統計學講義

第六部分：迴歸分析

1 多元線性迴歸模型

1.1 模型設定

定義 1.1 (多元線性迴歸模型). 設有 n 個觀測值，每個觀測包含一個依變數 Y_i 和 k 個自變數 $X_{i1}, \ldots, X_{ik}$。**多元線性迴歸模型**假設：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, \ldots, n$$

其中 β_0 為截距，$\beta_1, \ldots, \beta_k$ 為斜率係數，ε_i 為誤差項。

定義 1.2 (矩陣表示法). 將模型寫成矩陣形式：

$$\boxed{\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}}$$

其中：

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n\times 1}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{pmatrix}_{n\times(k+1)}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{(k+1)\times 1}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n\times 1}$$

$\mathbf{X}$ 稱為**設計矩陣**（design matrix），第一行全為 1 對應截距項。

性質 1.1 (古典假設).
(A1) **線性關係**：$\mathsf{E}[\mathbf{Y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$
(A2) **誤差期望為零**：$\mathsf{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$
(A3) **球面誤差**（spherical errors）：$\mathrm{var}(\boldsymbol{\varepsilon}|\mathbf{X}) = \sigma^2\mathbf{I}_n$
- 同質變異數（homoscedasticity）：$\mathrm{var}(\varepsilon_i) = \sigma^2$（常數）
- 誤差不相關：$\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$ 對 $i \neq j$

(A4) **滿秩假設**：$\mathrm{rank}(\mathbf{X}) = k + 1$（$\mathbf{X}$ 的行向量線性獨立）
(A5) **常態假設**（用於精確推論）：$\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$

註. 假設 (A1)–(A4) 稱為 **Gauss-Markov 假設**，保證 OLS 為 BLUE：

- **OLS**：Ordinary Least Squares，普通最小平方法
- **BLUE**：Best Linear Unbiased Estimator，最佳線性不偏估計量

假設 (A5) 用於有限樣本下的精確 t 和 F 檢定。

2 最小平方法

2.1 OLS 估計量

定義 2.1 (最小平方法). **普通最小平方法**（Ordinary Least Squares, OLS）選擇 $\hat{\boldsymbol{\beta}}$ 以最小化殘差平方和：

$$\mathrm{SSE}(\boldsymbol{\beta}) = \sum_{i=1}^{n}(Y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

引理 2.1 (矩陣微分公式). 設 $\boldsymbol{\beta}$ 為 $p \times 1$ 向量，$\mathbf{a}$ 為 $p \times 1$ 常數向量，$\mathbf{A}$ 為 $p \times p$ 常數矩陣。則：

(i) $\dfrac{\partial}{\partial \boldsymbol{\beta}}(\mathbf{a}^\top \boldsymbol{\beta}) = \dfrac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}^\top \mathbf{a}) = \mathbf{a}$

(ii) $\dfrac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}^\top \mathbf{A}\boldsymbol{\beta}) = (\mathbf{A} + \mathbf{A}^\top)\boldsymbol{\beta}$
當 $\mathbf{A}$ 為對稱矩陣（$\mathbf{A} = \mathbf{A}^\top$）時，簡化為 $2\mathbf{A}\boldsymbol{\beta}$。

證明 (i) 設 $\mathbf{a} = (a_1, \ldots, a_p)^\top$，則 $\mathbf{a}^\top \boldsymbol{\beta} = \sum_{j=1}^p a_j \beta_j$。對 β_i 微分：$\frac{\partial}{\partial \beta_i}(\mathbf{a}^\top \boldsymbol{\beta}) = a_i$。故 $\frac{\partial}{\partial \boldsymbol{\beta}}(\mathbf{a}^\top \boldsymbol{\beta}) = \mathbf{a}$。

(ii) 設 $\mathbf{A} = (a_{ij})$，則 $\boldsymbol{\beta}^\top \mathbf{A}\boldsymbol{\beta} = \sum_{i=1}^p \sum_{j=1}^p a_{ij}\beta_i\beta_j$。對 β_ℓ 微分：

$$\frac{\partial}{\partial \beta_\ell}(\boldsymbol{\beta}^\top \mathbf{A}\boldsymbol{\beta}) = \sum_{j=1}^p a_{\ell j}\beta_j + \sum_{i=1}^p a_{i\ell}\beta_i = (\mathbf{A}\boldsymbol{\beta})_\ell + (\mathbf{A}^\top \boldsymbol{\beta})_\ell$$

故 $\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}^\top \mathbf{A}\boldsymbol{\beta}) = \mathbf{A}\boldsymbol{\beta} + \mathbf{A}^\top \boldsymbol{\beta} = (\mathbf{A} + \mathbf{A}^\top)\boldsymbol{\beta}$。

當 $\mathbf{A} = \mathbf{A}^\top$ 時，$(\mathbf{A} + \mathbf{A}^\top)\boldsymbol{\beta} = 2\mathbf{A}\boldsymbol{\beta}$。 □

定理 2.1 (OLS 估計量). 在假設 (A4) 下，OLS 估計量為：

$$\boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y}}$$

證明 方法一：微分法

展開 SSE：

$$\begin{aligned}\mathrm{SSE}(\boldsymbol{\beta}) &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{Y}^\top \mathbf{Y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}\end{aligned}$$

(第二步利用 $\mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta}$ 為純量，故等於其轉置 $\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y}$。)

對 $\boldsymbol{\beta}$ 微分。由引理 2.1，$\mathbf{X}^\top \mathbf{X}$ 為對稱矩陣，故：

$$\frac{\partial \mathrm{SSE}}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

