統計學講義

第五部分：假設檢定

參考書籍

Jeffrey S. Rosenthal, *Probability and Statistics: The Science of Uncertainty*, 2nd Edition
Chapter 8: Hypothesis Testing; Chapter 9: Further Topics

https://utstat.utoronto.ca/mikevans/jeffrosenthal/

1 假設檢定的基本概念

1.1 統計假設

定義 1.1 (統計假設). **統計假設**（statistical hypothesis）是關於母體參數或母體分配的陳述。

- **虛無假設**（null hypothesis），記為 H_0：欲檢驗的假設，通常代表「無效果」、「無差異」或現狀，且包含等號。
- **對立假設**（alternative hypothesis），記為 H_1 或 H_a：與 H_0 相對的假設，通常代表研究者想要證明的主張。

例題 1.1 (假設的設定).
- 檢驗藥物是否有效：H_0：藥物無效 vs H_1：藥物有效
- 檢驗產品平均重量：$H_0 : \mu = 500$ vs $H_1 : \mu \neq 500$
- 檢驗新製程是否提高良率：$H_0 : p \leqslant 0.9$ vs $H_1 : p > 0.9$

1.2 檢定類型

定義 1.2 (單尾與雙尾檢定). 設母體參數為 θ，檢定值為 θ_0：

(i) **雙尾檢定**（two-tailed test）：
$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

(ii) **右尾檢定**（right-tailed test）：
$$H_0 : \theta \leqslant \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

(iii) **左尾檢定**（left-tailed test）：
$$H_0 : \theta \geqslant \theta_0 \quad \text{vs} \quad H_1 : \theta < \theta_0$$

註. 在實務中，H_0 常簡寫為 $\theta = \theta_0$，即使對立假設是單尾的。這是因為檢定統計量的分配通常在 $\theta = \theta_0$ 時計算。

1.3 檢定程序

性質 1.1 (假設檢定的步驟).
1. **建立假設**：根據問題設定 H_0 和 H_1
2. **選擇顯著水準**：決定 α（常用 0.05、0.01、0.10）
3. **計算檢定統計量**：根據樣本資料計算適當的統計量
4. **決定拒絕域或計算 p 值**
5. **做出結論**：拒絕或不拒絕 H_0

2 檢定錯誤與檢定力

2.1 兩類錯誤

	H_0 為真	H_0 為假
不拒絕 H_0	正確決策	**型二錯誤**（β）
拒絕 H_0	**型一錯誤**（α）	正確決策（檢定力）

定義 2.1 (型一錯誤與型二錯誤)**.**
- **型一錯誤**（Type I Error）：H_0 為真卻拒絕 H_0（偽陽性；拒絕不該拒絕的）

$$\alpha = \mathsf{P}(\text{拒絕}H_0 \mid H_0 \text{ 為真})$$

  α 稱為**顯著水準**（significance level）。
- **型二錯誤**（Type II Error）：H_0 為假卻不拒絕 H_0（偽陰性；接受不該接受的）

$$\beta = \mathsf{P}(\text{不拒絕}H_0 \mid H_0 \text{ 為假})$$

註.
- 型一錯誤與型二錯誤是**互相牽制**的：在固定樣本量下，降低 α 通常會增加 β。
- 一般優先控制型一錯誤（固定 α），因為型一錯誤的後果通常較嚴重（如錯誤地宣稱新藥有效）。
- 增加樣本量可以同時降低 α 和 β。

2.2 檢定力

定義 2.2 (檢定力)**. 檢定力**（power）是當 H_0 為假時，正確拒絕 H_0 的機率：

$$\text{Power} = 1 - \beta = \mathsf{P}(\text{拒絕}H_0 \mid H_0 \text{ 為假})$$

定理 2.1 (檢定力的影響因素)**.** 檢定力會隨以下因素而增加：

(i) 效應量（effect size）增加：真實參數值離 H_0 假設值越遠
(ii) 樣本量 n 增加
(iii) 顯著水準 α 增加
(iv) 母體變異數 σ^2 減少

2.3 型二錯誤機率的計算

性質 2.1 (計算 β 的步驟)**.**
1. 在 H_0 下，找出**不拒絕域**（接受域）
2. 假設真實參數值為 θ_1（H_1 下的某個特定值）
3. 在 $\theta = \theta_1$ 下，計算統計量落在不拒絕域的機率

