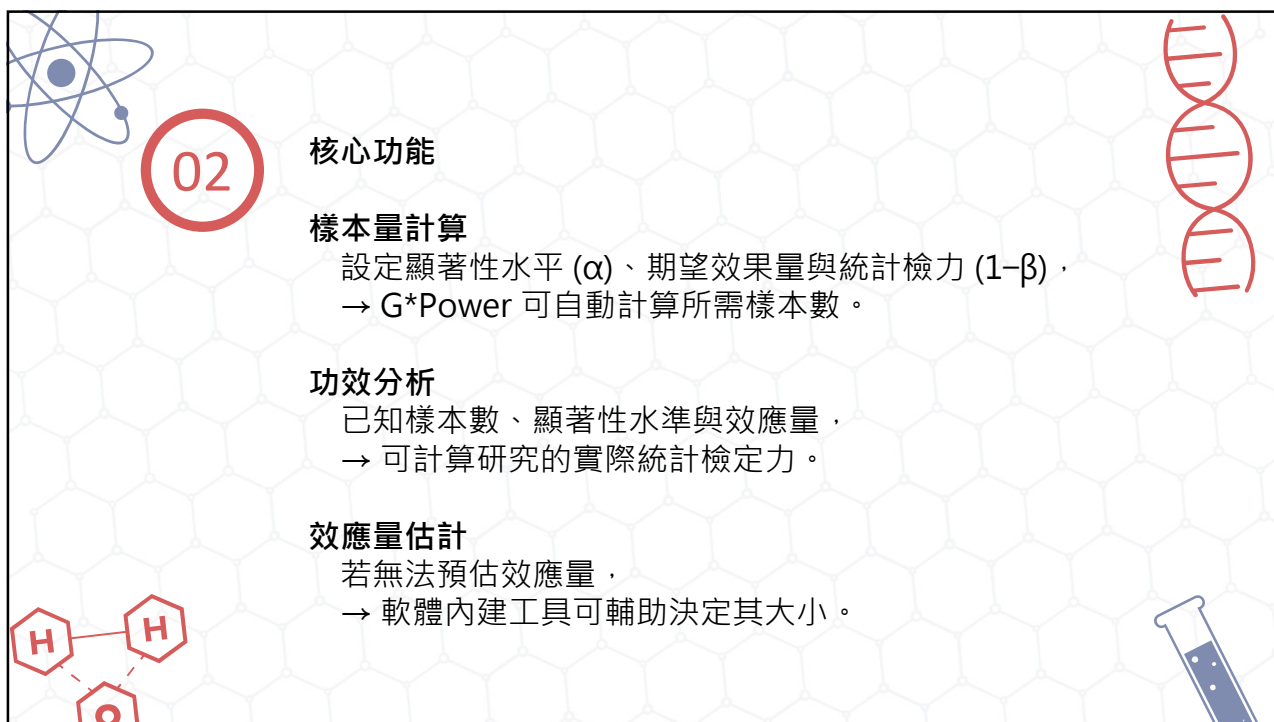
Download G*Power 3.1.9.7 for Windows XP, Vista, 7, 8, 10 and 11 (about 20 MB).



01 用途

G*Power 是用於計算**檢定力 (Power Analysis)** 的免費軟體，可為各種假說檢定（學生 t 檢定、F 檢定、卡方檢定等）進行功效分析或試驗所需樣本數估算。

- 計算功效時，使用者需提供 5 個變數中的 4 個：
組數、觀察數、效果量(Effect size)、顯著性水準 (α)、統計檢力 ($1-\beta$)。
- 若無法從文獻估算效果量，G*Power 內建工具可協助決定其大小。



02 核心功能

樣本量計算
設定顯著性水平 (α)、期望效果量與統計檢力 ($1-\beta$)，
→ G*Power 可自動計算所需樣本數。

功效分析
已知樣本數、顯著性水準與效應量，
→ 可計算研究的實際統計檢定力。

效應量估計
若無法預估效應量，
→ 軟體內建工具可輔助決定其大小。

樣本估計要做的事情（概覽）

- 
- 確認實驗單位 (experimental unit) (是「個體動物」還是「籠/窩/同胎」等)。這會直接決定 n 的定義。
 - 選一個主要結果 (primary outcome)，其它結果為次要。樣本數以主要結果為基準。
 - 決定統計參數：顯著水準 α (通常 0.05)、檢定力 (power, 通常 80% 或 90%)、試驗型別 (雙尾或單尾)、要偵測的最小具生物意義差異 (effect size) 與變異度 (SD)。
 - 用合適的統計模型／公式或軟體計算 n (t-test、ANOVA、比例、存活分析、重複量測等不同公式不同計算方式)。常用工具：G*Power、NC3Rs EDA、各大學/醫院的 power calculator。
 - 如果無法估 SD 或 effect size，考慮用 Resource Equation (Mead) 法作為探索性研究的替代方法 (把 $E = \text{error degrees of freedom}$ 控在 10–20)。
 - 把可能的動物流失率 (如死亡、樣本失效) 考慮進去並做上調 (例如預估 10–20% 流失則額外增加相應比例)。
 - 在倫理審查/IACUC/實驗記錄中寫清楚樣本數的估算過程、假設與參考資料 (依 ARRIVE/報告準則)。



03

支援的檢定類型

- ◆ 學生 t 檢定 (t-test)
- ◆ F 檢定 (ANOVA, Regression)
- ◆ 卡方檢定 (χ^2 test)
- ◆ 其他多種假說檢定方式



 **在實驗設計中的優點**

優點項目	說明
 免費且易於使用	降低研究成本，介面友好、操作直觀。
 確保研究的準確性與可靠性	避免樣本過少導致功效不足，或樣本過多造成資源浪費。
 資源規劃的優化	在設計初期即能確定合理樣本數，避免後期樣本不足。
 支援多樣實驗設計	多統計檢定類型，適用於不同研究領域需求。




未來高醫 引領未來





高雄醫學大學
KAOSHIUNG MEDICAL UNIVERSITY

常用公式（最常見情況）

- 兩組獨立樣本（等變異、連續結果）、雙尾 t 檢定的粗略 a priori 樣本數（每組 n）：

$$n = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma^2}{\Delta^2}$$

- $Z_{1-\alpha/2}$ ：雙尾 α 的常態臨界值（ $\alpha=0.05 \rightarrow 1.96$ ）
- $Z_{1-\beta}$ ：power 對應的常態值（power=0.8 $\rightarrow$ 0.84；power=0.9 $\rightarrow$ 1.28）
- σ ：兩組共同假設的標準差（SD）
- Δ ：兩組平均數之間要偵測的最小差（實務上要是「生物學上有意義」的差）
- （此式為近似公式；實務上可用軟體做更精確的 t 分布計算）。

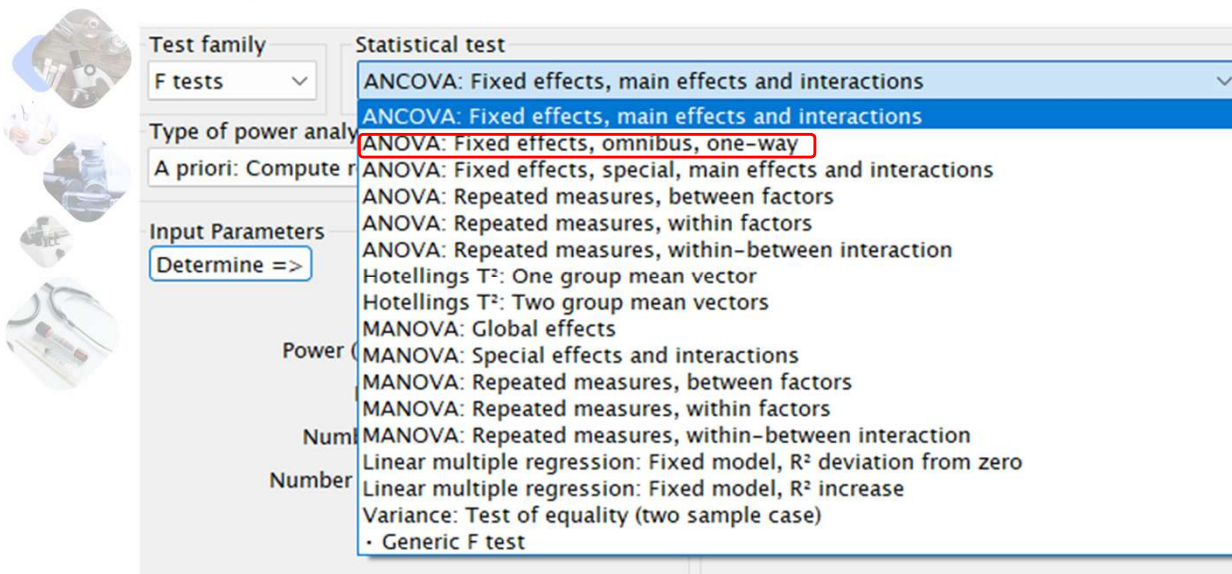
實例

- 情境：兩組獨立老鼠 (treatment vs control) · 期望偵測平均差 $\Delta=1.5$ $\Delta=1.5$ 單位；估 SD = 1.2； $\alpha=0.05$ (雙尾) · power = 80%。
- 取常態值： $Z_{1-\alpha/2}=1.96$ $Z_{1-\beta}=0.84$ 。
- Z 相加： $1.96+0.84=2.80$ 。
- 平方： $2.80^2=2.80 \times 2.80=7.84$ 。
- SD 平方： $\sigma^2=1.2^2=1.44$ 。
- 代入分子 (先算出 $(Z_{sum})^2 \times \sigma^2$ ： $7.84 \times 1.44=11.2896$ 。
- 乘以 2 (公式前面有 2)： $2 \times 11.2896=22.5792$ 。
- $\Delta^2=1.5^2=2.25$ 。
- 最後除以 Δ^2 ： $22.5792/2.25=10.0352$ 。
- 無條件進位取整數動物數 → 每組 **n = 11** (通常向上取整以確保樣本足夠)。
- 結論：在上述假設下，兩組各需 **11 隻** 老鼠 (總 N = 22)，若預估有 10% 流失率，則每組改成 $11/(1-0.10) \approx 12.211 / (1-0.10) \approx 12.2 \rightarrow$ **每組保守取 13 隻**。(計算與取整依實際倫理委員會與實驗限制可再微調)。



Test family	Statistical test
Exact	Correlation: Bivariate normal model
Type of power analysis	Correlation: Bivariate normal model
A priori: Compute	Linear multiple regression: Random model
Input Parameters	Proportion: Difference from constant (binomial test, one sample case)
Determine =>	Proportions: Inequality, two dependent groups (McNemar)
	Proportions: Inequality, two independent groups (Fisher's exact test)
	Proportions: Inequality, two independent groups (unconditional)
	Proportions: Inequality (offset), two independent groups (unconditional)
	Proportion: Sign test (binomial test)
	• Generic binomial test
α err prob	0.05
Power ($1-\beta$ err prob)	0.95
Correlation ρ H0	0
Total sample size	?
Actual power	?

