

驗證性資料分析

Confirmatory Data Analysis (CDA)



吳漢銘 副教授

國立政治大學 統計學系

wuhm@g.nccu.edu.tw

<http://www.hmwu.idv.tw/>

大綱



- 驗證性資料分析(CDA)定義
- 與探索性資料分析 (EDA) 的差異
- 何時使用 CDA ?
- CDA 分析流程
- CDA 所依賴的基本假設
- CDA 實務應用案例
- 驗證性資料分析方法(舉例)
- 諮詢流程與分工架構

驗證性資料分析(CDA)



- 驗證性資料分析 (Confirmatory Data Analysis, CDA) 是一種以**假設檢定**為核心的統計分析方法，其目的是根據先前提出的**理論模型**或**研究假設**，對資料進行**統計推論**：模型檢定、假設檢驗與統計推論。
- 與探索性資料分析 (EDA) 的差異
 - EDA 是「資料導向」，適合在對資料不熟時使用，了解資料特性與潛在趨勢。
 - CDA 是「假設導向」，適合用來驗證研究假設，進行統計檢定與推論。
 - 目的：CDA 驗證假設，EDA 發掘模式。
 - 推理方式：EDA 為歸納式，CDA 為演繹式。
 - 兩者互補，缺一不可，完整的資料分析流程應包含 EDA + CDA。

比較面向	EDA：探索性資料分析	CDA：驗證性資料分析
分析目標	發掘資料中的模式、異常與潛在結構	檢定事先提出的統計假設
推理方式	歸納式 (Inductive)	演繹式 (Deductive)
開始點	從資料本身出發，無明確假設	基於理論、設計或研究問題提出明確假設
常用方法	敘述統計、圖形化分析 (histogram, scatterplot, PCA 等)	t 檢定、ANOVA、迴歸分析、卡方檢定等
工具目的	了解資料分布、異常值、群集等資訊	判斷假設是否成立，進行統計推論
分析結果	幫助建立新假設、形成研究問題	支持或否定原始假設 (H_0 vs H_1)
應用階段	資料分析前期，準備或探索性研究階段	資料分析後期，驗證性研究與報告階段
範例問題	「資料中是否存在群集或趨勢？」	「這兩組的平均數是否有顯著差異？」



驗證性資料分析 在科學研究中的角色

4/13

1. 支持理論驗證

- CDA 是假設驅動的分析方法，用於檢驗是否有統計證據支持某個理論或假設。
- 幫助研究者：驗證理論模型是否成立、比較不同處理或條件下的效果差異、檢視變項間的關係強度與方向。
- 例：心理學實驗中檢定「某種教學策略是否能顯著提高學習成效」。

EDA → 發現問題

CDA → 驗證假設

它們互為前後，CDA 為研究提供嚴謹性與學術有效性。

2. 提供統計推論依據

- CDA 使用如 t 檢定、ANOVA、迴歸分析、卡方檢定等方法來推斷：結果是否具統計顯著性 (p-value)、效果大小 (effect size)、模型解釋力 (R^2 、pseudo R^2)
- 在學術論文中，CDA 是支撐結論與主張的重要數據證據。

3. 結合實驗設計與資料分析

- CDA 通常與良好的實驗設計配對進行：隨機分派 (Randomization)、對照組 (Control group)、事前設定的樣本大小與檢定力 (Power analysis)
- 可讓研究具有可重複性 (Reproducibility) 與可推論性 (Generalizability)。

4. 作為科學方法的關鍵步驟

- 科學方法的流程：提出假設 → 設計實驗 → 收集資料 → CDA 驗證 → 得出結論
- CDA 是從「觀察」轉為「推論」的橋樑，協助研究結果從個案經驗上升到理論層次。



何時使用 CDA ?

CDA特點：

- 針對事先設定的研究假設進行檢定。
- 分析結果通常用於支持或否定某假設。
- 強調模型的推論力與檢定力（如 p 值、效應量）。

何時使用 CDA ?