這給出**正規方程式**（normal equations）：

$$\mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{Y}$$

由假設 (A4)，$\mathbf{X}^\top \mathbf{X}$ 可逆，故：

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y}$$

二階條件：Hessian 矩陣 $\frac{\partial^2 \mathrm{SSE}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} = 2\mathbf{X}^\top \mathbf{X}$ 為正定（當 $\mathrm{rank}(\mathbf{X}) = k + 1$ 時），確認為最小值。

方法二：投影法

$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ 是 $\mathbf{Y}$ 在 $\mathbf{X}$ 的行空間（column space）上的正交投影。正交條件要求殘差 $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$ 與 $\mathbf{X}$ 的每一行正交：

$$\mathbf{X}^\top(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \implies \mathbf{X}^\top \mathbf{Y} = \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\beta}} \implies \hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y}$$

□

定義 2.2 (帽子矩陣與殘差製造矩陣). 定義**帽子矩陣**（hat matrix）：

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$$

則擬合值為 $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$，殘差為 $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathbf{M}\mathbf{Y}$，其中 $\mathbf{M} = \mathbf{I} - \mathbf{H}$ 為**殘差製造矩陣**（residual maker matrix）。

定義 2.3 (冪等矩陣). 矩陣 $\mathbf{A}$ 稱為**冪等矩陣**（idempotent matrix），若 $\mathbf{A}^2 = \mathbf{A}$。

引理 2.2 (冪等矩陣的跡與秩). 設 $\mathbf{A}$ 為 $n \times n$ 冪等矩陣，則 $\text{tr}(\mathbf{A}) = \text{rank}(\mathbf{A})$。

證明 由於 $\mathbf{A}^2 = \mathbf{A}$，$\mathbf{A}$ 的特徵值 λ 滿足 $\lambda^2 = \lambda$，即 $\lambda(\lambda - 1) = 0$，故 $\lambda \in \{0, 1\}$。

設 $\mathbf{A}$ 有 r 個特徵值為 1，其餘 $n - r$ 個為 0。則：

- $\text{tr}(\mathbf{A}) = r \cdot 1 + (n - r) \cdot 0 = r$
- $\text{rank}(\mathbf{A}) = r$（等於非零特徵值個數）

故 $\text{tr}(\mathbf{A}) = \text{rank}(\mathbf{A})$。 □

定理 2.2 (標準常態向量的二次型分配). 設 $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_n)$，$\mathbf{A}$ 為 $n \times n$ 對稱冪等矩陣（$\mathbf{A}^\top = \mathbf{A}$，$\mathbf{A}^2 = \mathbf{A}$）。則：

$$\mathbf{z}^\top\mathbf{A}\mathbf{z} \sim \chi^2_r, \quad \text{其中} r = \text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A})$$

證明 **步驟一：對稱冪等矩陣的譜分解**

由於 $\mathbf{A}$ 對稱，存在正交矩陣 $\mathbf{P}$（$\mathbf{P}^\top\mathbf{P} = \mathbf{P}\mathbf{P}^\top = \mathbf{I}$）使得：

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^\top$$

其中 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ 為特徵值對角矩陣。

由引理 2.2，$\mathbf{A}$ 的特徵值只能是 0 或 1。設有 r 個特徵值為 1，不失一般性令：

$$\mathbf{\Lambda} = \text{diag}(\underbrace{1, \ldots, 1}_{r \text{ 個}}, \underbrace{0, \ldots, 0}_{n-r \text{ 個}})$$

步驟二：正交變換保持標準常態性

令 $\mathbf{w} = \mathbf{P}^\top\mathbf{z}$。由於 $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_n)$：

- $\mathsf{E}[\mathbf{w}] = \mathbf{P}^\top\mathsf{E}[\mathbf{z}] = \mathbf{0}$
- $\text{var}(\mathbf{w}) = \mathbf{P}^\top\text{var}(\mathbf{z})\mathbf{P} = \mathbf{P}^\top\mathbf{I}\mathbf{P} = \mathbf{I}_n$

故 $\mathbf{w} \sim N(\mathbf{0}, \mathbf{I}_n)$，即 $w_1, \ldots, w_n$ 為獨立標準常態變數。

步驟三：計算二次型

$$\begin{aligned}
\mathbf{z}^\top\mathbf{A}\mathbf{z} &= \mathbf{z}^\top\mathbf{P}\mathbf{\Lambda}\mathbf{P}^\top\mathbf{z} \\
&= (\mathbf{P}^\top\mathbf{z})^\top\mathbf{\Lambda}(\mathbf{P}^\top\mathbf{z}) \\
&= \mathbf{w}^\top\mathbf{\Lambda}\mathbf{w} \\
&= \sum_{i=1}^n \lambda_i w_i^2 \\
&= \sum_{i=1}^r 1 \cdot w_i^2 + \sum_{i=r+1}^n 0 \cdot w_i^2 \\
&= \sum_{i=1}^r w_i^2
\end{aligned}$$

步驟四：結論

由於 $w_1, \ldots, w_r$ 為獨立標準常態變數，$\sum_{i=1}^r w_i^2 \sim \chi_r^2$。 □

性質 2.1 (**H** 和 **M** 的性質)**.**
(i) **H** 和 **M** 皆為**對稱冪等矩陣**（symmetric idempotent）：$\mathbf{H}^\top = \mathbf{H}$，$\mathbf{H}^2 = \mathbf{H}$；$\mathbf{M}^\top = \mathbf{M}$，$\mathbf{M}^2 = \mathbf{M}$
(ii) $\mathbf{HX} = \mathbf{X}$，$\mathbf{MX} = \mathbf{0}$
(iii) $\mathbf{HM} = \mathbf{MH} = \mathbf{0}$
(iv) $\mathrm{tr}(\mathbf{H}) = k+1$，$\mathrm{tr}(\mathbf{M}) = n-k-1$