未來高醫 引領未來



Test family: F tests

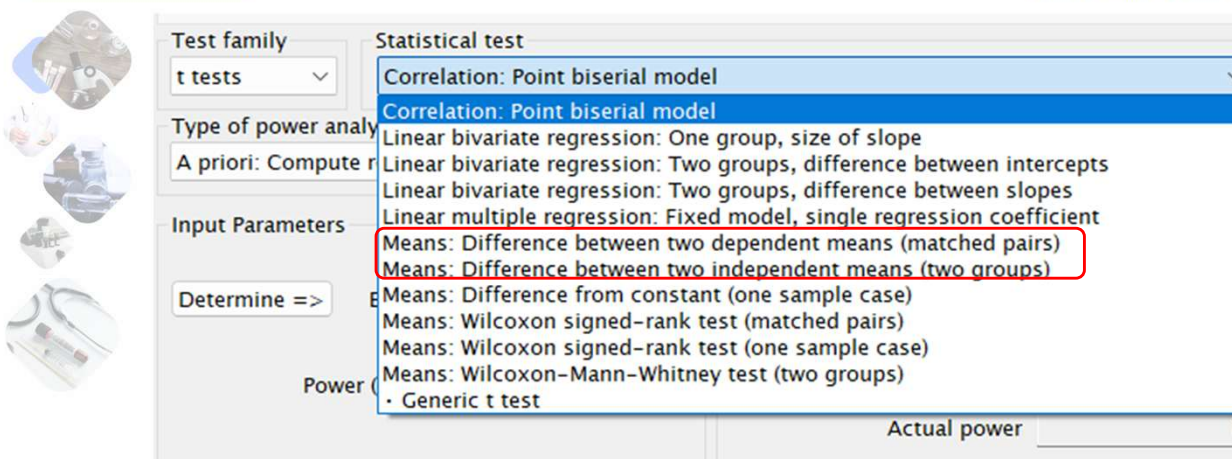
Type of power analysis: A priori: Compute power

Input Parameters: Determine =>

Statistical test:

- ANCOVA: Fixed effects, main effects and interactions
- ANOVA: Fixed effects, omnibus, one-way
- ANOVA: Fixed effects, special, main effects and interactions
- ANOVA: Repeated measures, between factors
- ANOVA: Repeated measures, within factors
- ANOVA: Repeated measures, within-between interaction
- Hottellings T²: One group mean vector
- Hottellings T²: Two group mean vectors
- MANOVA: Global effects
- MANOVA: Special effects and interactions
- MANOVA: Repeated measures, between factors
- MANOVA: Repeated measures, within factors
- MANOVA: Repeated measures, within-between interaction
- Linear multiple regression: Fixed model, R² deviation from zero
- Linear multiple regression: Fixed model, R² increase
- Variance: Test of equality (two sample case)
- Generic F test

未來高醫 引領未來



Test family: t tests

Type of power analysis: A priori: Compute power

Input Parameters: Determine =>

Statistical test:

- Correlation: Point biserial model
- Correlation: Point biserial model
- Linear bivariate regression: One group, size of slope
- Linear bivariate regression: Two groups, difference between intercepts
- Linear bivariate regression: Two groups, difference between slopes
- Linear multiple regression: Fixed model, single regression coefficient
- Means: Difference between two dependent means (matched pairs)
- Means: Difference between two independent means (two groups)
- Means: Difference from constant (one sample case)
- Means: Wilcoxon signed-rank test (matched pairs)
- Means: Wilcoxon signed-rank test (one sample case)
- Means: Wilcoxon-Mann-Whitney test (two groups)
- Generic t test

Actual power: ?

未來高醫 引領未來



Test family	Statistical test	
χ^2 tests	Goodness-of-fit tests: Contingency tables	
Type of power analysis	Goodness-of-fit tests: Contingency tables	
A priori: Compute r	Variance: Difference from constant (one sample case)	
	• Generic χ^2 test	
Input Parameters		
Determine =>	Effect size w	0.3
	α err prob	0.05
	Power (1- β err prob)	0.95
	Df	5
Output Parameters		
	Noncentrality parameter λ	?
	Critical χ^2	?
	Total sample size	?
	Actual power	?

未來高醫 引領未來



Test family	Statistical test	
z tests	Correlation: Tetrachoric model	
Type of power analysis	Correlation: Tetrachoric model	
A priori: Compute r	Correlations: Two dependent Pearson r's (common index)	
	Correlations: Two dependent Pearson r's (no common index)	
	Correlations: Two independent Pearson r's	
	Logistic regression	
	Poisson regression	
	Proportions: Difference between two independent proportions	
	• Generic z test	
Input Parameters		
Determine =>	α err prob	0.05
	Actual power	?

Effect of root-extracts of *Ficus benghalensis* (Banyan) in pain in animal models

Dipesh Raj Panday, Gajendra Prasad Rauniar

Department of Clinical Pharmacology and Therapeutics, B.P. Koirala Institute of Health Sciences, Dharan, Nepal

Panday DR, Rauniar GP. Effect of root-extracts of *Ficus benghalensis* (Banyan) in pain in animal models. J Neurosci Rural Pract 2016;7:210-5.

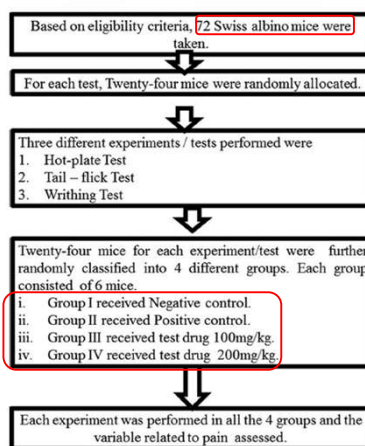


Figure 1: Study flow chart

- Sample size was calculated using free sample size calculating software G*Power version 3.1.9.2 (Franz, Universitat Kiel, Germany).
- With power of 80%, 0.05 level of statistical significance and effect size of 0.8, sample size for each test was calculated to be 24. Seventy-two
- mice were randomly assigned into one of the **three experimental groups**. It was a quantitative experimental study in mice.

Test family		Statistical test	
F tests		ANOVA: Fixed effects, omnibus, one-way	
Type of power analysis			
A priori: Compute required sample size – given α , power, and effect size			
Input Parameters			
Determine =>	Effect size f	0.8	
	α err prob	0.05	
	Power (1 – β err prob)	0.80	
	Number of groups	4	
Output Parameters			
	Noncentrality parameter λ	15.3600000	
	Critical F	3.0983912	
	Numerator df	3	
	Denominator df	20	
	Total sample size	24	
	Actual power	0.8592125	

G*POWER軟體 - Effect size估計介面

Input Parameters
Determine => Effect size f 0.8
α err prob 0.05
Power (1-β err prob) 0.80
Number of groups 4

Select procedure
Effect size from means

Number of groups 3
SD σ within each group 1

Group	Mean	Size
1	0	5
2	0	5
3	0	5

Equal n 5
Total sample size 15
Calculate Effect size f ?
Calculate and transfer to main window
Close

Select procedure
Effect size from variance

From variances
variance explained by special effect 1
Variance within groups 1
Partial η² 0.5
Calculate Effect size f ?
Calculate and transfer to main window
Close

Sample size calculations for select animal studies using G*Power.

Test family
t tests

Statistical test
Correlation: Point biserial model

Type of power analysis
A priori: Compute required sample size - given α, power, and effect size

☐ n1 != n2
Mean group 1 0
Mean group 2 1
SD σ within each group 0.5

☒ n1 = n2
Mean group 1 143.26
Mean group 2 192.84
SD σ group 1 54.50
SD σ group 2 48.90

Calculate Effect size d 0.9575908
Calculate and transfer to main window
Close

Test family
t tests

Statistical test
Means: Difference between two independent means (two groups)

Type of power analysis
A priori: Compute required sample size - given α, power, and effect size

Input Parameters
Determine => Effect size d 0.9575908
α err prob 0.05
Power (1-β err prob) 0.80
Allocation ratio N2/N1 1

Output Parameters
Noncentrality parameter δ 2.9514931
Critical t 2.0280940
Df 36
Sample size group 1 19
Sample size group 2 19
Total sample size 38
Actual power 0.8191075

無條件進位取整數動物數 → 每組 **n = 19**

結論：兩組各需 19 隻 老鼠（總 N = 38），若預估有 10% 流失率，則每組改成 $19/(1-0.10) \approx 21.11$ → 每組保守取 22 隻。（計算與取整依實際倫理委員會與實驗限制可再微調）。

How to calculate sample size in animal and human studies. Zhang X, Hartmann P. Front Med (Lausanne). 2023 Aug 17;10:1215927. doi: 10.3389/fmed.2023.1215927. eCollection 2023.

n1 = n2

Mean group 1	1.65
Mean group 2	3.37
SD σ group 1	0.85
SD σ group 2	1.59

Calculate Effect size d 1.349154

Calculate and transfer to main window Close

Test family: t tests Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size – given α , power, and effect size

Input Parameters		Output Parameters	
Tail(s)	Two	Noncentrality parameter δ	3.0168012
Determine =>	Effect size d: 1.3491545	Critical t	2.1009220
α err prob	0.05	Df	18
Power (1- β err prob)	0.80	Sample size group 1	10
Allocation ratio N2/N1	1	Sample size group 2	10
		Total sample size	20
		Actual power	0.8139797

X-Y plot for a range of values Calculate

How to calculate sample size in animal and human studies, Zhang X, Hartmann P. Front Med (Lausanne). 2023 Aug 17;10:1215927. doi: 10.3389/fmed.2023.1215927. eCollection 2023.

Test family: t tests Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size – given α , power, and effect size

Input Parameters		Output Parameters	
Tail(s)	Two	Noncentrality parameter δ	3.6948118
Determine =>	Effect size d: 1.3491545	Critical t	2.7632625
α err prob	0.01	Df	28
Power (1- β err prob)	0.80	Sample size group 1	15
Allocation ratio N2/N1	1	Sample size group 2	15
		Total sample size	30
		Actual power	0.8152957

判定標準更嚴格($p < 0.01$)・樣本數增加

Sample size calculations for select animal studies using G*Power. type I error of 0.01 and power of 0.8,

How to calculate sample size in animal and human studies, Zhang X, Hartmann P. Front Med (Lausanne). 2023 Aug 17;10:1215927. doi: 10.3389/fmed.2023.1215927. eCollection 2023.

統計檢力增加(0.95)，樣本數增加

Test family		Statistical test	
t tests		Means: Difference between two independent means (two groups)	
Type of power analysis			
A priori: Compute required sample size – given α , power, and effect size			
Input Parameters		Output Parameters	
Determine =>	Tail(s) Two	Noncentrality parameter δ	3.8159852
	Effect size d 1.3491545	Critical t	2.0422725
	α err prob 0.05	Df	30
	Power (1- β err prob) 0.95	Sample size group 1	16
	Allocation ratio N2/N1 1	Sample size group 2	16
		Total sample size	32
		Actual power	0.9582960

Sample size calculations for select animal studies using G*Power. type I error of 0.05 and power of 0.95.