例題 2.1. 設 $X_1, \ldots, X_{25} \overset{\text{i.i.d.}}{\sim} N(\mu, 100)$（即 $\sigma = 10$）。檢定 $H_0 : \mu = 50$ vs $H_1 : \mu > 50$，顯著水準 $\alpha = 0.05$。若真實 $\mu = 54$，求型二錯誤機率 β。

解答. 步驟 1：在 $H_0 : \mu = 50$ 下找不拒絕域

檢定統計量：$Z = \dfrac{\bar{X} - 50}{10/\sqrt{25}} = \dfrac{\bar{X} - 50}{2}$

右尾檢定，拒絕域：$Z > z_{0.05} = 1.645$

即 $\bar{X} > 50 + 1.645 \times 2 = 53.29$ 時拒絕 H_0

不拒絕域：$\bar{X} \leqslant 53.29$

步驟 2：在真實 $\mu = 54$ 下計算 β

當 $\mu = 54$ 時，$\bar{X} \sim N(54, 4)$（標準差 2）

$$\begin{aligned}
\beta &= \mathsf{P}(\bar{X} \leqslant 53.29 \mid \mu = 54) \\
&= P\left(Z \leqslant \frac{53.29 - 54}{2}\right) \\
&= \mathsf{P}(Z \leqslant -0.355) \\
&= \Phi(-0.355) \approx 0.361
\end{aligned}$$

結論：型二錯誤機率約為 36.1%，檢定力 $= 1 - 0.361 = 0.639$（約 64%）。

例題 2.2. 延續例 2.1，畫出檢定力函數 Power(μ)，並求使檢定力達到 0.90 所需的樣本量。

解答. 檢定力函數：

對於任意 $\mu > 50$：

$$\begin{aligned}
\text{Power}(\mu) &= \mathsf{P}(\bar{X} > 53.29 \mid \mu) \\
&= P\left(Z > \frac{53.29 - \mu}{2}\right) \\
&= 1 - \Phi\left(\frac{53.29 - \mu}{2}\right)
\end{aligned}$$

當 $\mu = 54$：Power $= 1 - \Phi(-0.355) = 0.639$

當 $\mu = 56$：Power $= 1 - \Phi(-1.355) = 0.912$

求所需樣本量（使 Power $= 0.90$ 當 $\mu = 54$）：

設樣本量為 n，標準誤 $= 10/\sqrt{n}$

拒絕域臨界值：$c = 50 + 1.645 \times \dfrac{10}{\sqrt{n}}$

$$\text{Power} = P\left(Z > \frac{c - 54}{10/\sqrt{n}}\right) = 0.90$$

需要 $\dfrac{c - 54}{10/\sqrt{n}} = -z_{0.10} = -1.282$

代入 $c = 50 + 1.645 \times \dfrac{10}{\sqrt{n}}$：

$$\begin{aligned}
\frac{50 + 1.645 \times \frac{10}{\sqrt{n}} - 54}{10/\sqrt{n}} &= -1.282 \\
\frac{-4 + \frac{16.45}{\sqrt{n}}}{10/\sqrt{n}} &= -1.282 \\
\frac{-4\sqrt{n} + 16.45}{10} &= -1.282
\end{aligned}$$

$$
\begin{aligned}
-4\sqrt{n} + 16.45 &= -12.82 \\
\sqrt{n} &= \frac{16.45 + 12.82}{4} = 7.32 \\
n &= 53.6
\end{aligned}
$$

故至少需要 $n = 54$ 個樣本。

3 p 值

3.1 p 值的定義

定義 3.1 (p 值)**. p 值**（p-value）是在 H_0 為真的假設下，觀察到與當前樣本統計量**一樣極端或更極端**的結果之機率。

- 右尾檢定：$\text{p-value} = \mathsf{P}(T \geqslant t_{\text{obs}} \mid H_0)$
- 左尾檢定：$\text{p-value} = \mathsf{P}(T \leqslant t_{\text{obs}} \mid H_0)$
- 雙尾檢定：$\text{p-value} = 2 \times \mathsf{P}(T \geqslant |t_{\text{obs}}| \mid H_0)$

其中 t_{obs} 是觀察到的檢定統計量值。

性質 3.1 (p 值的決策規則)**.**
- 若 $\text{p-value} < \alpha$，則拒絕 H_0
- 若 $\text{p-value} \geqslant \alpha$，則不拒絕 H_0