證明 (i) **H 的對稱性：**

$$\mathbf{H}^\top = [\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]^\top = \mathbf{X}[(\mathbf{X}^\top\mathbf{X})^{-1}]^\top\mathbf{X}^\top = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{H}$$

(利用 $(\mathbf{X}^\top\mathbf{X})^{-1}$ 對稱，因為 $\mathbf{X}^\top\mathbf{X}$ 對稱。)

H 的冪等性：

$$\mathbf{H}^2 = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \cdot \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{H}$$

$\mathbf{M} = \mathbf{I} - \mathbf{H}$ 的對稱性和冪等性由 **H** 直接推得。

(ii) $\mathbf{HX} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X} = \mathbf{XI} = \mathbf{X}$

$\mathbf{MX} = (\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{X} - \mathbf{X} = \mathbf{0}$

(iii) $\mathbf{HM} = \mathbf{H}(\mathbf{I} - \mathbf{H}) = \mathbf{H} - \mathbf{H}^2 = \mathbf{H} - \mathbf{H} = \mathbf{0}$

(iv) $\mathrm{tr}(\mathbf{H})$ **的計算：**

由引理 2.2，$\mathrm{tr}(\mathbf{H}) = \mathrm{rank}(\mathbf{H})$。

H 將任意向量投影到 **X** 的行空間（維度 $k+1$），故 $\mathrm{rank}(\mathbf{H}) = k+1$。

因此 $\mathrm{tr}(\mathbf{H}) = k+1$。

$\mathrm{tr}(\mathbf{M})$ **的計算：**

$$\mathrm{tr}(\mathbf{M}) = \mathrm{tr}(\mathbf{I} - \mathbf{H}) = \mathrm{tr}(\mathbf{I}) - \mathrm{tr}(\mathbf{H}) = n - (k+1) = n-k-1$$

□

註 (誤差與殘差的區別)**. 誤差**（error）$\boldsymbol{\varepsilon}$ 與**殘差**（residual）**e** 是不同的概念：

- **誤差** $\varepsilon_i = Y_i - \mathbf{x}_i^\top\boldsymbol{\beta}$：真實模型中 Y_i 與其期望值的偏離，**不可觀測**（因為真實參數 $\boldsymbol{\beta}$ 未知）
- **殘差** $e_i = Y_i - \mathbf{x}_i^\top\hat{\boldsymbol{\beta}} = Y_i - \hat{Y}_i$：觀測值與擬合值的偏離，**可觀測**

兩者的關係：由 $\mathbf{e} = \mathbf{MY} = \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{M}\boldsymbol{\varepsilon}$（因為 $\mathbf{MX} = \mathbf{0}$），殘差是誤差經過矩陣 **M** 變換後的結果。

性質 2.2 (正規方程式的意涵)**.** OLS 殘差滿足：

(i) $\mathbf{X}^\top\mathbf{e} = \mathbf{0}$（殘差與所有自變數正交）
(ii) $\sum_{i=1}^n e_i = 0$（殘差和為零）
(iii) $\hat{\mathbf{Y}}^\top\mathbf{e} = 0$（殘差與擬合值正交）

證明 (i) 由正規方程式 $\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^\top\mathbf{Y}$，移項得：

$$\mathbf{X}^\top\mathbf{Y} - \mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \implies \mathbf{X}^\top(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \implies \mathbf{X}^\top\mathbf{e} = \mathbf{0}$$

(ii) 設計矩陣 **X** 的第一行為全 1 向量 $\mathbf{1} = (1, 1, \ldots, 1)^\top$。由 (i)，$\mathbf{X}^\top\mathbf{e} = \mathbf{0}$ 的第一個分量為：

$$\mathbf{1}^\top\mathbf{e} = \sum_{i=1}^n e_i = 0$$

注意：此性質僅在模型包含截距項時成立。

(iii) 由 $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$：

$$\hat{\mathbf{Y}}^\top \mathbf{e} = (\mathbf{X}\hat{\boldsymbol{\beta}})^\top \mathbf{e} = \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{e} = \hat{\boldsymbol{\beta}}^\top \cdot \mathbf{0} = 0$$

此性質說明擬合值與殘差在 n 維空間中正交，這正是 OLS 的幾何意義：$\hat{\mathbf{Y}}$ 是 $\mathbf{Y}$ 在 $\mathbf{X}$ 行空間上的正交投影。 □

2.2 OLS 估計量的統計性質

定理 2.3 (不偏性)**.** 在假設 (A1)–(A2) 下：

$$\boxed{\mathsf{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = \boldsymbol{\beta}}$$

證明 將 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 代入 $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$：

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}
\end{aligned}$$

取條件期望：

$$\mathsf{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathsf{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \cdot \mathbf{0} = \boldsymbol{\beta}$$

□

定理 2.4 (變異數-共變異數矩陣)**.** 在假設 (A1)–(A4) 下：

$$\boxed{\mathrm{var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}}$$

證明 由 $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}$：

$$\begin{aligned}
\mathrm{var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) &= \mathsf{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top|\mathbf{X}] \\
&= \mathsf{E}\left[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}|\mathbf{X}\right] \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathsf{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top|\mathbf{X}]\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \quad \text{（由假設 A3）} \\
&= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}
\end{aligned}$$