How to calculate sample size in animal and human studies, Zhang X, Hartmann P. Front Med (Lausanne). 2023 Aug 17;10:1215927. doi: 10.3389/fmed.2023.1215927. eCollection 2023.

Sample size calculations (療效百分比差異) for select human studies using G*Power.

Test family		Statistical test	
z tests		Proportions: Difference between two independent proportions	
Type of power analysis			
A priori: Compute required sample size – given α , power, and effect size			
Input Parameters		Output Parameters	
	Tail(s) Two	Critical z	-1.9599640
Placebo	Proportion p2 0.1	Sample size group 1	11
Drug A	Proportion p1 0.65	Sample size group 2	11
	α err prob 0.05	Total sample size	22
	Power (1- β err prob) 0.80	Actual power	0.8039601
	Allocation ratio N2/N1 1		

How to calculate sample size in animal and human studies, Zhang X, Hartmann P. Front Med (Lausanne). 2023 Aug 17;10:1215927. doi: 10.3389/fmed.2023.1215927. eCollection 2023.

統計檢力增加(0.95)，樣本數增加

Test family		Statistical test	
z tests		Proportions: Difference between two independent proportions	
Type of power analysis			
A priori: Compute required sample size – given α , power, and effect size			
Input Parameters		Output Parameters	
Tail(s)	Two	Critical z	-1.9599640
Proportion p2	0.1	Sample size group 1	18
Proportion p1	0.65	Sample size group 2	18
α err prob	0.05	Total sample size	36
Power (1- β err prob)	0.95	Actual power	0.9607731
Allocation ratio N2/N1	1		

How to calculate sample size in animal and human studies. Zhang X, Hartmann P. Front Med (Lausanne). 2023 Aug 17;10:1215927. doi: 10.3389/fmed.2023.1215927. eCollection 2023.



Effect Size（效果量）

衡量研究中差異或關聯程度的重要統計指標

- ✓ **量化影響強度：**效果量提供一個客觀的數值來衡量變數間關聯的強度，或是比較不同處理組之間的差異大小，例如在社會科學和醫學研究中。
- ✓ **不受樣本數影響：**統計顯著性(p 值)很容易受樣本數影響，當樣本數很大時，即使效果很小也可能達到統計顯著。效果量則提供一個不受樣本數影響的指標，可以更真實地反映研究的實質意義。
- ✓ **補充統計檢定：**效果量與統計檢定是互補的概念，用於評估研究結果的實際重要性。

一、定義

- 效果量（Effect Size）是衡量研究中兩個變項之間實際差異或關聯程度的統計指標。
- 它補足了 p 值僅反映『是否有差異』的不足，讓研究者了解『差異有多大、是否有實質意義』。
-  簡單說：
- p 值 → 告訴你「差異是否存在」
- 效果量 → 告訴你「差異有多大」

二、為何重要

- 避免樣本數誤導：大樣本即使微小差異也可能顯著。
- 比較研究結果：效果量是跨研究比較與統合分析的基礎。
- 臨床與實務意義：提供統計顯著之外的實際重要性。

三、常見效果量指標

類型	指標	用途	概念/公式	Cohen基準
平均數差異	Cohen's d	比較兩組平均值	$d = (M_1 - M_2) / SD_{pooled}$	0.2小 0.5中 0.8大
變異解釋比例	η^2 / R^2	ANOVA或迴歸分析	解釋變異佔總變異比例	0.01小 0.06中 0.14大
相關程度	r	連續變項關聯	-1~1，越接近±1越強	0.1小 0.3中 0.5大
分類資料	OR / RR	群組事件機率比較	1=無差異；>1=增加；<1=減少	依研究領域而定

常見的標準化效果量指標與解釋

Cohen's d 與 η^2 效果量大小分類

Cohen's d

- 衡量兩組平均數差異的標準化值，常用於 t 檢定與 ANOVA
- 效果量分類：
 - $|d| \approx 0.2$ ：小效果
 - $|d| \approx 0.5$ ：中等效果
 - $|d| \approx 0.8$ ：大效果

Test family: t tests
Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size - given α , power, and effect size

Input Parameters: Tail(s): Two
Determine => Effect size d: 0.2
 α err prob: 0.05
Power (1 - β err prob): 0.8
Allocation ratio N2/N1: 1

Output Parameters: Noncentrality parameter δ : 2.8071338
Critical t: 1.9629867
Df: 786
Sample size group 1: 394
Sample size group 2: 394
Total sample size: 788
Actual power: 0.8005931

t tests: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size - given α , power, and effect size

Input Parameters: Tail(s): Two
Determine => Effect size d: 0.5
 α err prob: 0.05
Power (1 - β err prob): 0.8
Allocation ratio N2/N1: 1

Output Parameters: Noncentrality parameter δ : 2.8284271
Critical t: 1.9789706
Df: 126
Sample size group 1: 64
Sample size group 2: 64
Total sample size: 128
Actual power: 0.8014596

Test family: t tests
Statistical test: Means: Difference between two independent means (two groups)

Type of power analysis: A priori: Compute required sample size - given α , power, and effect size

Input Parameters: Tail(s): Two
Determine => Effect size d: 0.8
 α err prob: 0.05
Power (1 - β err prob): 0.8
Allocation ratio N2/N1: 1

Output Parameters: Noncentrality parameter δ : 2.8844410
Critical t: 2.0085591
Df: 50
Sample size group 1: 26
Sample size group 2: 26
Total sample size: 52
Actual power: 0.8074866

Cohen's f^2 / η^2 (eta-squared)

- 衡量變異數分析中自變數對總變異的解釋比例。
- η^2 越大，自變數效果越大。
- 效果量分類：
 - $\eta^2 = 0.04$ ：小效果
 - $\eta^2 = 0.25$ ：中等效果
 - $\eta^2 = 0.64$ ：大效果

例子說明

-  t檢定：實驗組=80、控制組=70、SD=10 $\rightarrow d=(80-70)/10=1.0 \rightarrow$ 大效果量
-  表示實驗組明顯優於控制組。
-  迴歸分析：模型 $R^2=0.25 \rightarrow$ 表示自變項可解釋25%變異，屬中等解釋力。

Summary

- 效果量應與 p 值並列報告，才能呈現完整研究結果
- 統計顯著不一定代表實質重要
- 建議在報告中同時討論效果量的理論與臨床意義

未來高醫 引領未來

高雄醫學大學
KAOSHIUNG MEDICAL UNIVERSITY



如何事先決定效果量？

- 最常使用的三種方法
- (1)根據之前的文獻
- (2)根據前驅研究 (Pilot study)
- (3)預估中等的效果量，
- 使用者只需保持G-power的預設即為中度效果量，效果量的中等效果=0.25，
 - 每個統計方法所採取的效果量都會有不一樣的公式，因此每一種效果算出來也會有不一樣的範圍及標準
 - 研究者在使用時，必須先確認所使用的效果量為哪一種，至於各效果量所對應的標準，可以參考

• 【Kottrik, J. W., & Williams, H. A. (2003). The incorporation of effect size in information technology, learning, and performance research. Information Technology, Learning, and Performance Journal, 21(1), 1-7.】

One-Way ANOVA G*Power

- is free for everyone
(<http://www.gpower.hhu.de/>)

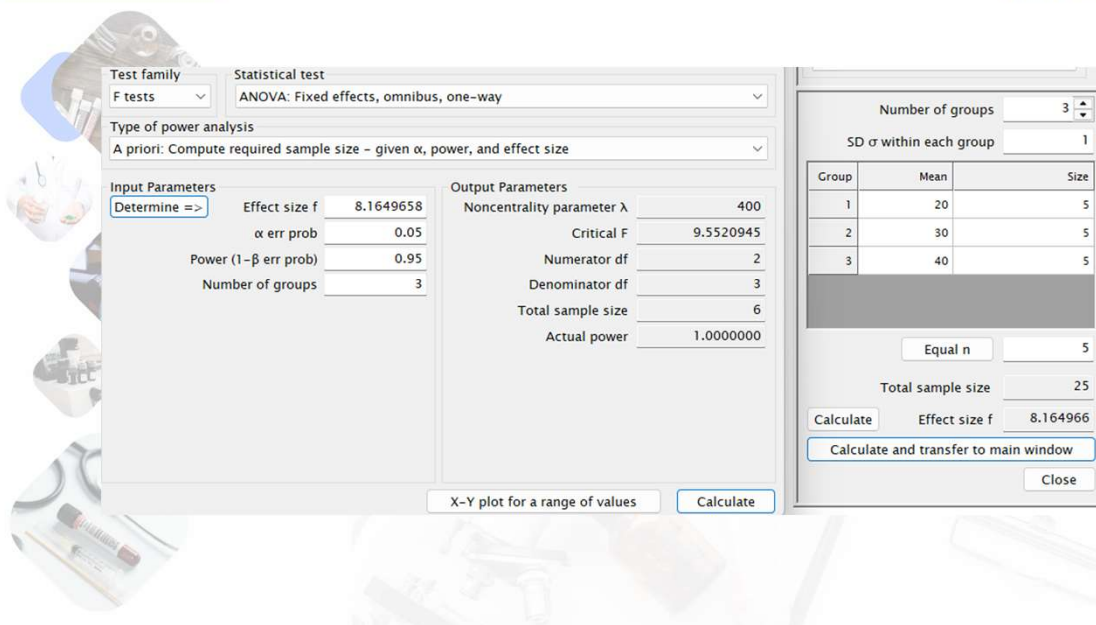


未來高醫 引領未來



- (1) Test family選擇「F tests」。
- (2) Statistical test選擇「ANOVA: Fixed effects, omnibus, one-way」。
- (3) 透過Determine，可將手邊已有的數據計算出效果量，若不需要的話。
- (4) 效果量的計算，G-power提供三種方式，主要分為從平均數得到效果量，以及從變異數得到效果量。
- (5) 第一種方式透過平均數計算效果量裡，需要先設定組別數，以及全體標準差（這部分也可以透過公式，根據各組的樣本數與標準差計算出）。
- (6) 接著填入每組平均數與樣本數，若是各組樣本數相同，可以透過Equal n做快速設定。
- (7) 按下「Calculate and transfer to main window」，可計算出效果量，並自動轉入左邊的效果量欄位。

未來高醫 引領未來



Test family: F tests
Statistical test: ANOVA: Fixed effects, omnibus, one-way

Type of power analysis: A priori: Compute required sample size – given α , power, and effect size

Input Parameters:
 Determine => Effect size f: 8.1649658
 α err prob: 0.05
 Power (1- β err prob): 0.95
 Number of groups: 3

