

# 淺談假設檢定與效果量

Qi-Wen Ding



# Null hypothesis significance test (NHST)

現今的統計檢定混合了Fisher與Neyman-Pearson的思想：

## 研究規劃階段 (Neyman-Pearson)

1. 設定 $H_0$ 與 $H_1$  (及預期的效果量)
2. 選擇合適的檢定統計量
3. 設定顯著水準 $\alpha$
4. 計算所需樣本數，讓檢定維持足夠的power

## 研究決策階段 (Fisher)

5. 計算 $p$ -value
6. 和 $\alpha$ 比較，並下決策

然而實務上經常忽略研究規劃的步驟...

# 對統計檢定（之誤用）的批評

- 實務執行時過度忽略power的重要性
- 只要樣本數夠大，統計檢定便很容易「顯著」，即便效果非常小 (Amrhein, Greenland & McShane, 2019)
- 即便透過統計檢定推翻 $H_0$ ，仍無法提供更多關於參數的資訊 (Cohen, 1994; Kirk, 1996; Meehl, 1967)
- $p$ -value不是 $P(H_0 \text{ 為真})$ ，也不是 $P(H_0 \text{ 為真} | \text{資料})$
- ...

# 效果量 (effect size) 的定義

因此，不少學者建議：  
在詮釋結果時，更應關注效果的強度

(e.g., APA, 2010, 2020; Citrome & Ketter, 2013; Grissom & Kim, 2012; Kirk, 1996; Maxwell, Camp, & Arvey, 1981)

## Kelley & Preacher (2012)

- **Effect size** is defined as a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest.
- The question of interest might refer to central tendency, variability, association, difference, odds, rate, duration, discrepancy, proportionality, superiority, or degree of fit or misfit, ...

# 效果量之類型

1

*d*-type

描述分數差異

- Cohen's *d*
- Raw mean difference
- ...

2

*r*-type

描述關聯強度

- Pearson's correlation
- Regression coefficient
- ...

3

Others

其他沒有明確分類者

- Relative risk
- Odds ratio
- ...

各指標之優缺點、統計特性與信賴區間計算可參考補充材料表1 (<https://osf.io/n69xs/>)

# 效果量之用途

- 執行事前統計檢定力分析，以估計有效樣本數  
(Cohen, 1962, 1969)
- 量化研究所觀察到的效果強度
- 有利後續學者進行整合分析 (meta-analysis)

# 學者對效果量的期待

- 其點估計應和樣本數及統計顯著性無關  
(Kelley & Preacher, 2012; Rosenthal, 1994)
- 應具有良好的統計性質，如：不偏性、有效性與一致性  
(Kelley & Preacher, 2012)
- 應易於理解 (Pek & Flora, 2018)，能表達研究結果的實用性 (Kirk, 1996)
- 能彰顯研究結果在理論、實務或臨床上的重要性  
(APA, 2010; Thompson, 2002; Wilkinson & the Task Force on Statistical Inference, 1999)



# 效果量的報告準則

1. 作者應在研究結果裡報告效果量，以便讀者體認研究效果的強度或結果的重要性  
(APA, 2010, p. 34; APA, 2020, p. 89)
2. 若測量單位有實務意涵，可報告具有原始單位意義的效果量  
(APA, 2010, p. 34; APA, 2020, p. 89; Wilkinson & the Task Force on Statistical Inference, 1999, p. 599)
3. 應同時報告效果量、效果量的信賴區間或標準誤、統計檢定的結果及其顯著性  
(AERA, 2006, p. 37; APA, 2010, p. 34; APA, 2020, p. 89)
4. 提供讀者足夠的資訊、評估結果的實用性，或是結果在理論或臨床上的重要性  
(APA, 2010, p. 34; APA, 2020, p. 89)

# 效果量的報告準則

5. 與過去類似的研究所報告的效果量相比較，以助讀者評估研究結果橫跨不同樣本、設計與分析之穩定性

(Wilkinson & the Task Force on Statistical Inference, 1999, p. 599)

6. 在解釋效果量時，需考量自身研究的情境脈絡（如：實驗組和控制組參與者的特性、實驗組和控制組所經歷的事件）

(APA, 2010, p. 36; APA, 2020, p. 90)

7. 當統計顯著性與效果量的強度有落差時，需解釋造成此落差的可能原因

- 例：是否由於樣本數不足導致統計檢定力低，因此檢定未達顯著，但效果量值卻很高。

(Chen et al., 2020, p. 576)