□

註. $(\mathbf{X}^\top\mathbf{X})^{-1}$ 的對角元素乘以 σ^2 給出各 $\hat{\beta}_j$ 的變異數，非對角元素乘以 σ^2 給出共變異數。

定理 2.5 (Gauss-Markov 定理)**.** 在假設 (A1)–(A4) 下，OLS 估計量 $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的**最佳線性不偏估計量**（Best Linear Unbiased Estimator, BLUE）。

即：對於任意線性不偏估計量 $\tilde{\boldsymbol{\beta}} = \mathbf{C}\mathbf{Y}$（其中 $\mathbf{C}$ 為常數矩陣），有

$$\mathrm{var}(\tilde{\beta}_j) \geqslant \mathrm{var}(\hat{\beta}_j), \quad \forall j$$

或更一般地，$\mathrm{var}(\tilde{\boldsymbol{\beta}}) - \mathrm{var}(\hat{\boldsymbol{\beta}})$ 為半正定矩陣。

證明 設 $\tilde{\boldsymbol{\beta}} = \mathbf{CY}$ 為 $\boldsymbol{\beta}$ 的線性不偏估計量，其中 $\mathbf{C}$ 為 $(k+1)\times n$ 常數矩陣。

步驟一：不偏性的必要條件

$\mathsf{E}[\tilde{\boldsymbol{\beta}}] = \mathbf{C}\mathsf{E}[\mathbf{Y}] = \mathbf{CX}\boldsymbol{\beta} = \boldsymbol{\beta}$ 對所有 $\boldsymbol{\beta}$ 成立，故必須 $\mathbf{CX} = \mathbf{I}_{k+1}$。

步驟二：分解 C

令 $\mathbf{C} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{D}$，其中 $\mathbf{D} = \mathbf{C} - (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$。

由 $\mathbf{CX} = \mathbf{I}$：

$$\mathbf{DX} = \mathbf{CX} - (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X} = \mathbf{I} - \mathbf{I} = \mathbf{0}$$

步驟三：計算 $\mathrm{var}(\tilde{\boldsymbol{\beta}})$

$$\begin{aligned}
\mathrm{var}(\tilde{\boldsymbol{\beta}}) &= \mathrm{var}(\mathbf{CY}) = \mathbf{C}\,\mathrm{var}(\mathbf{Y})\mathbf{C}^\top = \sigma^2\mathbf{CC}^\top \\
&= \sigma^2\left[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{D}\right]\left[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} + \mathbf{D}^\top\right] \\
&= \sigma^2\left[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{D}^\top + \mathbf{DX}(\mathbf{X}^\top\mathbf{X})^{-1} + \mathbf{DD}^\top\right]
\end{aligned}$$

由 $\mathbf{DX} = \mathbf{0}$，中間兩項為零（$(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{D}^\top = [(\mathbf{DX})(\mathbf{X}^\top\mathbf{X})^{-1}]^\top = \mathbf{0}$）：

$$\mathrm{var}(\tilde{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} + \sigma^2\mathbf{DD}^\top = \mathrm{var}(\hat{\boldsymbol{\beta}}) + \sigma^2\mathbf{DD}^\top$$

由於 $\mathbf{DD}^\top$ 為半正定矩陣，$\mathrm{var}(\tilde{\boldsymbol{\beta}}) - \mathrm{var}(\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{DD}^\top \succeq \mathbf{0}$。

等號成立當且僅當 $\mathbf{D} = \mathbf{0}$，即 $\mathbf{C} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$，此時 $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$。 □

2.3 σ^2 的估計

引理 2.3 (二次型的期望值). 設 $\mathbf{z}$ 為 $n\times 1$ 隨機向量，$\mathsf{E}[\mathbf{z}] = \boldsymbol{\mu}$，$\mathrm{var}(\mathbf{z}) = \boldsymbol{\Sigma}$。設 $\mathbf{A}$ 為 $n\times n$ 常數矩陣。則：

$$\mathsf{E}[\mathbf{z}^\top\mathbf{A}\mathbf{z}] = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top\mathbf{A}\boldsymbol{\mu}$$

特別地，當 $\mathsf{E}[\mathbf{z}] = \mathbf{0}$ 時，$\mathsf{E}[\mathbf{z}^\top\mathbf{A}\mathbf{z}] = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})$。

證明 $\mathbf{z}^\top\mathbf{A}\mathbf{z} = \sum_{i=1}^n\sum_{j=1}^n a_{ij}z_iz_j$ 為純量。取期望：

$$\mathsf{E}[\mathbf{z}^\top\mathbf{A}\mathbf{z}] = \sum_{i=1}^n\sum_{j=1}^n a_{ij}\mathsf{E}[z_iz_j]$$

由 $\mathrm{cov}(z_i, z_j) = \mathsf{E}[z_iz_j] - \mu_i\mu_j$，得 $\mathsf{E}[z_iz_j] = \Sigma_{ij} + \mu_i\mu_j$。代入：

$$\begin{aligned}
\mathsf{E}[\mathbf{z}^\top\mathbf{A}\mathbf{z}] &= \sum_{i,j} a_{ij}(\Sigma_{ij} + \mu_i\mu_j) \\
&= \sum_{i,j} a_{ij}\Sigma_{ij} + \sum_{i,j} a_{ij}\mu_i\mu_j \\
&= \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top\mathbf{A}\boldsymbol{\mu}
\end{aligned}$$