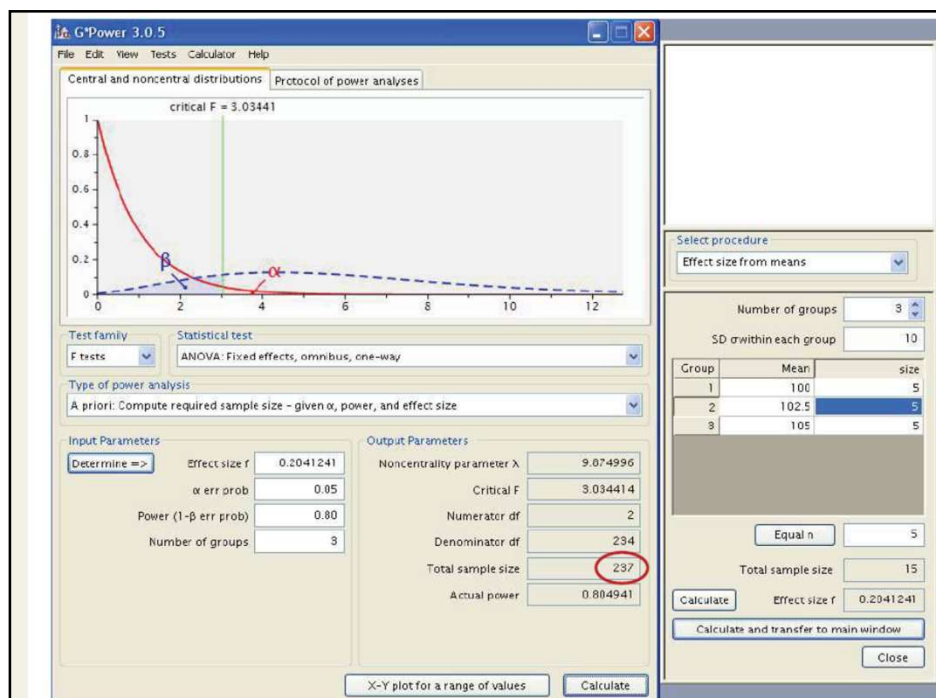
Output Parameters:
 Noncentrality parameter λ : 400
 Critical F: 9.5520945
 Numerator df: 2
 Denominator df: 3
 Total sample size: 6
 Actual power: 1.0000000

Number of groups: 3
 SD σ within each group: 1

Group	Mean	Size
1	20	5
2	30	5
3	40	5

Equal n: 5
 Total sample size: 25
 Calculate
 Effect size f: 8.164966
 Calculate and transfer to main window
 Close

X-Y plot for a range of values
 Calculate



G*Power 3.0.5

Central and noncentral distributions
 critical F = 3.03441

Test family: F tests
 Statistical test: ANOVA: Fixed effects, omnibus, one-way

Type of power analysis: A priori: Compute required sample size – given α , power, and effect size

Input Parameters:
 Determine => Effect size f: 0.2041241
 α err prob: 0.05
 Power (1- β err prob): 0.80
 Number of groups: 3

Output Parameters:
 Noncentrality parameter λ : 9.074996
 Critical F: 3.034414
 Numerator df: 2
 Denominator df: 234
 Total sample size: 237
 Actual power: 0.804941

Select procedure: Effect size from means

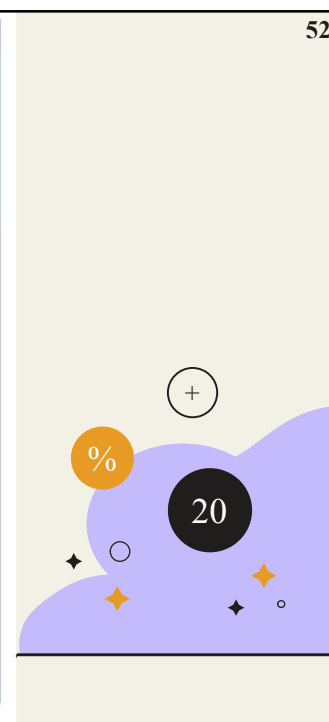
Number of groups: 3
 SD σ within each group: 10

Group	Mean	size
1	100	5
2	102.5	5
3	105	5

Equal n: 5
 Total sample size: 15
 Calculate
 Effect size f: 0.2041241
 Calculate and transfer to main window
 Close

X-Y plot for a range of values
 Calculate

52



The parameters examined included:

- Primary outcome variable: Patient-reported outcomes: VAS score on pain, swelling, bruising and bleeding.
- Secondary outcome variables: Clinical recordings of flap closure, pain, swelling, suppuration, and implant stability by calibrated examiners.

The sequence of events is presented in Fig. 1.

Statistical analysis

Sample size calculation and power analysis

For the sample size calculation, the power analysis was performed for a one-way fixed effects analysis of variance (ANOVA) with 4 levels. The criterion for significance was set at $\alpha = 0.05$ (type I error) and at $\beta = 0.20$ (type II error). The analysis of variance is nondirectional (i.e., two-tailed) indicating that an

effect in either direction will be interpreted. If an effect of 2.5 VAS units difference (largest mean – smallest mean) among the four treatment groups is expected, the sample size is 45 cases per group. However, if an effect of 2.0 VAS units is expected, the sample size is 70 cases per group. Calculating a dropout rate of 20%, $n = 84$ per group resulting in a power of 0.80 for an effect size of 2.0 VAS units at a level of significance of $\alpha = 0.05$.

Statistical and analytical methods

Chi-square (or Chi-square exact) tests were used to compare the percentage distribution of postsurgical complications at weeks 1, 2, 4, and 8 among the 4 treatment groups. Repeated measures ANOVA were performed on the VAS scores with the use of multivariate tests (Wilk's Lambda) for the effect of time and the interaction effect between treatment groups and time (all effects considered to be fixed). The above analyses were per-

Reference: Tan WC, Ong M, Han J, Mattheos N, Pjetursson BE, Tsai AY, Sanz I, Wong MC, Lang NP; ITI Antibiotic Study Group. Effect of systemic antibiotics on clinical and patient-reported outcomes of implant therapy - a multicenter randomized controlled clinical trial. Clin Oral Implants Res. 2014 Feb;25(2):185-93

line and SN line. We applied G*Power version 3.1.9.2 to estimate the sample size (Franz, Universitat Kiel, Germany) (1). With power ($1-\beta$) of 90%, α of 0.05, and the estimated effect size of 0.73, the total sample size calculated was only 30—which was 10 for each group. In this study, we recruited 120 patients to achieve sufficient power of at least 90%.

Cheng JH, Chen CM, Chen PH, Chou ST, Pan CY, Tseng YC. Comparison of Pharyngeal Airway between Mandibular Setback Surgery Patients (Skeletal Class III) and Nonsurgery Patients (Skeletal Classes I and II). Biomed Res Int. 2019 Mar 17;2019:5012037. doi: 10.1155/2019/5012037. eCollection 2019.

55

估算sample size and power using G*Power (Version 3.1.9.4)



56

- 1) 『(顯著性)』： α 通常直接設定為0.05
- 2) 『(統計檢定力 (power))』：Power通常直接設定為0.80 『(P(Y=1))』——結果/依變項 (Outcome) 的生機率，因此必須在文獻的敘述統計裡找
- 3) 到Y=1的百分比，或是母群體Y=1所佔的比例。E.g.發生或有疾病的人數比例（有疾病+無疾病=100%）。
- 4) 『OR值 (odds ratio)』——即X影響Y的效果量，通常一篇研究都會有一個最主要的研究因子(X)，因此需要針對此變項找到參考文獻所計算出來的OR值，通常會出現在卡方檢定或logistic regression部分。
- 6) 『自變項X的敘述性統計量』——雖然G-power裡提供X多種分配的設定，不過此處我們只舉兩個比較常見的範例：
 - (6-1)當X為連續變項時，我們必須在文獻的敘述統計裡找到X的平均數(mean)與標準差(standard deviation)，該軟體預設平均數=0、標準差=1
 - (6-22)當X為二分類(binominal variable)類別變項時，必須在文獻的敘述統計裡找到X=1的百分比，軟體預設0.5（即百分比為50%）。

57

$$OR = (151 \times 7) / (36 \times 63) = 0.4660493$$

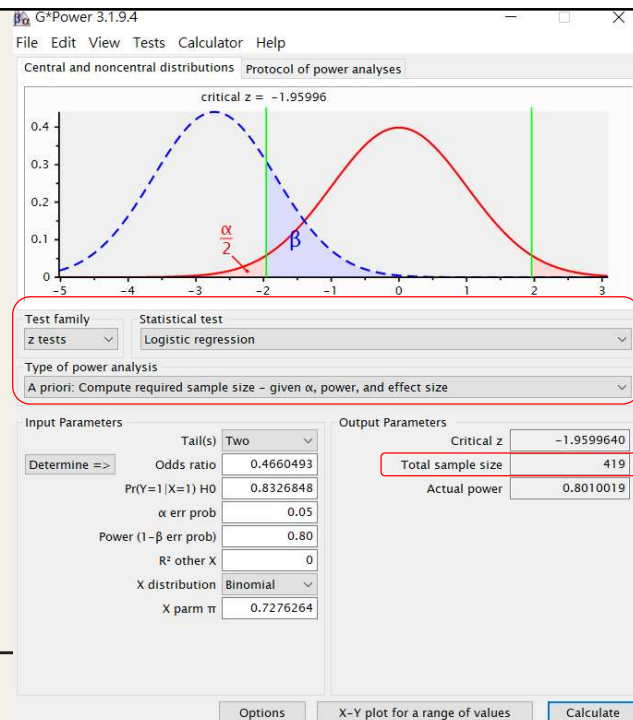
TABLE 3: Success rate of miniscrew implants at T2.

	Success rate at T2 (%)	Number of 2nd outcome successes/Total number	Unadjusted odds ratio (95% CI)	Unadjusted P value	Adjusted odds ratio (95% CI)	Adjusted P value
Overall success	83.3	214/257				
Age at surgery						
<20 years	80.7	151/187	0.47 (0.20 to 1.10)	0.082	0.49 (0.14 to 1.73)	0.270
≥20 years	90.0	63/70	1		1	
Gender				0.482		0.109
Female	81.4	92/113	1		1	
Male	84.7	122/144	1.27 (0.66 to 2.44)		2.12 (0.85 to 5.31)	
Sagittal skeletal malocclusion				0.025*		0.487
Class I	89.5	77/86	1		1	
Class II	83.5	101/121	0.59 (0.23 to 1.54)	0.438	0.88 (0.25 to 3.19)	0.852
Class III	68.6	24/35	0.26 (0.08 to 0.79)	0.014*	0.25 (0.02 to 2.56)	0.240
Vertical skeletal malocclusion				0.028*		0.043*
High angle	79.6	82/103	1		1	
Average	92.6	87/94	3.18 (1.13 to 8.98)	0.025*	4.22 (1.35 to 13.16)	0.013*
Low angle	75.0	12/16	0.77 (0.19 to 3.13)	1	2.20 (0.36 to 13.33)	0.391
Dental malocclusion				0.260		0.770
Class I	88.9	48/54	1		1	
Class II Div. I	84.0	126/150	0.69 (0.22 to 2.16)	1	0.81 (0.18 to 3.70)	0.787
Class III	75.0	30/40	0.39 (0.03 to 2.66)	1	1.25 (0.09 to 17.71)	0.339
Class II Div. 2	71.4	5/7	0.30 (0.10 to 1.47)	1	0.22 (0.01 to 5.02)	0.868

Reference: Yi Lin S, Mimi Y, Ming Tak C, Kelvin Weng Chiong F, Hung Chew W. A study of success rate of miniscrew implants as temporary anchorage devices in singapore. Int J Dent. 2015;2015:294670. doi: 10.1155/2015/294670. Epub 2015 Mar 10. Erratum in: Int J Dent. 2016;2016:5035379.