- 當你已經有一個明確的研究假設
 - 比較兩組是否有顯著差異、預測變項對結果的影響
- 當你的研究設計為實驗性或準實驗性
 - 有控制組 / 實驗組、前測與後測設計
- 當你需對資料進行統計推論
 - 用於發表論文、做政策建議或預測未來趨勢
- 當你有理論架構或事先設定模型
 - 例如使用 SEM、迴歸、卡方等方法來驗證理論
- 作為探索性資料分析（EDA）之後的確認步驟
 - EDA 發現模式後，使用 CDA 驗證是否顯著

CDA 分析流程



1. **設定研究假設**：定義 H_0 與 H_1 ，例：藥物 A 對血壓有效。
 2. **選擇統計模型**：如 t 檢定、ANOVA、迴歸、SEM。
 3. **資料檢查與前處理**：檢查常態性、處理遺漏值與異常值。
 4. **執行統計分析**：模型建構、參數估計與假設檢定。(含模型選擇及診斷)
 5. **結果解釋與結論**：是否拒絕 H_0 ？效果量與意義解讀。
 6. **撰寫報告與重現性**：完整記錄流程與程式碼。
- **注意事項:**
- CDA 必須在資料分析前就設定明確假設，不可事後根據資料結果隨意修改假設。
 - 分析結果應包含：p 值、信賴區間、效果量 (Effect Size) 及模型診斷與限制討論。
 - 強調實證支持與可重現性。



效果量 (Effect Size) 是什麼?

□ 效果量 (Effect Size)

統計分析中用來衡量實際效應的大小或強度的指標，它不像 p 值僅僅判斷結果是否顯著，而是幫助我們理解「效果有多大」，也就是研究變項之間的關聯強度或差異程度。

□ 為什麼需要效果量？

- p 值告訴你「有沒有差異」。
- 效果量告訴你「差異有多大」。
- 在樣本很大的情況下，即使差異極小， p 值仍可能顯著。
- 因此效果量能提供實務意義 (practical significance) 。

常見的效果量指標：

類型	名稱	用途	解釋
Cohen's d	兩組平均數差異	t 檢定	差異以標準差為單位表示
η^2 (Eta squared)	變異解釋比例	ANOVA	解釋總變異的比例
r 或 r^2	相關或決定係數	相關、迴歸分析	r 表示關聯強度， r^2 表示預測力
OR (Odds Ratio) 或 RR (Risk Ratio)	群體比率差異	邏輯斯迴歸、流行病學研究	機率或風險大小比值
β (Standardized Beta)	標準化迴歸係數	多元迴歸	變數對結果的影響強度

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p}$$

$$\eta^2 = \frac{SS_{\text{effect}}}{SS_{\text{total}}}$$



CDA 所依賴的基本假設

驗證性資料分析 (CDA, Confirmatory Data Analysis) 是一種以統計推論為核心的方法，因此它的正確性與可行性高度依賴一系列統計模型的基本假設。這些假設若未滿足，將可能導致誤導性的結論。若假設違反：

- 應進行前置檢查 (如 Shapiro-Wilk、Levene' s Test)
- 或採用穩健方法 (如非參數檢定、引入控制變項、資料轉換)

CDA 所依賴的基本假設 (依據所選模型可能略有不同) :

1. 樣本獨立性 (Independence of Observations)

- 每個觀測值是彼此獨立的。
- 違反此假設常見於重複量測或群組內資料 (需改用混合模型或群組分析) 。

2. 常態分配 (Normality)

- 資料或誤差項需近似常態分佈 (尤其對 t 檢定、ANOVA、迴歸中的推論有效性很關鍵) 。
- 若樣本數足夠大 (如 $n > 30$)，根據中央極限定理 (CLT) 可以放寬此限制。

3. 變異數同質性 (Homogeneity of Variance)

- 各組的變異數應該相近 (對 ANOVA 重要) 。
- 若違反此假設，可改用 Welch's ANOVA 或非參數方法。

4. 線性關係 (Linearity)

- 自變數與因變數之間應具有線性關係 (迴歸分析特別重要) 。
- 可透過散佈圖或殘差分析來檢查。

5. 無多重共線性 (No Multicollinearity)

- 自變數之間不應過度高度相關 (迴歸分析中重要) 。
- 可透過 VIF (變異數膨脹因子) 檢查。

6. 誤差獨立且隨機 (Independence and Randomness of Errors)

- 模型的誤差項應獨立且服從常態分配，期望值為 0。

7. 正確的模型設定 (Model Specification)

- 所建立的模型需合理反映理論假設與資料特性。
- 包括變項選擇、轉換、交互作用是否適當等。



CDA 實務應用案例

1. 醫學研究：藥物是否有效？

- 假設：新藥 A 可有效降低病患血壓
- 分析方法：獨立樣本 t 檢定或 ANCOVA
- 資料來源：臨床試驗中的控制組與實驗組血壓數據
- 目的：驗證新藥是否顯著優於安慰劑 (placebo)

2. 教育領域：教學法是否提升學習成效？

- 假設：採用翻轉教學的班級，其學習成效高於傳統教學班級
- 分析方法：t 檢定、ANOVA 或混合設計模型
- 資料來源：期末測驗成績、前測與後測分數
- 目的：評估教學方法對學習成效的影響