(第一項利用 $\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) = \sum_i(\mathbf{A}\boldsymbol{\Sigma})_{ii} = \sum_i\sum_j a_{ij}\Sigma_{ji} = \sum_{i,j} a_{ij}\Sigma_{ij}$，因 $\boldsymbol{\Sigma}$ 對稱。) □

定理 2.6 (MSE 的不偏性)**.**

$$s^2 = \mathrm{MSE} = \frac{\mathrm{SSE}}{n-k-1} = \frac{\mathbf{e}^\top \mathbf{e}}{n-k-1}$$

是 σ^2 的不偏估計量。

證明　步驟一：將 SSE **表示為誤差的二次型**

由 $\mathbf{e} = \mathbf{MY} = \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{MX}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\varepsilon} = \mathbf{M}\boldsymbol{\varepsilon}$（因為 $\mathbf{MX} = \mathbf{0}$）。

故：

$$\mathrm{SSE} = \mathbf{e}^\top \mathbf{e} = (\mathbf{M}\boldsymbol{\varepsilon})^\top (\mathbf{M}\boldsymbol{\varepsilon}) = \boldsymbol{\varepsilon}^\top \mathbf{M}^\top \mathbf{M}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\top \mathbf{M}\boldsymbol{\varepsilon}$$

（最後一步利用 $\mathbf{M}$ 對稱冪等：$\mathbf{M}^\top\mathbf{M} = \mathbf{MM} = \mathbf{M}$。）

步驟二：計算 $\mathsf{E}[\mathrm{SSE}]$

應用引理 2.3，取 $\mathbf{z} = \boldsymbol{\varepsilon}$，$\mathbf{A} = \mathbf{M}$。

由假設 (A2)，$\mathsf{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$（即 $\boldsymbol{\mu} = \mathbf{0}$）。由假設 (A3)，$\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$（即 $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$）。

由引理 2.3：

$$\begin{aligned}
\mathsf{E}[\mathrm{SSE}] &= \mathsf{E}[\boldsymbol{\varepsilon}^\top \mathbf{M}\boldsymbol{\varepsilon}] \\
&= \mathrm{tr}(\mathbf{M} \cdot \sigma^2\mathbf{I}) + \mathbf{0}^\top \mathbf{M}\mathbf{0} \\
&= \sigma^2 \,\mathrm{tr}(\mathbf{MI}) + 0 \\
&= \sigma^2 \,\mathrm{tr}(\mathbf{M})
\end{aligned}$$

步驟三：計算 $\mathrm{tr}(\mathbf{M})$

由性質 2.1(iv)，$\mathrm{tr}(\mathbf{M}) = n-k-1$。

故 $\mathsf{E}[\mathrm{SSE}] = \sigma^2(n-k-1)$。

步驟四：結論

$$\mathsf{E}[\mathrm{MSE}] = \mathsf{E}\left[\frac{\mathrm{SSE}}{n-k-1}\right] = \frac{\mathsf{E}[\mathrm{SSE}]}{n-k-1} = \frac{\sigma^2(n-k-1)}{n-k-1} = \sigma^2$$

故 MSE 是 σ^2 的不偏估計量。 □

3　簡單線性迴歸：$k = 1$ 的特例

當只有一個自變數時（$k = 1$），多元迴歸退化為**簡單線性迴歸**：

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

此時設計矩陣為：

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}$$

定理 3.1 (簡單迴歸的顯式公式)**.** 當 $k = 1$ 時，OLS 估計量為：

$$\boxed{\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

$$\boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}}$$

其中 $S_{XX} = \sum(X_i - \bar{X})^2$，$S_{XY} = \sum(X_i - \bar{X})(Y_i - \bar{Y})$。

證明　計算 $\mathbf{X}^\top\mathbf{X}$ 和 $\mathbf{X}^\top\mathbf{Y}$：

$$\mathbf{X}^\top\mathbf{X} = \begin{pmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix} = \begin{pmatrix} n & n\bar{X} \\ n\bar{X} & \sum X_i^2 \end{pmatrix}$$

$$\mathbf{X}^\top\mathbf{Y} = \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix} = \begin{pmatrix} n\bar{Y} \\ \sum X_i Y_i \end{pmatrix}$$

$\mathbf{X}^\top\mathbf{X}$ 的行列式：$\det(\mathbf{X}^\top\mathbf{X}) = n\sum X_i^2 - n^2\bar{X}^2 = n(\sum X_i^2 - n\bar{X}^2) = n \cdot S_{XX}$

$(\mathbf{X}^\top\mathbf{X})^{-1} = \frac{1}{n \cdot S_{XX}} \begin{pmatrix} \sum X_i^2 & -n\bar{X} \\ -n\bar{X} & n \end{pmatrix}$

由 $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$：

$$\hat{\beta}_1 = \frac{1}{n \cdot S_{XX}} \left[-n\bar{X} \cdot n\bar{Y} + n \sum X_i Y_i \right] = \frac{n(\sum X_i Y_i - n\bar{X}\bar{Y})}{n \cdot S_{XX}} = \frac{S_{XY}}{S_{XX}}$$

由正規方程式第一列：$n\hat{\beta}_0 + n\bar{X}\hat{\beta}_1 = n\bar{Y} \implies \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$。 □

推論 3.1 (簡單迴歸估計量的變異數)**.**

$$\mathrm{var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{XX}}, \qquad \mathrm{var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} \right)$$