58

以研究主題為年齡的影響為例：
Recognizing the peak bone mass
(age 30) as a cutoff point to
achieve the success of
orthodontic implants



59

G*Power sample size(樣本大小)

Q1如何使用GPower計算相關係數所需要的樣本量?
以下以九步驟假設檢定的先行檢定力分析下去計算:

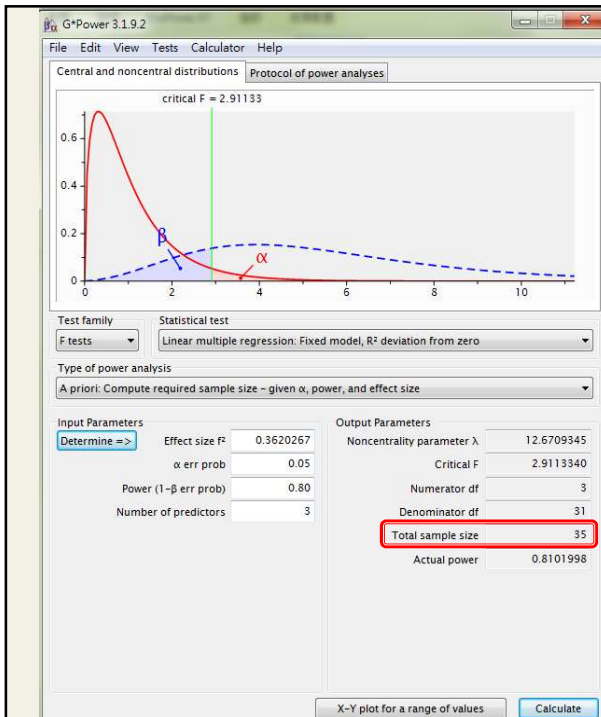
- (1)Test family: exact
- (2)Statistical test: Correlation: Bivariate normal model
- (3)Type of power analysis: A priori
- (4)Tails (建議使用雙尾)
- (5)Correlation H1 (此處指效果量)
- (6)alpha (顯著水準)
- (7)Power (檢定力)
- (8)Correlation H⁰ (虛無假設值 通常為零)
- (9)承上 在效果量為0.3 顯著水準0.05 檢定力0.80的條件下 研究樣本量(全部)至少為84

<http://mypaper.pchome.com.tw/readingstatistic/post/1321024507#47>

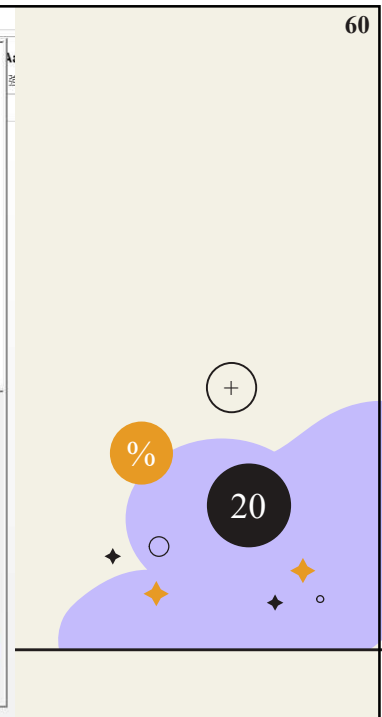
Test family	Statistical test
Exact	Correlation: Bivariate normal model
Type of power analysis	
A priori: Compute required sample size - given α , power, and effect size	

Test family	Statistical test
Exact	Correlation: Bivariate normal model
Type of power analysis	
A priori: Compute required sample size - given α , power, and effect size	
Input Parameters	
Tail(s): Two	
Determine =>	Correlation ρ H1: 0.3
α err prob	0.05
Power (1- β err prob)	0.80
Correlation ρ H0	0
Output Parameters	
Lower critical r	-0.2145669
Upper critical r	0.2145669
Total sample size	84
Actual power	0.8003390

60



From correlation coefficient	
Squared multiple correlation ρ^2	0.2658
From predictor correlations	
Number of predictors	3
Squared multiple correlation ρ^2	?
Specify matrices	
Calculate	Effect size F²: 0.3620267
Calculate and transfer to main window	
Close	





常見實務建議與注意事項

- 先做文獻/試驗室的資料搜尋：找到類似模型的平均值與 SD，或用 pilot data（但 pilot 太小時估 SD 可能偏差大）。
- 選擇「生物上有意義」的 **effect size**，而不只依統計顯著性設定極小差異。
- 多重比較：若多組比較或多個主要結果，須考慮 Type I error 的控制（如 Bonferroni 或事前定義主要比較）。多重比較會影響所需樣本數。
- 重複測量設計通常需要較少動物（因為每隻動物多次測量），但計算式較複雜且需考慮組內相關性（correlation）。
- 報告透明：在提交倫理審查或論文時，應提供樣本數估算的所有假設（ α 、power、 Δ 、SD、估算來源、軟體或公式）並引用 ARRIVE/ NICE 指南。
- 使用可靠工具：建議用 G*Power，NC3Rs 的 EDA power calculators 或等工具來做 a priori 計算；EDA 對動物研究特別友好並結合設計建議。

Parametric/non-parametric methods

- Parametric methods:
 - Underlying normal/binominal distribution
 - Underlying central-limit theorem ($N \geq 30$)
- Non-parametric methods:
 - Normal/binominal distribution inapplicable
 - Because of small sample size ($N < 30$)
 - central-limit theorem inapplicable

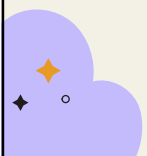
Central limit theorem

- The central limit theorem states that if sample size are large enough, the distribution will be approximately normal. The general rule of $n \geq 30$ applies.



Nonparametric methods

- Drawback
 - power is lost relative to using parametric procedure
- Comparison of two sample (analog to t test)
 - Independent samples
 - Wilcoxon rank-sum test (Mann-Whitney U test)
 - Paired samples
 - The sign test
 - Wilcoxon signed-rank test
- Comparison of k independent samples (analog to ANOVA)
 - Kruskal-Wallis test
- Correlation test
 - Spearman's correlation (analog to Pearson's correlation)



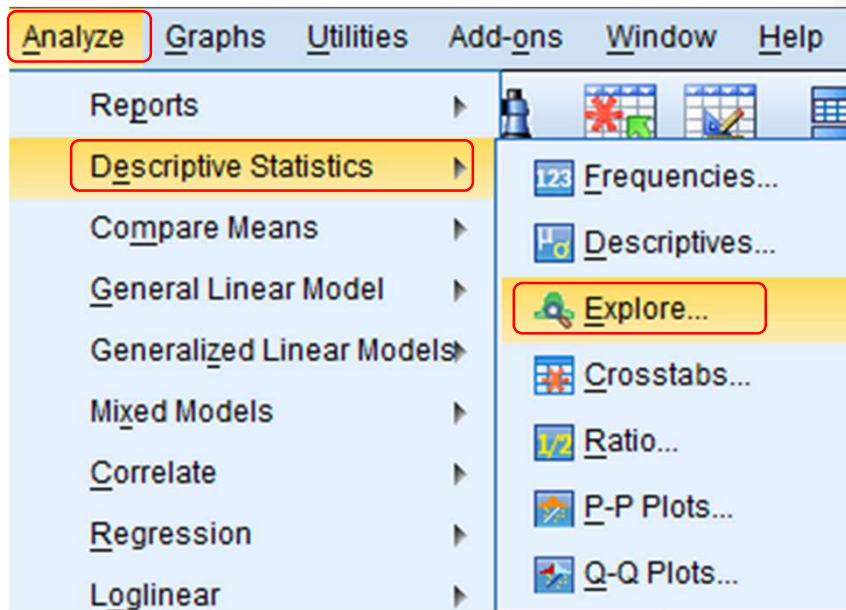
Statistical Methods to Test Hypotheses

資料屬性	兩群獨立樣本	三群以上獨立樣本	前後測比較	重複測量	迴歸分析
連續資料	Independent t test	ANOVA	Paired t test	重複量數分析	線性迴歸相關分析
類別資料	(1) Fisher's exact test (2) Yate's 校正數	Chi-square (χ^2) test	McNemar's test	Cochran's Q test	Logistic regression
序位資料 (連續無母數)	Mann-Whitney U test	Kruskal-Wallis test	Wilcoxon signed-rank test	Friedman statistic	
存活分析	Kaplan-Meier analysis Log rank test				Cox hazard regression

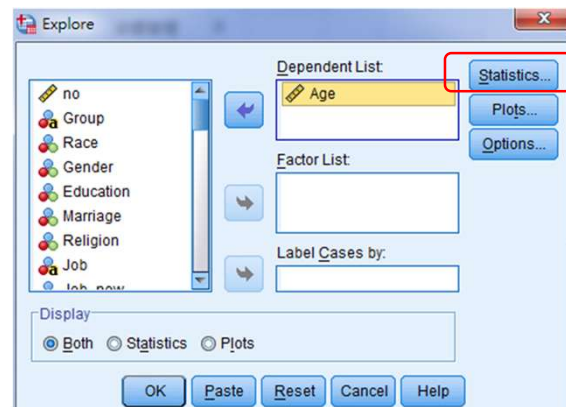
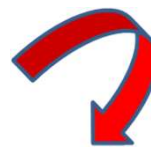
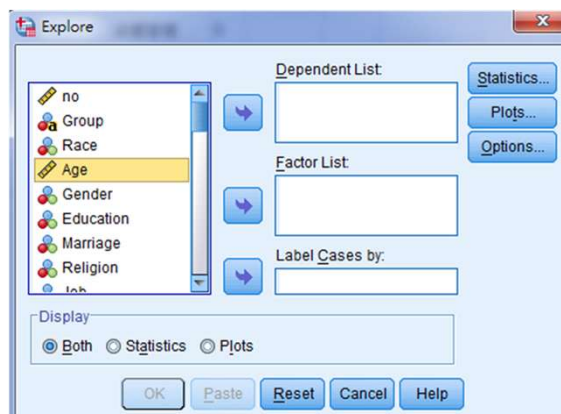
Normality Test