定理 3.1 (p 值與拒絕域的等價性)**.** $\text{p-value} < \alpha \iff$ 檢定統計量落在拒絕域內。

證明. 以右尾檢定為例。設檢定統計量為 T，臨界值為 c_α（滿足 $\mathsf{P}(T > c_\alpha \mid H_0) = \alpha$）。

觀察到 t_{obs}，則：

$$
\begin{aligned}
\text{p-value} < \alpha &\iff \mathsf{P}(T \geqslant t_{\text{obs}} \mid H_0) < \alpha \\
&\iff \mathsf{P}(T \geqslant t_{\text{obs}} \mid H_0) < \mathsf{P}(T > c_\alpha \mid H_0) \\
&\iff t_{\text{obs}} > c_\alpha \\
&\iff t_{\text{obs}} \text{ 落在拒絕域內}
\end{aligned}
$$

3.2 p 值的解釋

註 (p 值的正確與錯誤解釋)**. 正確解釋：**

- p 值是在 H_0 為真時，觀察到目前結果（或更極端結果）的機率
- p 值越小，反對 H_0 的證據越強

錯誤解釋（常見誤解）：

- **錯**：p 值是 H_0 為真的機率
- **錯**：p 值是結果由隨機造成的機率
- **錯**：$1 - \text{p-value}$ 是效應存在的機率

例題 3.1. 設 $X_1, \ldots, X_{16} \overset{\text{i.i.d.}}{\sim} N(\mu, 64)$（$\sigma = 8$）。樣本平均 $\bar{x} = 53$。檢定 $H_0 : \mu = 50$ vs $H_1 : \mu \neq 50$。求 p 值並在 $\alpha = 0.05$ 下做結論。

解答. 檢定統計量：

$$
z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{53 - 50}{8/\sqrt{16}} = \frac{3}{2} = 1.5
$$

雙尾檢定的 p 值：

$$\text{p-value} = 2 \times \mathsf{P}(Z \geqslant 1.5) = 2 \times (1 - \Phi(1.5)) = 2 \times 0.0668 = 0.1336$$

由於 $\text{p-value} = 0.1336 > 0.05 = \alpha$，不拒絕 H_0。

結論：在 $\alpha = 0.05$ 下，沒有足夠證據說明 $\mu \neq 50$。

4 單樣本平均數檢定

4.1 z 檢定（σ 已知）

定理 4.1 (單樣本 z 檢定)**.** 設 $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$，$\sigma^2$ 已知。檢定 $H_0 : \mu = \mu_0$。

檢定統計量：

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1) \quad （在 H_0 下）$$

拒絕域（顯著水準 α）：

- 雙尾（$H_1 : \mu \neq \mu_0$）：$|Z| > z_{\alpha/2}$
- 右尾（$H_1 : \mu > \mu_0$）：$Z > z_\alpha$
- 左尾（$H_1 : \mu < \mu_0$）：$Z < -z_\alpha$

說明 (為何 Z 服從標準常態？)**.** 由於 $X_i \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$，樣本平均數 $\bar{X} \sim N(\mu, \sigma^2/n)$。標準化後：

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

在 $H_0 : \mu = \mu_0$ 下，以 μ_0 代入 μ，得到檢定統計量。

4.2 t 檢定（σ 未知）

定理 4.2 (單樣本 t 檢定)**.** 設 $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$，$\sigma^2$ 未知。檢定 $H_0 : \mu = \mu_0$。

檢定統計量：

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \quad （在 H_0 下）$$

拒絕域（顯著水準 α）：

- 雙尾（$H_1 : \mu \neq \mu_0$）：$|T| > t_{\alpha/2, n-1}$
- 右尾（$H_1 : \mu > \mu_0$）：$T > t_{\alpha, n-1}$
- 左尾（$H_1 : \mu < \mu_0$）：$T < -t_{\alpha, n-1}$

說明 (為何 T 服從 t 分配？)**.** 將 z 檢定中的 σ 替換成樣本標準差 S。由 t 分配的定義：

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{S/\sigma} = \frac{Z}{\sqrt{(n-1)S^2/\sigma^2/(n-1)}} = \frac{Z}{\sqrt{V/(n-1)}}$$

其中 $Z \sim N(0, 1)$，$V = (n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$，且 Z 與 V 獨立（常態母體下 $\bar{X}$ 與 S^2 獨立），故 $T \sim t_{n-1}$。

例題 4.1. 某便利商店宣稱每日庫存量平均為 500 件。品管人員隨機抽查 25 天，得平均庫存量 485 件，樣本標準差 30 件。在 $\alpha = 0.05$ 下，檢定庫存量是否低於宣稱值。