證明　由 $\mathrm{var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$：

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{n \cdot S_{XX}} \begin{pmatrix} \sum X_i^2 & -n\bar{X} \\ -n\bar{X} & n \end{pmatrix}$$

$\mathrm{var}(\hat{\beta}_1) = \frac{\sigma^2 \cdot n}{n \cdot S_{XX}} = \frac{\sigma^2}{S_{XX}}$

$\mathrm{var}(\hat{\beta}_0) = \frac{\sigma^2 \sum X_i^2}{n \cdot S_{XX}} = \frac{\sigma^2(\sum X_i^2 - n\bar{X}^2 + n\bar{X}^2)}{n \cdot S_{XX}} = \frac{\sigma^2 S_{XX}}{n \cdot S_{XX}} + \frac{\sigma^2 n\bar{X}^2}{n \cdot S_{XX}} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} \right)$ □

4　判定係數與 ANOVA

4.1　平方和分解

定理 4.1 (變異數分解)**.**

$$\boxed{\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}}$$

其中：

- $\mathrm{SST} = \mathbf{Y}^\top\mathbf{M}_0\mathbf{Y} = \sum(Y_i - \bar{Y})^2$（總平方和, Total Sum of Squares），$\mathbf{M}_0 = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$
- $\mathrm{SSR} = \hat{\mathbf{Y}}^\top\mathbf{M}_0\hat{\mathbf{Y}} = \sum(\hat{Y}_i - \bar{Y})^2$（迴歸平方和, Regression Sum of Squares）
- $\mathrm{SSE} = \mathbf{e}^\top\mathbf{e} = \sum(Y_i - \hat{Y}_i)^2$（殘差平方和, Error Sum of Squares）

證明 $\mathbf{Y} - \bar{Y}\mathbf{1} = (\mathbf{Y} - \hat{\mathbf{Y}}) + (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) = \mathbf{e} + (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1})$

平方展開：$\text{SST} = \mathbf{e}^\top\mathbf{e} + (\hat{\mathbf{Y}} - \bar{Y}\mathbf{1})^\top(\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) + 2\mathbf{e}^\top(\hat{\mathbf{Y}} - \bar{Y}\mathbf{1})$

交叉項：$\mathbf{e}^\top(\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}) = \mathbf{e}^\top\hat{\mathbf{Y}} - \bar{Y}\mathbf{e}^\top\mathbf{1} = 0 - \bar{Y} \cdot 0 = 0$（由正規方程式）

故 $\text{SST} = \text{SSE} + \text{SSR}$。 □

4.2 判定係數

定義 4.1 (判定係數 R^2)**.**

$$\boxed{R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}}$$

R^2 表示自變數能解釋依變數總變異的比例，$0 \leqslant R^2 \leqslant 1$。

定義 4.2 (調整後判定係數)**.**

$$\boxed{\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2) = 1 - \frac{\text{SSE}/(n-k-1)}{\text{SST}/(n-1)} = 1 - \frac{\text{MSE}}{s_Y^2}}$$

其中 $s_Y^2 = \text{SST}/(n-1)$ 為 Y 的樣本變異數。

註 (自由度的來源)**.** 調整後 R^2 的分子分母分別除以相應的自由度：

- **分母** $n-1$：$\text{SST} = \sum(Y_i - \bar{Y})^2$ 有 n 項，但受到一個約束 $\sum(Y_i - \bar{Y}) = 0$，故自由度為 $n-1$。
- **分子** $n-k-1$：$\text{SSE} = \sum e_i^2$ 有 n 項，但受到 $k+1$ 個約束（正規方程式 $\mathbf{X}^\top\mathbf{e} = \mathbf{0}$ 有 $k+1$ 個方程），故自由度為 $n-(k+1) = n-k-1$。

調整後 R^2 對增加的自變數進行懲罰。若增加不顯著變數，$\bar{R}^2$ 可能下降。$\bar{R}^2$ 可為負。

定理 4.2 (簡單迴歸中 $R^2 = r^2$)**.** 在簡單線性迴歸（$k=1$）中，$R^2 = r_{XY}^2$，其中 r_{XY} 為 X 與 Y 的樣本相關係數。

證明 $R^2 = \text{SSR}/\text{SST} = \hat{\beta}_1^2 S_{XX}/S_{YY} = (S_{XY}/S_{XX})^2 \cdot S_{XX}/S_{YY} = S_{XY}^2/(S_{XX}S_{YY}) = r_{XY}^2$ □

5 假設檢定

5.1 抽樣分配

定理 5.1 (OLS 估計量的抽樣分配)**.** 在假設 (A1)–(A5) 下：

$$\hat{\boldsymbol{\beta}}|\mathbf{X} \sim N\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\right)$$

證明 由 $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}$，$\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\varepsilon}$ 的線性變換。

在假設 (A5) 下，$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$。

常態分配的線性變換仍為常態：若 $\mathbf{z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，則 $\mathbf{A}\mathbf{z} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$。

取 $\mathbf{A} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$，$\mathbf{z} = \boldsymbol{\varepsilon}$，$\mathbf{b} = \boldsymbol{\beta}$：

$$\begin{aligned}\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon} &\sim N\left(\boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \cdot \mathbf{0}, (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\right) \\ &= N\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\right)\end{aligned}$$

□

定理 5.2 (SSE/σ^2 的分配)**.** 在假設 (A1)–(A5) 下：

$$\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-k-1}$$

且 SSE 與 $\hat{\boldsymbol{\beta}}$ 獨立。

證明　步驟一：標準化

令 $\mathbf{z} = \boldsymbol{\varepsilon}/\sigma$，則 $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_n)$（標準常態向量）。