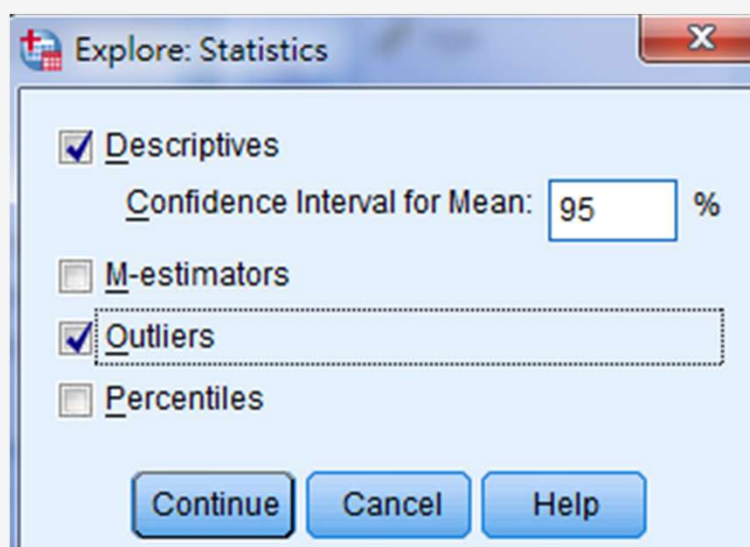
67



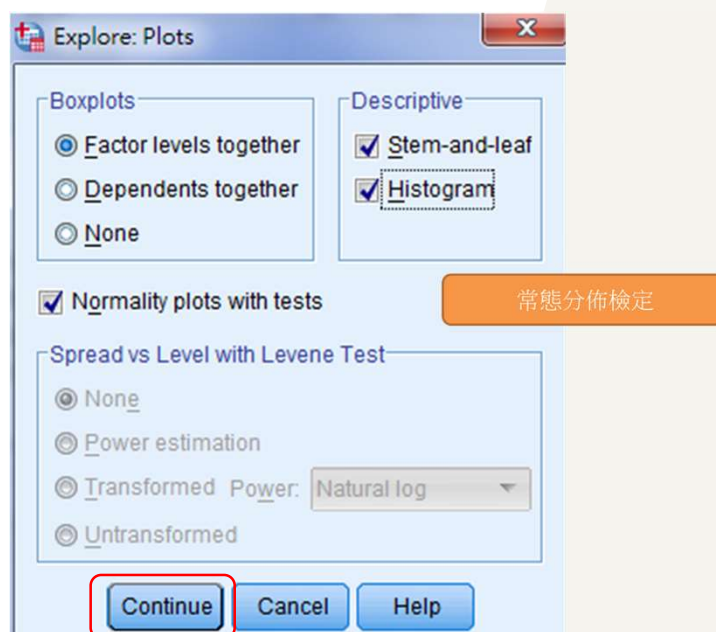
68



69



70



71

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Age	168	100.0%	0	.0%	168	100.0%

Descriptives

		Statistic	Std. Error
Age	Mean 平均數	51.07	.851
	95% Confidence Interval for Mean 平均數的95%信賴區間 (下限, 上限)	49.39	
	Lower Bound	52.74	
	Upper Bound		
	5% Trimmed Mean 刪除兩極端各5%觀察值之平均數	50.79	
	Median 中位數	50.50	
	Variance 變異數	121.559	
	Std. Deviation 標準差	11.025	
	Minimum 最小值	22	
	Maximum 最大值	89	
	Range 範圍	67	
	Interquartile Range 四分位全距	15	
	Skewness 偏態	.374	.187
	Kurtosis 峰度	.209	.373

72

Extreme Values

			Case Number	Value
Age	Highest	1	34	89
		2	111	78
		3	45	76
		4	65	76
		5	80	73 ^a
	Lowest	1	118	22
		2	147	30
		3	160	32
		4	158	33
		5	81	33 ^b

a. Only a partial list of cases with the value 73 are shown in the table of upper extremes.

b. Only a partial list of cases with the value 33 are shown in the table of lower extremes.

73

Test of Normality

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Age	.062	168	.200*	.989	168	.218

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

H_0 = Normality

If you accept, then assume normality

If you reject, then do not assume normality

- (1) $N \leq 2000$ elements, we use the Shapiro-Wilk test,
- (2) $N > 2000$, Kolmogorov-Smirnov test is used.
- (3) In our case, since we have only 168 cases, the Shapiro-Wilk test is used.
- (4) The p-value is 0.218, and we can reject the alternative hypothesis and conclude that the data comes from a normal distribution.

74

SPSS → Analyze → Descriptive Statistics → Explore → Plots

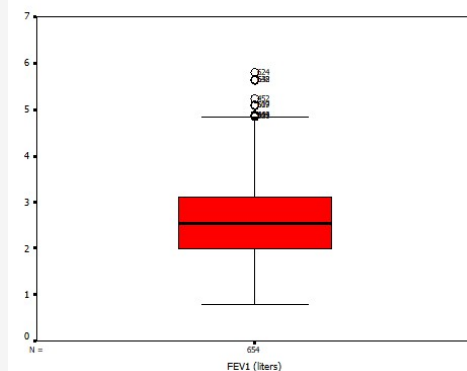
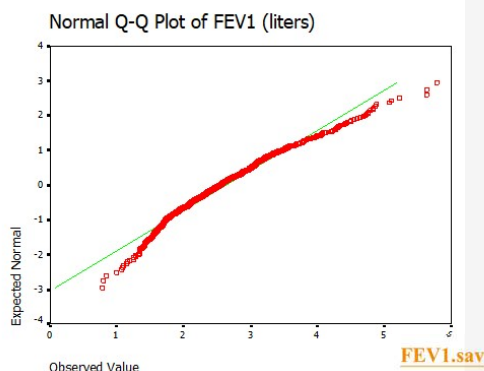
Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FEV1 (liters)	.051	654	.000

a. Lilliefors Significance Correction

Q: What can we do next?

1. Use Nonparametric
or
2. Transformed Data



75

SPSS → Analyze → Descriptive Statistics → Explore → Plots

Data Nature log
Transformation

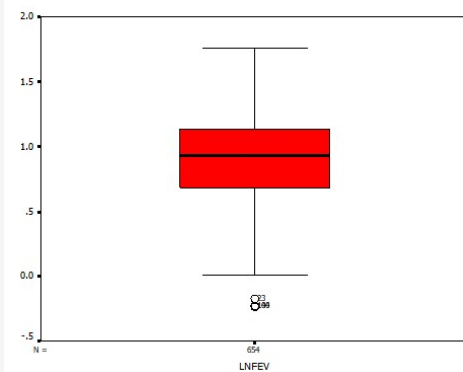
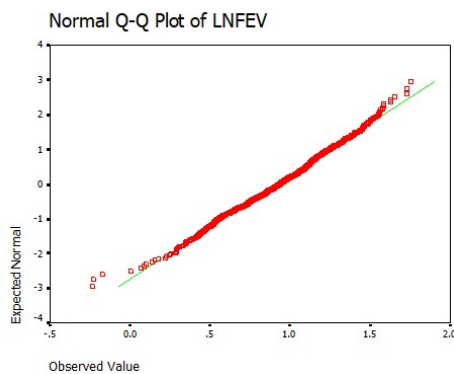
Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
LNFEV	.030	654	.200*

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Q: Can we use
Independent t-test
now?



常態性轉換

76

常態性轉換

非常態性模式	轉換方法
正偏態	取平方根。 $Y' = \sqrt{Y}$ 或 $X' = \sqrt{X}$
負偏態	取自然對數。 $Y' = \ln Y$ 或 $X' = \ln X$
低闊峰	取倒數。 $Y' = \frac{1}{Y}$ 或 $X' = \frac{1}{X}$
高峽峰	取平方。 $Y' = Y^2$ 或 $X' = X^2$

※也可利用增加樣本數來解決

「動物實驗樣本數試算網站」

2789-8700 iacuc@gate.sinica.edu.tw
中央研究院實驗動物照護及使用委員會

1



中央研究院
ACADEMIA SINICA

動物實驗線上申請電燈系統
ePROTOCOL
enabling paperless research compliance

version 2.0.16 U131.3

ePROTOCOL

Academia Sinica IACUC (中央研究院動物實驗管理小組) set this e-Protocol Management System to strengthen the laboratory animal research environment and enhance the protection of animal welfare involved in the research to meet the ethical standards and the regulatory research compliance requirements.

e-Protocol Management System is a total web based solution that automates the research approval processes. It is used for electronic protocol submission, protocol routing / reviewing / tracking that would enable investigators, IACUC members, and research administrators to process protocol applications online.....

登入前請務必先關閉快顯或彈出廣告攔截功能(或勾選"永遠允許快顯") 如何解除廣告攔截或快顯封鎖

Pop-up blocking software must be disabled (pop-ups must be allowed) while using eProtocol. See [instruction](#) for allowing pop-ups only for this site in your browser.

Tel: (02) 2789-8700 Email: iacuc@gate.sinica.edu.tw 歡迎計畫主持人來信申請帳號 (only PI can apply the account)

一般研究人員: 如何新增其他實驗操作人員 如何解除Avira Antivir 假警報 如何下載問卷 如何開發增補案(amendment)

如何回復審查意見 如何辦理展延(舊案提查) 動物實驗倫理3P說明樣本 如何新增實驗操作步驟

所長及動物房負責人專用: 線上查核使用說明

審查委員專用: 線上審查操作說明 增補案審查操作說明



User ID

Password

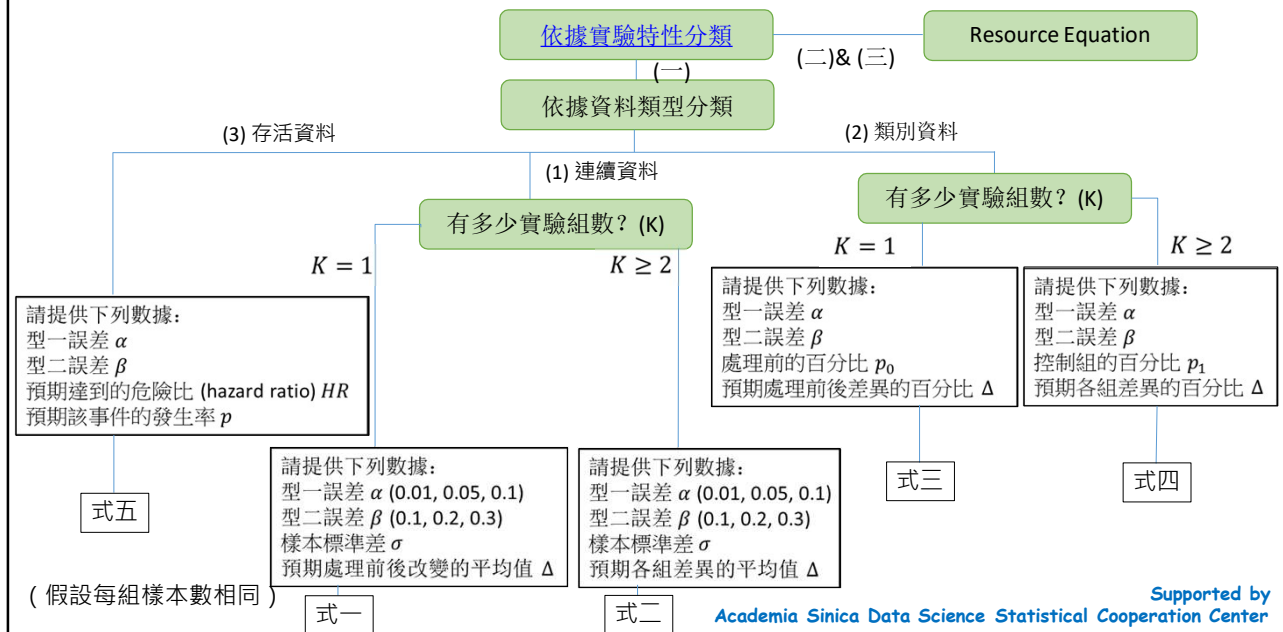
Login Forgot Password

實驗動物使用線上申報管理系統(Animal Use Monitoring System)