3. 心理學研究：心理療法是否改善情緒？

- 假設：認知行為治療可顯著減輕焦慮症狀
- 分析方法：重複量數 ANOVA (Repeated Measures ANOVA)
- 資料來源：治療前中後的焦慮量表得分
- 目的：驗證時間與治療交互作用對情緒的影響

CDA 的核心在於：根據既有理論提出假設 → 使用適當統計方法 → 檢驗是否支持假設

4. 行銷分析：行銷策略是否提升銷售？

- 假設：A 廣告版本的點擊率高於 B 版本
- 分析方法：卡方檢定 (Chi-square test)、邏輯斯迴歸
- 資料來源：兩版本的曝光次數與點擊數
- 目的：驗證哪一種行銷手法更有效

5. 公共政策評估：政策是否有效？

- 假設：最低工資上調後失業率未顯著上升
- 分析方法：差異中的差異法 (Difference-in-Differences, DiD)
- 資料來源：政策前後各地區的失業率數據
- 目的：驗證政策因果效果是否成立



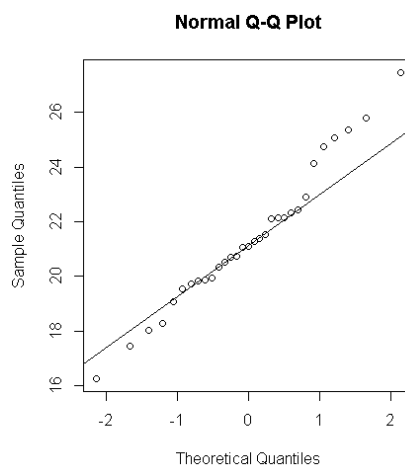
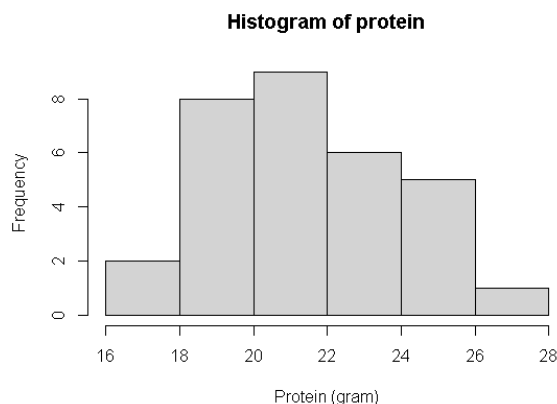
範例:消保官想了解能量棒上的標示 「此能量棒含20公克的蛋白質」是否正確? (t檢定)

$$H_0: \mu = 20, \quad H_1: \mu \neq 20, \alpha = 0.05.$$

31根能量棒的蛋白質含量(克數):

20.70, 27.46, 22.15, 19.85, 21.29, 24.75, 20.75, 22.91, 25.34, 20.33, 21.51, 21.22,
22.14, 19.56, 21.10, 18.04, 24.12, 19.95, 19.72, 18.28, 16.26, 17.46, 20.
25.06, 22.44, 19.08, 19.88, 21.39, 22.33, 25.79

營養成分 每份(50克)	
熱量	190大卡
蛋白質	20克
碳水化合物	17克
總脂肪	6克
飽和脂肪	3.5克
膽固醇	15毫克
鈉	180毫克
膳食纖維	<1克
糖	2克
糖醇	8克



```
> ks.test(log(protein), "pnorm")

One-sample Kolmogorov-Smirnov test

data:  log(protein)
D = 0.99735, p-value = 3.331e-16
alternative hypothesis: two-sided
```

```
> shapiro.test(protein)

Shapiro-Wilk normality test

data:  protein
W = 0.9768, p-value = 0.7191
```

```
> t.test(protein, mu = 20)

One Sample t-test

data:  protein
t = 3.0668, df = 30, p-value = 0.004553
alternative hypothesis: true mean is not equal to 20
95 percent confidence interval:
 20.46771 22.33229
sample estimates:
mean of x
 21.4
```

拒絕「平均蛋白質公克數等於 20」的虛無假設。標示資訊不正確，且蛋白質公克數的母體實際上平均數大於 20。

標籤資訊應該更新，或製造流程應該改善，以製造出平均含 20 公克蛋白質的能量棒。

卡方獨立性檢定

- H_0 : In the population, the two categorical variables are independent.
- H_a : In the population, two categorical variables are dependent.