由 $\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon}$：

$$\frac{\text{SSE}}{\sigma^2} = \frac{\boldsymbol{\varepsilon}^\top \mathbf{M}\boldsymbol{\varepsilon}}{\sigma^2} = \mathbf{z}^\top \mathbf{M}\mathbf{z}$$

步驟二：二次型的分配

由定理 2.2，若 $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_n)$ 且 $\mathbf{A}$ 為對稱冪等矩陣，則 $\mathbf{z}^\top \mathbf{A}\mathbf{z} \sim \chi^2_r$，其中 $r = \text{tr}(\mathbf{A})$。

$\mathbf{M}$ 為對稱冪等矩陣，$\text{tr}(\mathbf{M}) = n - k - 1$，故：

$$\frac{\text{SSE}}{\sigma^2} = \mathbf{z}^\top \mathbf{M}\mathbf{z} \sim \chi^2_{n-k-1}$$

步驟三：獨立性

$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}$ 與 $\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon}$ 的獨立性：

兩者皆為 $\boldsymbol{\varepsilon}$ 的線性變換。由常態分配的性質，獨立性等價於共變異數為零：

$$\begin{aligned}
\text{cov}\left((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}, \mathbf{M}\boldsymbol{\varepsilon}\right) &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \text{var}(\boldsymbol{\varepsilon})\mathbf{M}^\top \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\sigma^2\mathbf{I})\mathbf{M} \\
&= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{M} = \mathbf{0}
\end{aligned}$$

（最後一步利用 $\mathbf{X}^\top\mathbf{M} = (\mathbf{M}\mathbf{X})^\top = \mathbf{0}^\top = \mathbf{0}$。） □

5.2 整體 F 檢定

表 1: 多元迴歸 ANOVA 表

來源	SS	df	MS	F
迴歸	SSR	k	$\text{MSR} = \text{SSR}/k$	$F = \text{MSR}/\text{MSE}$
殘差	SSE	$n-k-1$	$\text{MSE} = \text{SSE}/(n-k-1)$	
總和	SST	$n-1$		

定理 5.3 (整體 F 檢定)**.** 檢定 $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$：

$$\boxed{F = \frac{\text{MSR}}{\text{MSE}} = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{k,n-k-1}} \quad \text{（在 } H_0 \text{ 下）}$$

證明 F **統計量的公式：**

$F = \frac{\text{SSR}/k}{\text{SSE}/(n-k-1)}$。由 $R^2 = \text{SSR}/\text{SST}$ 和 $1 - R^2 = \text{SSE}/\text{SST}$：

$$F = \frac{R^2 \cdot \text{SST}/k}{(1-R^2)\cdot \text{SST}/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

F **分配的推導：**

在 H_0 下，可以證明 $\text{SSR}/\sigma^2 \sim \chi^2_k$ 且與 SSE 獨立。由定理 5.2，$\text{SSE}/\sigma^2 \sim \chi^2_{n-k-1}$。

F 分配的定義：若 $U \sim \chi^2_m$，$V \sim \chi^2_n$ 獨立，則 $(U/m)/(V/n) \sim F_{m,n}$。

故：

$$F = \frac{(\text{SSR}/\sigma^2)/k}{(\text{SSE}/\sigma^2)/(n-k-1)} = \frac{\text{SSR}/k}{\text{SSE}/(n-k-1)} \sim F_{k,n-k-1}$$

□

5.3 個別係數 t 檢定

定理 5.4 (個別係數檢定). 對於 $H_0 : \beta_j = 0$：

$$\boxed{t = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)} \sim t_{n-k-1}}$$

其中 $\text{SE}(\hat{\beta}_j) = s\sqrt{[(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj}}$，$s = \sqrt{\text{MSE}}$。

證明 由定理 5.1，$\hat{\beta}_j \sim N(\beta_j, \sigma^2[(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj})$。

標準化：$Z = \frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{[(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj}}} \sim N(0,1)$

由定理 5.2，$\text{SSE}/\sigma^2 \sim \chi^2_{n-k-1}$，且與 $\hat{\beta}_j$ 獨立。

t 分配的定義：若 $Z \sim N(0,1)$，$V \sim \chi^2_\nu$ 獨立，則 $Z/\sqrt{V/\nu} \sim t_\nu$。

令 $V = \text{SSE}/\sigma^2$，$\nu = n-k-1$：

$$\begin{aligned} t = \frac{Z}{\sqrt{V/\nu}} &= \frac{\frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{[(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj}}}}{\sqrt{\frac{\text{SSE}/\sigma^2}{n-k-1}}} \\ &= \frac{\hat{\beta}_j - \beta_j}{\sqrt{\frac{\text{SSE}}{n-k-1}} \cdot \sqrt{[(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj}}} \\ &= \frac{\hat{\beta}_j - \beta_j}{s\sqrt{[(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj}}} \sim t_{n-k-1} \end{aligned}$$

在 $H_0 : \beta_j = 0$ 下，$t = \hat{\beta}_j/\text{SE}(\hat{\beta}_j)$。 □

定理 5.5 (簡單迴歸中 $t^2 = F$). 在簡單線性迴歸（$k=1$）中，檢定 $H_0 : \beta_1 = 0$ 時：

$$t^2 = F$$

證明 $t = \hat{\beta}_1/(s/\sqrt{S_{XX}})$，故 $t^2 = \hat{\beta}_1^2 S_{XX}/s^2 = \hat{\beta}_1^2 S_{XX}/\text{MSE} = \text{SSR}/\text{MSE} = F$ □