中央研究院實驗動物整合管理系統(建置中, 僅供委員、管理人員使用)

動物實驗樣本數試算網頁  <http://mouse.imb.sinica.edu.tw/IACUCEP/samplesizecal.action>

小鼠樣本數計算公式



中央研究院實驗動物整合管理系統
TEL : (02)2789-8700
E-mail: iacuc@gate.sinica.edu.tw

動物實驗樣本數 (N值) 試算

◎ 請填寫您要規劃動物使用量的Protocol編號:

例如: 20-08-1234

下一步驟

Powered by 中央研究院 統計所 資料科學統計合作社

請注意: 系統將根據 Protocol 編號產生試算結果 (文字+PDF), 請將結果複製至 Protocol 內容中或列為附件, 方便加速審查作業。

動物實驗樣本數（N值）試算

◎ 請選擇實驗的特性：

一般實驗：

- (一) 此實驗所收集的數據要進行顯著性分析(例如：t-test, ANOVA, Chi-square test, log rank test)，且可提供計算樣本數的重要數據。(例如：預期的差異，標準差，hazard ratio等)

僅限先導試驗(Pilot Study only)：

- (二) 此實驗所收集的數據要進行顯著性分析(例如：t-test, ANOVA, Chi-square test, log rank test)，但無法提供計算樣本數的重要數據。
- (三) 此實驗所收集的數據暫不進行顯著性分析。

下一步驟

請注意：(二)(三)僅限於少量數量的先導試驗(Pilot Study only)，若使用於一般試驗，Protocol審查時將不予通過。

返回第一步驟 返回上一步驟 Protocol: 20-08-1234 → 一般實驗的顯著性分析

動物實驗樣本數（N值）試算

◎ 請選擇實驗的資料類型：

- 連續資料 (例如：動物體重、腫瘤大小、進食量、血壓、血糖等)。
- 類別資料 (例如：有無發生、有無發病、是否判定死亡等)。
- 存活資料 (例如：自起始點到事件發生或是觀察終了的時間)。

下一步驟

說明：

1. 類別資料也可以是基於連續資料產生的判定結果(例如：血糖降到某一特定數值或腫瘤直徑縮小至特定值時，即可判定有效)。
2. 存活資料一般多用在流行病學的研究。



中央研究院
ACADEMIA SINICA

中央研究院實驗動物整合管理系統
TEL : (02)2789-8700
E-mail: iacuc@gate.sinica.edu.tw

返回第一步驟
返回上一步驟
Protocol: 20-08-1234 → 一般實驗的顯著性分析 → 連續資料

動物實驗樣本數 (N值) 試算

◎ 請問有多少的實驗組數(K) :

 組。 (K=1或K≥2)之整數

下一步驟

說明：

- K=1係指實驗沒有對照組，是測量比較同組動物處理前/後、手術前/後、用藥前/後的生理數值資料 (例如：注射特定藥物前後的血糖值、腫瘤直徑(體積)等)。
- K≥2則表示實驗有對照組，是測量比較實驗組與對照組間之差異。

注意：K≥2時，組數的多寡並不會影響樣本數 (N值) 的大小

自己跟自己比



中央研究院
ACADEMIA SINICA

中央研究院實驗動物整合管理系統
TEL : (02)2789-8700
E-mail: iacuc@gate.sinica.edu.tw

返回第一步驟
返回上一步驟
Protocol: 20-08-1234 → 一般實驗的顯著性分析 → 連續資料 → 1組

動物實驗樣本數 (N值) 試算

◎ 請提供下列數據：

型一誤差 α

一般建議選用 0.05

型二誤差 β

一般建議選用 0.2

標準差 σ

範圍與限制：{x≥0}

預期處理前後改變的平均值 Δ

範圍與限制：{x≠0}

下一步驟

說明：

- 型一誤差 α ：即所謂的「顯著水準」或「偽陽性的出現機率」，一般建議選用 0.05。
- 型二誤差 β ：(1- β) 即所謂「檢定力」，一般建議選用 0.2，即當檢定力=0.8時。
- 標準差 σ 即變異數 σ^2 之平方根，指測定的數值在該特定族群或品系的變異或離散程度，通常可由前導試驗、文獻或已知之類似藥物資料甚至經驗來得到。
- 預期(定)實驗能達到測定的數值在處理前 vs 處理後之改變(差異)的平均值，例如：某藥物預期可以讓糖尿病模式鼠的血糖值平均下降40mg/dL。

注意：標準差 σ 與預期改變的平均值所使用的「單位」要相同

返回第一步驟 返回上一步驟 Protocol 20-08-1234 → 一般實驗的顯著性分析 → 連續資料 → 2組

動物實驗樣本數 (N值) 試算

注意：K≥2時，組數的多寡並不會影響樣本數 (N值) 的大小

◎ 請提供下列數據：

型一誤差 α

型二誤差 β

標準差 σ 範圍與限制： $\{x \geq 0\}$

預期各組差異的平均值 Δ 範圍與限制： $\{x \neq 0\}$

下一步驟

說明：

1. 型一誤差 α ：即所謂的「顯著水準」或「偽陽性的出現機率」，一般建議選用 0.05。
2. 型二誤差 β ： $(1-\beta)$ 即所謂「檢定力」，一般建議選用 0.2，即當檢定力=0.8時。
3. 標準差 σ 即變異數 σ^2 之平方根，指測定的數值在該特定族群或品系的變異或離散程度，通常可由前導試驗、文獻或已知之類似藥物資料甚至經驗來得到。
4. 預期(定)實驗能達到測定的數值在 實驗組 vs 對照組 間差異的平均值，例如：某藥物預期可以讓糖尿病模式鼠的血糖值比對照組平均下降40mg/dL。

注意：標準差 σ 與 預期改變的平均值所使用的「單位」要相同

藉由測量大鼠體重以檢測某藥物是否影響食慾

決定 α 值 → 5% → **0.05**
 決定 β 值 → 20% → **0.2**
 先前實驗得知大鼠平均體重為400g, 標準差 σ 為23g

研究人員設定“預期的差異值達20g”即該藥物影響食慾

連續資料・無對照組(K=1)

因族群內個體差異大，故針對單一個體測量實驗前後的差異 (自己跟自己比)

- 實驗組數 K = 1
- 雙尾顯著水準 $\alpha = 0.05$
- 檢定力 $1-\beta = 1-0.2 = 0.8$
- 預期達到目標的差異 $\Delta = 20$
- 標準差 $\sigma = 23$

依據公式

$$n = \frac{\sigma^2(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}})^2}{\Delta^2}$$

$n = (\sigma^2)(Z_{1-\beta} + Z_{1-\alpha/2})^2 / \Delta^2$
 所得各組最少的樣本數為 **11**。

連續資料・有對照組(K≥2)

- 實驗組數 K = 2
- 雙尾顯著水準 $\alpha = 0.05$
- 檢定力 $1-\beta = 1-0.2 = 0.8$
- 預期達到目標的差異 $\Delta = 20$
- 標準差 $\sigma = 23$

依據公式

$$n = \frac{2\sigma^2(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}})^2}{\Delta^2}$$

$n = (2\sigma^2)(Z_{1-\beta} + Z_{1-\alpha/2})^2 / \Delta^2$
 所得各組最少的樣本數為 **21**。

樣本數減少

中央研究院
ACADEMIA SINICA

中央研究院實驗動物整合管理系統
TEL : (02)2789-8700
E-mail: iacuc@gate.sinica.edu.tw

返回第一步驟 返回上一步驟 Protocol: 20-08-1234 → 一般實驗的顯著性分析 → 連續資料 → 2組

動物實驗樣本數 (N值) 試算

◎ 計算結果如下：

▼ 請點此下載試算結果的PDF檔。

複製以下的文字內容至剪貼簿 將以下的文字內容複製到其他的應用程式文字輸入框。

Protocol #: 20-08-1234
試算ID : 2_20-08-1234_1596526157147
適用公式 #: 2
樣本數試算的參數條件：

- 實驗組數 $K = 2$
- 雙尾顯著水準 $\alpha = 0.05$
- 檢定力 $1 - \beta = 1 - 0.2 = 0.8$
- 預期達到目標的差異 $\Delta = 20$
- 標準差 $\sigma = 23$

依據公式

中央研究院
ACADEMIA SINICA

中央研究院實驗動物整合管理系統
TEL : (02)2789-8700
E-mail: iacuc@gate.sinica.edu.tw

返回第一步驟 返回上一步驟 Protocol: 20-08-1234 → 一般實驗的顯著性分析

動物實驗樣本數 (N值) 試算

◎ 請選擇實驗的資料類型：

- 連續資料 (例如：動物體重、腫瘤大小、進食量、血壓、血糖等)。
- 類別資料 (例如：有無發生、有無發病、是否判定死亡等)。
- 存活資料 (例如：自起始點到事件發生或是觀察終了的時間)。

下一步驟

說明：

- 類別資料也可以是基於連續資料產生的判定結果 (例如：血糖降到某一特定數值或腫瘤直徑縮小至特定值時，即可判定有效)。
- 存活資料一般多用在流行病學的研究。

中央研究院實驗動物整合管理系統

TEL : (02)2789-8700

E-mail: iacuc@gate.sinica.edu.tw

中央研究院
ACADEMIA SINICA

Protocol: 20-08-1234 → 一般實驗的顯著性分析 → 類別資料

動物實驗樣本數 (N值) 試算

◎ 請問有多少的實驗組數(K) :

組。 (K=1或 $K \geq 2$)之整數

說明：

1. $K=1$ 係指實驗沒有對照組，是測量比較同組動物處理前/後、手術前/後、用藥前/後的生理數值資料 (例如：注射特定藥物前後的血糖值、腫瘤直徑(體積)等)。
2. $K \geq 2$ 則表示實驗有對照組，是測量比較實驗組與對照組間之差異。

