

## 數量方法與金融應用第二次作業

繳交日期 June 13, 2026

Spring 2026

1. 某研究探討企業 ESG 表現對公司績效 (ROA) 的影響，模型如下：

$$ROA_i = \beta_0 + \beta_1 \times ESG_i + \beta_2 \times SIZE_i + \beta_3 \times LEV_i + \beta_4 \times R\&D_i + \varepsilon_i$$

研究者發現：

- 整體 F 檢定顯著；
- ESG、SIZE、LEV、R&D 四個變數的 t 檢定係數都不顯著；
- ESG 與 SIZE 的相關係數為 0.92；
- $VIF(ESG) = 14.8$ ；

請回答以下問題：

- a. 解釋什麼是多重共線性，對 OLS 的迴歸估計有何影響？
- b. 為何個別 t 檢定不顯著，但整體 F 檢定可能顯著？
- c. 說明 VIF 的定義，變數 ESG 的  $VIF=14.8$  與迴歸分析有何關聯？
- d. 請比較以下三種改善共線性的方法，並比較其優缺點：請比較以下方法：刪除變數、主成分分析、ridge regression。
- e. 若研究目的是「預測」而非「因果推論」，是否仍須嚴格處理共線性問題？請說明。

2. 某研究探討企業數位化程度 (Digitalization) 對公司獲利能力 (ROA) 的影響：

$$ROA_i = \beta_0 + \beta_1 \times Digital_i + \beta_2 \times Size_i + \beta_3 \times Debt_i + \varepsilon_i$$

研究者得到以下結果：

- $Adjusted R^2 = 0.68$ 、 $R^2 = 0.68$
- 整體 F 檢定顯著、部分變數 t 檢定不顯著
- 殘差圖呈現非線性型態、Q-Q plot 顯示尾部偏離常態

請回答：

- a.  $Adjusted R^2$  與一般  $R^2$  有何差異？
- b. 為何整體 F 檢定顯著，但部分變數的 t 檢定可能不顯著？
- c. 殘差圖出現非線性型態，可能代表哪些問題？
- d. 誤差項不符合常態分配時，對 OLS 推論有何影響？
- e. 請提出至少三種模型診斷或改善方法，並說明其適用情境。

註：學號除以 2，餘數為 0、1 分別回答第一題、第二題。

3. 某金融研究以橫斷面資料分析 CEO 薪酬：

$$\log(\text{Salary}_i) = \beta_0 + \beta_1 \times \text{Performance}_i + \beta_2 \times \text{FirmSize}_i + \varepsilon_i$$

研究者發現殘差圖呈現明顯扇形擴散（隨著 Y 值而擴大）。

請回答：

- 什麼是誤差的異質變異（heteroskedasticity，或譯為變異數異質性）？
- 在迴歸分析中異質變異違反哪一項假設？
- 異質變異對迴歸分析的估計有什麼影響，例如：參數估計值的不偏性、誤差變異數的估計、整體迴歸方程式的假設檢定。
- 有不少異質變異的檢定方法，包括：White test、Breusch-Pagan test 等，請列舉三種這類型檢定並說明各方法的基本概念。
- 誤差異質變異的調整方法包括以下幾種：robust standard error、weighted least squares、variable transformation，請說明這幾種方法比較適合哪些場合，以及使用限制（或不合適時機）？

4. 某研究分析消費者是否會購買金融商品（1=購買，0=未購買）：

$$P(Y_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{Income}_i + \beta_2 \text{Age}_i + \beta_3 \text{Risk}_i)}}$$

估計結果如下：

| Variable        | Coefficient |
|-----------------|-------------|
| Income          | 0.80        |
| Age             | -0.03       |
| Risk Preference | 1.25        |

請回答：

- 為何不適合直接使用線性機率模型（LPM）？
- Logistic regression 的基本概念為何？
- 如何解釋 Income 和 Age 的係數？
- 如何解釋 Risk Preference 的 odds ratio？（提示： $e^{1.25}=3.49$ ）
- OLS 的 marginal effect 與 odds ratio 有何不同？
- 在商學研究中哪些情境適合使用 logistic regression？

5. 某研究欲建立企業財務危機預測模型，原始資料包含 25 個可能解釋變數，包括：財務比率、公司治理指標、市場變數、總體經濟指標。研究者擔心變數間高度相關、模型過於複雜，尤其本研究的樣本數有限，過多變數可能導致 overfitting，因此考慮使用變數選取(variable selection)建立最終模型。

請回答：

- 為何在迴歸分析中需要進行變數選取(variable selection)？
- 說明以下方法的基本概念及操作流程：Best Subset Selection、Forward Selection、Backward Elimination、Stepwise Regression。
- Best Subset 與 Stepwise Regression 的主要差異為何？
- 在以下情況中哪種方法較適合，請說明理由：
  - 變數數量很多但樣本數有限
  - 重視模型解釋性
  - 重視預測能力
  - 計算資源有限
- 使用 Stepwise Regression 是否需要考慮：overfitting、multicollinearity、model instability、inflated Type I error？
- 除了 stepwise 請再列舉兩種現代變數選取方法，並簡述其基本概念。

註：以下為第三題至第五題的安排→吳鈞麟(第三題)、歐陽宏(第四題)、林諺琮(第五題)、陳威成(第三題)、翁振源(第四題)。

### 加分題：

6. 某研究比較三個模型預測企業破產機率：

| Model   | Adjusted R <sup>2</sup> | AIC | BIC |
|---------|-------------------------|-----|-----|
| Model A | 0.42                    | 820 | 845 |
| Model B | 0.48                    | 790 | 835 |
| Model C | 0.51                    | 805 | 870 |

- AIC 與 BIC 的基本概念為何？
- 為何模型不能只比較 R<sup>2</sup>？
- AIC 與 BIC 對模型複雜度的懲罰有何不同？
- 若研究目的為解釋、預測應如何選擇模型？
- 請說明 overfitting、underfitting 的概念與可能影響。