6 進階主題

6.1 含交互作用項的模型

定義 6.1 (交互作用). 模型 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$ 中，X_1 對 Y 的**邊際效應**為：

$$\frac{\partial Y}{\partial X_1} = \beta_1 + \beta_3 X_2$$

此效應取決於 X_2 的值——這就是「交互作用」的含義。

例題 6.1 (廣告與定價的交互作用). 模型：$Y = 20 + 2.5X_1 - 1.2X_2 + 0.04X_1X_2$（$Y$: 銷售，$X_1$: 廣告，$X_2$: 定價）

廣告的邊際效應 $= 2.5 + 0.04X_2$。當定價 X_2 越高，廣告效果越強。

6.2 無截距模型

定義 6.2 (無截距迴歸——多元形式). 考慮模型 $Y_i = \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i$（無截距項），矩陣形式為 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$，其中 $\mathbf{X}$ 為 $n \times k$ 矩陣（不含全 1 的行）。

OLS 估計量仍為 $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y}$。

性質 6.1 (無截距模型的特性).
- (i) $\sum_{i=1}^{n} e_i$ 不一定為零（因為 $\mathbf{X}$ 不包含截距項，正規方程式不含 $\sum e_i = 0$）
- (ii) $\mathbf{X}^\top \mathbf{e} = \mathbf{0}$ 仍成立（這是正規方程式的直接結果）
- (iii) Centered-$R^2 = 1 - \text{SSE}/\text{SST}$ 可能為負，因為 $\text{SST} + \text{SSE}$ 不一定等於 SSR
- (iv) 使用 Uncentered-$R^2 = 1 - \text{SSE}/\sum Y_i^2$ 較為適當

例題 6.2 (CAPM 模型). 資本資產定價模型（Capital Asset Pricing Model, CAPM）的實證形式為：

$$R_{it} - R_f = \beta_i (R_{mt} - R_f) + \varepsilon_{it}$$

其中 $R_{it} - R_f$ 為資產 i 的超額報酬，$R_{mt} - R_f$ 為市場超額報酬。這是典型的無截距迴歸。

若加入截距項（Jensen's alpha）：$R_{it} - R_f = \alpha_i + \beta_i(R_{mt} - R_f) + \varepsilon_{it}$，則 $\alpha_i \neq 0$ 表示該資產的表現優於（$\alpha > 0$）或劣於（$\alpha < 0$）市場基準。

7 計算範例

例題 7.1. 某分析師研究家庭垃圾產量的決定因素。資料：$n = 440$，$k = 3$（房屋大小、兒童數、成人數），$R^2 = 0.170$，$F = 29.80$。

- (a) 計算調整後 R^2
- (b) 計算殘差自由度
- (c) 進行整體 F 檢定（$\alpha = 0.05$）

解答. (a) 調整後 R^2：

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2) = 1 - \frac{439}{436}(0.830) = 1 - 1.0069 \times 0.830 = 1 - 0.836 = \boxed{0.164}$$

(b) 殘差自由度：$df_{\text{殘差}} = n - k - 1 = 440 - 3 - 1 = \boxed{436}$

(c) F 檢定：

驗算：$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0.170/3}{0.830/436} = \frac{0.0567}{0.00190} = 29.8$ ✓

臨界值：$F_{0.05,3,436} \approx 2.63$

決策：$29.80 > 2.63$，**拒絕** H_0。模型整體顯著。

例題 7.2. $n = 100$，$k = 4$，$R^2 = 0.36$。

解答. $F = \frac{0.36/4}{0.64/95} = \frac{0.09}{0.00674} = 13.35$

$\bar{R}^2 = 1 - \frac{99}{95}(0.64) = 1 - 0.667 = 0.333$

例題 7.3. Y 對 X 迴歸得 $\hat{Y} = 2 + 0.5X$，$R^2 = 0.64$。求 X 對 Y 迴歸的斜率。

解答. $r^2 = R^2 = 0.64$，$\hat{\beta}_1 = 0.5 > 0$，故 $r = 0.8$。

由 $\hat{\beta}_1 \cdot \hat{\gamma}_1 = r^2$：$\hat{\gamma}_1 = 0.64/0.5 = 1.28$

8 習題

習題 8.1. 已知 SST $= 100$，SSE $= 36$，$n = 12$，$k = 1$。求 R^2、$\bar{R}^2$、F。

解答. $R^2 = 1 - 36/100 = 0.64$

$\bar{R}^2 = 1 - \frac{11}{10}(0.36) = 1 - 0.396 = 0.604$

$F = \frac{0.64/1}{0.36/10} = \frac{0.64}{0.036} = 17.78$

習題 8.2. 迴歸輸出：$n = 200$，$k = 3$，$R^2 = 0.72$，截距 $\hat{\beta}_0 = 7.19$，SE$(\hat{\beta}_0) = 1.09$。

(a) 計算 $\bar{R}^2$

(b) 計算截距的 t 值

(c) 在 $\alpha = 0.05$ 下檢定整體模型顯著性

解答. (a) $\bar{R}^2 = 1 - \frac{199}{196}(0.28) = 1 - 0.284 = 0.716$

(b) $t = 7.19/1.09 = 6.59$

(c) $F = \frac{0.72/3}{0.28/196} = \frac{0.24}{0.00143} = 168$

$F_{0.05,3,196} \approx 2.65$。$168 > 2.65$，拒絕 H_0，模型整體顯著。