注意： $K \geq 2$ 時，組數的多寡並不會影響樣本數 (N值) 的大小

Protocol: 20-08-1234 → 一般實驗的顯著性分析 → 類別資料 → 1組

中央研究院
ACADEMIA SINICA

Protocol: 20-08-1234 → 一般實驗的顯著性分析 → 類別資料

動物實驗樣本數 (N值) 試算

◎ 請提供下列數據：

型一誤差 α

一般建議選用 0.05

型二誤差 β

一般建議選用 0.2

處理前的百分比 p_0 (%)

範圍與限制：{0≤x≤100}

預期處理前後差異的百分比 Δ (%)

範圍與限制：{-100≤x≤100}

說明：

1. 型一誤差 α ：即所謂的「顯著水準」或「偽陽性的出現機率」，一般建議選用 0.05。
2. 型二誤差 β ：(1- β) 即所謂「檢定力」，一般建議選用 0.2，即當檢定力=0.8時。
3. 處理前的百分比 p_0 ，指處理前類別事件的發生率，通常可由前導試驗、文獻或已知之類似藥物資料甚至經驗來得到。
4. 差異的百分比 Δ ，即預期(定)實驗能達到事件的發生率改變(差異)的平均值，增加為正值、減少為負值。

注意：請輸入百分比，例如：30%即輸入30，百分比的差異為5 %即輸入5，- 5 %即輸入-5

Approach 1：一定時間內實驗組是否發生該事件(是與否，二分變量)

不需要知道standard deviation

必須知道對照組發生的比率並且預估實驗組的發生比率

類別資料

某基因轉殖鼠自發癌症發生率為50%，預估測試藥物能降低發生率到25%

決定 α 值 $\rightarrow$ 5% $\rightarrow$ 0.05

決定 β 值 $\rightarrow$ 20% $\rightarrow$ 0.2

類別資料・無對照組(K=1)

- 實驗組數 $K = 1$
- 雙尾顯著水準 $\alpha = 0.05$
- 檢定力 $1 - \beta = 1 - 0.2 = 0.8$
- 處理前的百分比 $p_0 = 50\%$
- 預期達到目標的差異百分比 $\Delta = 25\%$

依據公式

$$n = \frac{p_0 q_0 (Z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} + Z_{1-\frac{\alpha}{2}})^2}{\Delta^2}, p_1 = p_0 + \Delta$$

$n = p_0 q_0 (Z_{1-\beta} \sqrt{p_1 q_1 / p_0 q_0} + Z_{1-\alpha/2})^2 / \Delta^2$,
 $p_1 = p_0 + \Delta$

所得各組最少的樣本數為 29。

← 樣本數減少

類別資料・有對照組(K $\geq$ 2)

- 實驗組數 $K = 2$
- 雙尾顯著水準 $\alpha = 0.05$
- 檢定力 $1 - \beta = 1 - 0.2 = 0.8$
- 控制組的百分比 $p_1 = 50\%$
- 預期達到目標的差異百分比 $\Delta = 25\%$

依據公式

$$n = \frac{(Z_{1-\beta} \sqrt{p_1 q_1 + p_2 q_2} + Z_{1-\frac{\alpha}{2}} \sqrt{2 p \bar{q}})^2}{\Delta^2}$$

$$p_2 = p_1 + \Delta, \bar{p} = \frac{p_1 + p_2}{2}, \bar{q} = 1 - \bar{p}$$

$n = (Z_{1-\beta} \sqrt{p_1 q_1 + p_2 q_2} + Z_{1-\alpha/2} \sqrt{2 * p_avg * q_avg})^2 / \Delta^2$,
 $p_2 = p_1 + \Delta, p_avg = (p_1 + p_2) / 2, q_avg = 1 - p_avg$

所得各組最少的樣本數為 58。


中央研究院
 ACADEMIA SINICA

中央研究院實驗動物整合管理系統
 TEL : (02)2789-8700
 E-mail: iacuc@gate.sinica.edu.tw

返回第一步驟
返回上一步驟
Protocol: 20-08-1234 → 一般實驗的顯著性分析

動物實驗樣本數 (N值) 試算

◎ 請選擇實驗的資料類型：

☐ 連續資料 (例如：動物體重、腫瘤大小、進食量、血壓、血糖等)。

☒ 類別資料 (例如：有無發生、有無發病、是否判定死亡等)。

☐ 存活資料 (例如：自起始點到事件發生或是觀察終了的時間)。

說明：

- 類別資料也可以是基於連續資料產生的判定結果 (例如：血糖降到某一特定數值或腫瘤直徑縮小至特定值時，即可判定有效)。
- 存活資料一般多用在流行病學的研究。

[返回第一步驟](#)
[返回上一步驟](#)
[Protocol:20-08-1234](#)
[→ 一般實驗的顯著性分析](#)
[→ 存活資料](#)

動物實驗樣本數（N值）試算

◎ 請提供下列數據：

型一誤差 α 一般建議選用 0.05
 型二誤差 β 一般建議選用 0.2
 預期達到的危險比 (hazard ratio) HR 範圍與限制：{ $x > 0$ }
 預期該事件的發生率 p (%) 範圍與限制：{ $0 < x \leq 100$ }

[下一步驟](#)

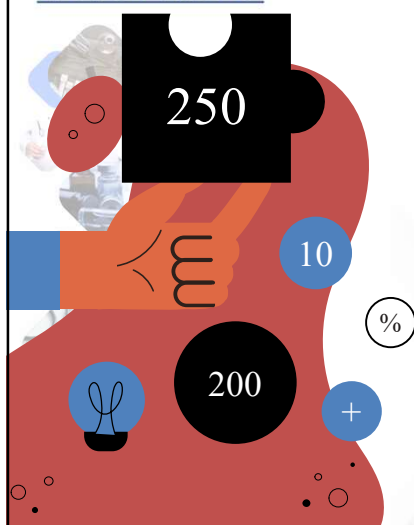
說明：

1. 型一誤差 α ：即所謂的「顯著水準」或「偽陽性的出現機率」，一般建議選用 0.05。
2. 型二誤差 β ：($1-\beta$) 即所謂「檢定力」，一般建議選用 0.2，即當檢定力=0.8時。
3. 預期達到的危險比 (hazard ratio) HR：例如某事件的危險率(hazard rate)為B，有特定因子介入後（如：使用新藥或疫苗）預期該事件的危險率(hazard rate)為A，危險比(hazard ratio)HR 即為 A/B，危險率增加時HR大於1，減少時HR介於0與1之間。
4. 預期該事件整體的發生率(盛行率，流行率) p，請輸入百分比，例如：30%即輸入30。

未來高醫 引領未來


高雄醫學大學
 KAOHSIUNG MEDICAL UNIVERSITY





◆ 參考資料:

- ◆ 李建宏著，生物統計學講義
- ◆ 楊志良著，生物統計學新論(增修二版)，藝軒圖書出版社，2007
- ◆ Jan W.Kuzma原著(史麗珠、林麗華編譯)，基礎生物統計學(Basic statistics for the health sciences)，學富文化事業有限公司，2008
- ◆ Bernard Rosner, Fundamentals of Biostatistics, 7th edition, 2011
- ◆ 臺北醫學大學動物實驗管理系統(IACUC)
<https://iacuc.tmu.edu.tw/2025/09/03/動物實驗樣品數量估算>
- ◆ Arifin, W. N., & Wan Mohammad, W. M. Z. (2017).
Sample Size Calculation in Animal Studies Using Resource Equation Approach.
Malaysian Journal of Medical Sciences, 24(5), 101–105.
<https://doi.org/10.21315/mjms2017.24.5.11>
- ◆ Mead, R. (1988).
The Design of Experiments: Statistical Principles for Practical Applications.
Cambridge University Press.

謝謝聆聽！

Q2什麼是統計檢定力(power)?

96

當有 effect 的時候，power 是偵測到這個 effect 的機率。power 是當事實上是要否定虛無假設，而拒絕虛無假設的機率 (it is the probability of rejecting the null hypothesis when it is in fact false)。再換個方式說，**power 是避免錯誤接受虛無假設的機率**。另外一個作 power analysis 的好處就是決定樣本數，一般來說，power 要多大呢？這要看領域了，通常 0.8 就可以了，有些人會訂在 0.8 到 0.9 之間。

<http://newgenerationresearcher.blogspot.tw/2011/06/power.html>

Q3.alpha=0.05，power=1-beta=0.8

因此數十年以來，量化研究都太過於強調型一錯誤，不過近十年來心理學領域的期刊已經開始規定必須報告型二錯誤或統計檢定力，一般說來大家都根據Cohen(1988)的建議，將型一錯誤容忍值定為5%且型二錯誤容忍值定為20%，也就是說如果我們得到的樣本資料所推論的p值小於0.05且power高於80%時，通常這樣的結果都是能夠被接受的，因為兩個類型錯誤發生機率都蠻低的就代表不管母群體結果為何，推論的結果的犯錯機率都是低的，因此我們會對這個結果非常的有信心！

Q4.「效果量」(effect size、簡單ES)

再來則是介紹「效果量」(effect size、簡單ES)，先以統計術語介紹：「樣本資料所得到的差異是幾個標準差」，再來以T檢定舉例以帮助大家理解，假使我們要比較男生與女生分數的差異，結果抽樣回來的結果顯示男生平均數55分且女生平均數為50分，非常剛好地男生與女生的標準差都是10分，此時根據效果量定義為「樣本資料所得到的差異(男生減掉女生為5分)是幾個標準差(標準差為)10分」，因此答案為5分除以10分=0.5，但是因為我們已經除以標準差，就代表我們已經將變項「去掉單位」，大家可注意到一個訊息：「效果量跟樣本數完全沒有關係」，大家可注意到剛剛d=0.5的計算完全沒有考慮樣本數，因此即使兩個研究，前者樣本數1000人且d=0.2，後者樣本數100人且d=0.5，此時若宣稱後者的效果大於前者的效果，這句話是成立的。

[http://dasanlin888.pixnet.net/blog/post/34468802-](http://dasanlin888.pixnet.net/blog/post/34468802-%E6%A8%A3%E6%9C%AC%E6%95%B8%E8%A6%8F%E5%8A%83%E5%88%9D%E6%8E%A2~%E6%9E%97%E6%98%9F%E5%B8%86%E9%A1%A7%E5%95%8F%E6%95%B4%E7%90%86-(sample-size)%26-p)

[http://dasanlin888.pixnet.net/blog/post/34468802-%E6%A8%A3%E6%9C%AC%E6%95%B8%E8%A6%8F%E5%8A%83%E5%88%9D%E6%8E%A2~%E6%9E%97%E6%98%9F%E5%B8%86%E9%A1%A7%E5%95%8F%E6%95%B4%E7%90%86-\(sample-size\)%26-p](http://dasanlin888.pixnet.net/blog/post/34468802-%E6%A8%A3%E6%9C%AC%E6%95%B8%E8%A6%8F%E5%8A%83%E5%88%9D%E6%8E%A2~%E6%9E%97%E6%98%9F%E5%B8%86%E9%A1%A7%E5%95%8F%E6%95%B4%E7%90%86-(sample-size)%26-p)