For testing independence in $I \times J$ contingency tables

$$H_0: \pi_{ij} = \pi_{i+}\pi_{+j} \quad \text{for all } i \text{ and } j$$

$\mu_{ij} = n\pi_{ij} = n\pi_{i+}\pi_{+j}$ as the expected frequency.

estimated expected frequencies.

$$\hat{\mu}_{ij} = n\pi_{i+}\pi_{+j} = n \left(\frac{n_{i+}}{n} \right) \left(\frac{n_{+j}}{n} \right) = \frac{n_{i+}n_{+j}}{n}$$

The *Pearson chi-squared statistic* for testing H_0 is

$$X^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}}$$

The X^2 statistic has approximately a chi-squared distribution, for large n . (WHY?)

Table 2.5. Cross Classification of Party Identification by Gender

Gender	Party Identification			Total
	Democrat	Independent	Republican	
Females	762 (703.7)	327 (319.6)	468 (533.7)	1557
Males	484 (542.3)	239 (246.4)	477 (411.3)	1200
Total	1246	566	945	2757

Note: Estimated expected frequencies for hypothesis of independence in parentheses. Data from 2000 General Social Survey.

```
> M <- as.table(rbind(c(762, 327, 468),
                      c(484, 239, 477)))
> dimnames(M) <- list(gender = c("F", "M"),
+                      party = c("Democrat",
+                                "Independent",
+                                "Republican"))
> M
      party
gender Democrat Independent Republican
F           762           327           468
M           484           239           477
> (res <- chisq.test(M))
      Pearson's Chi-squared test

data:  M
X-squared = 30.07, df = 2, p-value = 2.954e-07
```



驗證性資料分析方法(舉例)

群體比較分析

- 分析兩組或多組平均值是否有顯著差異
- 常見應用：治療組 vs. 控制組
- 單樣本 t 檢定、雙樣本 t 檢定 (獨立樣本)、配對樣本 t 檢定 (前後測)
- 變異數分析
- ANOVA 的事後分析: Tukey HSD、Bonferroni 校正、多重比較問題

迴歸分析與模型診斷

- 簡單與多元迴歸模型
- R^2 、AIC、BIC、殘差圖、Cook's D、共線性檢查
- Logistic 迴歸

諮詢流程與分工架構



在統計諮詢 (Statistical Consulting) 或資料分析專案中，諮詢流程與分工扮演非常關鍵的角色。良好的流程可提升合作效率、確保資料品質與分析的有效性。

1. 問題定義與需求訪談 (Initial Consultation)

- 目標：明確了解研究問題、決定研究目的與期望成果
- 訪談重點：
 - 研究問題是什麼？
 - 需要分析哪些變數？
 - 有哪些資料？是否已收集完畢？
- 分工建議：
 - 研究者/客戶：提出問題與研究背景
 - 統計顧問：引導問題明確化與可分析化

2. 研究設計與資料規劃 (Study Design)

- 目標：協助設計合適的資料收集方法與分析架構
- 關鍵工作：
 - 決定樣本數與抽樣方法
 - 設計問卷/實驗架構
- 分工建議：
 - 統計顧問：設計樣本數估計、資料結構規劃
 - 研究者：回饋可行性與研究需求

3. 資料準備與品質檢查 (Data Preparation)

- 目標：資料整理與確認可用性
- 重點任務：
 - 處理遺漏值、異常值
 - 整併格式、變數命名一致性
- 分工建議：
 - 研究者：提供原始資料與說明
 - 資料分析師：進行資料清理與格式轉換

角色	主要職責
研究者/領域專家	問題定義、資料收集與解釋結果
統計顧問	分析規劃、方法選擇與推論建議
資料分析師	程式實作、資料前處理與模型訓練
專案協調者 (如有)	協調進度與溝通橋樑

4. 資料分析與模型建立 (Statistical Analysis)

- 目標：根據假設與目標執行適當的統計方法
- 常見方法：t 檢定、ANOVA、迴歸、分類模型等
- 分工建議：
 - 資料分析師/統計顧問：執行分析、撰寫初步報告
 - 研究者：確認模型與變數設定的合理性

5. 解釋與報告撰寫 (Interpretation & Reporting)

- 目標：將結果轉化為易理解的研究結論與建議
- 分工建議：
 - 統計顧問：撰寫分析摘要與圖表
 - 研究者：整合為論文或簡報，撰寫討論與結論段落

6. 後續追蹤與修正建議 (Follow-up)

- 目標：根據審查意見或應用情境進行修正與補強
- 分工建議：
 - 統計顧問：協助補強分析或增加附錄資料
 - 研究者：完成回覆審查與最終整合