

Industrial Economics
Prof. Chien-Yuan Sher

Lecture I. 基本工具：迴歸分析入門

Part 1 迴歸分析所能解決的問題以及解決問題的方式

A. 目標：透過大量資料確認或否證兩個變數間是否有所關聯或是存在因果關係。

例，i. 天候不佳會否影響電影首周票房？若答案為是，程度為何？

ii. 公布斷層分布地圖，會否使位於斷層沿線的房價下跌？若然，程度為何？

iii. 十個技安（肥宅）同時同地點跳一下是否會引起地震？

PS.: 台灣孕震構造圖: <https://www.twreporter.org/a/dangerous-fault-architect>

台灣活動斷層分布查詢: <http://fault.moeacgs.gov.tw/MgFault/>

B. 方式或流程（對照關於平均數的計算與檢定）

i. 確立問題，建立假設或模型（依據背景知識，background knowledge）

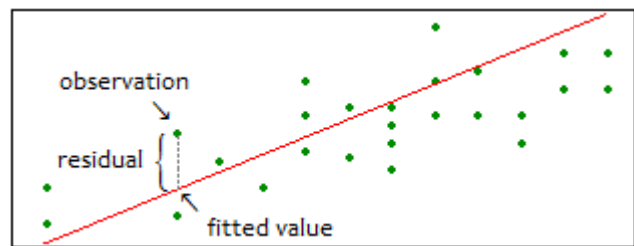
ii. 蒐集資料

iii. 由資料中計算估計值

iv. 進行假設檢定

Part 2 普通最小平方估計式 (ordinary least squares estimators, OLS)

A. 估計式之基本想法：最小平方法



B. 估計式之推導

C. 以向量來表示 OLS 估計式

D. 配適性的衡量： R^2

Part 3 假設檢定

1. 假設檢定的基本想法

2. OLS 估計式的分配

以 OLS 估計式算出來的值是隨機的，所以要知道其分配為何。

A. 傳統對殘差項的假設

B. 現代對殘差項的假設

- i. 殘差項條件分配的期望值為 0，或解釋變數與殘差項不相關

數學意義

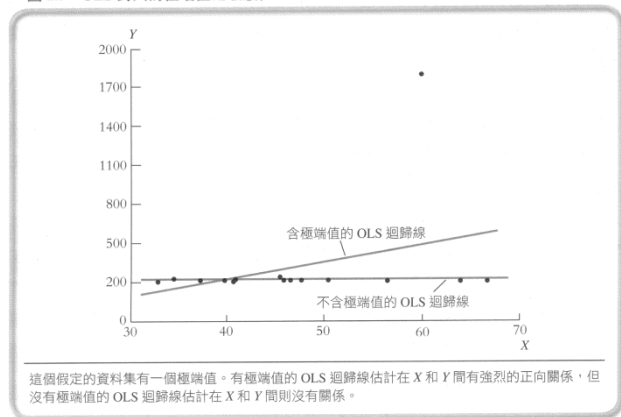
實做上之意涵

- ii. 每次抽取的樣本皆為獨立且相等之分配
實做上之意涵

以時間序列資料或空間資料為例

iii. 沒有極端值/分配不要太奇怪

圖 1.5 OLS 對大的極端值之敏感性



小結: 對殘差項做假設的重點在於要進行假設檢定，吾人需找出估計值的機率分配。

Part 4 其他需要注意的問題

1. 殘差項具同質性 (homoskedasticity) 或異質性 (heteroskedasticity) 的討論

2. 當解釋變數為二元變數時之處理：虛擬變數 (dummy variable)

3. 當有多個解釋變數時，需注意共線性的問題
 - A. 完全多重共線性 (perfect multicollinearity)

虛擬變數陷阱

虛擬變數與常數項共線

- B. 不完全多重共線性 (imperfect multicollinearity)

4. R^2 與 調整後的 R^2 (adjusted R^2)

Part 5 如何評估迴歸分析的結果

1. 內生性 (endogeneity) 或內在效度的問題：假性相關；
 - A. 遺漏變數偏誤 (omitted variable bias)：檢定所需的殘差項假設不成立
 - B. 函數形式錯誤設定 (misspecification)：遺漏重要變數或非線性關係
 - C. 同時因果關係 (simultaneous causality)：檢定所需的殘差項假設不成立
 - D. 變數的測量誤差 (errors-in-variable bias)：檢定所需的殘差項假設可能不成立
 - E. 樣本選擇 (sample selection bias)

2. 外在效度的問題：適當解讀、推廣與運用迴歸分析所得之結果；理解研究限制
 - A. 不要聽到「演算法」或其他看似高深的資料科學名詞就覺得好棒棒，要理解它背後的原理與限制

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176): 1203-1205.

- B. 如何適當理解「統計上顯著」的意涵；正確理解 p -value

林澤民 (2018). p 值的陷阱: 為何頂級方法論期刊做出禁用 p 值的決定? 關鍵評論網。
<https://www.thenewslens.com/article/88282>

- C. 將結果推廣至其他類型/其他年的樣本

3. 因果推論與干預

Lucas, R. (1976). *Econometric Policy Evaluation: A Critique*. pp. 19–46 in Brunner, K. and Meltzer, A (eds.), *The Phillips Curve and Labor Markets. Carnegie-Rochester Conference Series on Public Policy. 1*. New York: Elsevier.

Part 6 實際解讀迴歸分析的結果

表 4.1 利用美國加州小學學區的資料做測驗成績對生師比和控制變數學生特徵的迴歸結果

應變數：學區的平均測驗分數

自變數	(1)	(2)	(3)	(4)	(5)
生師比 (X_1)	-2.28** (0.52)	-1.10* (0.43)	-1.00** (0.27)	-1.31** (0.34)	-1.01** (0.27)
英文學習者的百分比 (X_2)		-0.650** (0.031)	-0.122** (0.033)	-0.488** (0.030)	-0.130** (0.036)
享有午餐補助資格的學生百分比 (X_3)			-0.547** (0.024)		-0.529** (0.038)
享有收入援助資格的學生百分比 (X_4)				-0.790** (0.068)	0.048 (0.059)
截距	698.9** (10.4)	686.0** (8.7)	700.2** (5.6)	698.0** (6.9)	700.4** (5.5)
總結統計式					
SER	18.58	14.46	9.08	11.65	9.08
R^2	0.049	0.424	0.773	0.626	0.773
n	420	420	420	420	420

這些迴歸結果是利用附錄 4.1 中描述的美國加州 K-8 學區的資料所估計，標準誤在係數下面的括號裡。利用雙尾檢定，單一個係數在 5% 的顯著水準下顯著為*，或 1% 的顯著水準下顯著為**。