解答. 步驟 1：建立假設

$H_0 : \mu \geqslant 500$ vs $H_1 : \mu < 500$（左尾檢定）

步驟 2：計算檢定統計量

σ 未知，使用 t 檢定：

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{485 - 500}{30/\sqrt{25}} = \frac{-15}{6} = -2.5$$

步驟 3：確定拒絕域

$df = 24$，$\alpha = 0.05$，左尾檢定

$t_{0.05,24} = 1.711$，拒絕域：$t < -1.711$

步驟 4：結論

$t = -2.5 < -1.711$，落在拒絕域內，拒絕 H_0。

p 值計算：p-value $= \mathsf{P}(T_{24} < -2.5) \approx 0.01$

結論：在 $\alpha = 0.05$ 下，有足夠證據顯示庫存量低於宣稱的 500 件。

5　雙樣本平均數檢定

5.1　獨立樣本 t 檢定

定理 5.1 (獨立樣本 t 檢定(假設變異數相等)). 設兩獨立樣本 $X_1, \ldots, X_{n_1} \overset{\text{i.i.d.}}{\sim} N(\mu_1, \sigma^2)$ 與 $Y_1, \ldots, Y_{n_2} \overset{\text{i.i.d.}}{\sim} N(\mu_2, \sigma^2)$。

檢定 $H_0 : \mu_1 - \mu_2 = \delta_0$（通常 $\delta_0 = 0$）。

合併變異數估計：

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

檢定統計量：

$$T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2} \quad \text{（在 } H_0 \text{ 下）}$$

證明. 由第三部分的抽樣分配結論，當 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 時：

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

且 $\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$。由 t 分配的定義，將 σ 用 S_p 估計後得到 $t_{n_1+n_2-2}$ 分配。

例題 5.1. 比較兩種教學法。A 法：$n_1 = 12$，$\bar{x}_1 = 78$，$s_1 = 8$；B 法：$n_2 = 15$，$\bar{x}_2 = 72$，$s_2 = 10$。假設成績服從常態分配且變異數相等。在 $\alpha = 0.05$ 下，檢定兩種教學法是否有顯著差異。

解答. 假設：$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$

合併變異數：

$$S_p^2 = \frac{11 \times 64 + 14 \times 100}{12 + 15 - 2} = \frac{704 + 1400}{25} = \frac{2104}{25} = 84.16$$

$S_p = 9.17$

檢定統計量：

$$t = \frac{78-72}{9.17\sqrt{\frac{1}{12}+\frac{1}{15}}} = \frac{6}{9.17 \times 0.391} = \frac{6}{3.59} = 1.67$$

臨界值：$df = 25$，$t_{0.025,25} = 2.060$

結論：$|t| = 1.67 < 2.060$，不拒絕 H_0。

在 $\alpha = 0.05$ 下，沒有足夠證據顯示兩種教學法有顯著差異。

5.2 配對樣本 t 檢定

定義 5.1 (配對樣本). 當兩組觀測值存在自然配對關係時，應使用**配對樣本 t 檢定**：

- 同一受試者的前後測量（before-after）
- 雙胞胎研究
- 配對比較實驗

定理 5.2 (配對樣本 t 檢定). 設配對觀測值為 $(X_1, Y_1), \ldots, (X_n, Y_n)$，令差異 $D_i = X_i - Y_i$。

假設 $D_i \overset{\text{i.i.d.}}{\sim} N(\mu_D, \sigma_D^2)$。檢定 $H_0 : \mu_D = 0$（即 $\mu_1 = \mu_2$）。

檢定統計量：

$$T = \frac{\bar{D} - 0}{S_D/\sqrt{n}} \sim t_{n-1} \quad \text{（在 } H_0 \text{ 下）}$$

其中 $\bar{D} = \dfrac{1}{n}\sum_{i=1}^n D_i$，$S_D^2 = \dfrac{1}{n-1}\sum_{i=1}^n (D_i - \bar{D})^2$。

註. 配對樣本 t 檢定本質上是對「差異」進行單樣本 t 檢定，自由度為 $n-1$（配對數減 1）。

例題 5.2. 8 輛汽車分別使用舊引擎和新引擎，測量燃油效率（km/L）如下：

車輛	1	2	3	4	5	6	7	8
新引擎	12	15	14	13	16	14	15	13
舊引擎	10	13	12	12	14	12	13	11
差異 D	2	2	2	1	2	2	2	2

在 $\alpha = 0.01$ 下，檢定新引擎是否顯著提高燃油效率。

解答. 假設：$H_0 : \mu_D \leqslant 0$ vs $H_1 : \mu_D > 0$（右尾檢定）

計算差異的統計量：

$$\bar{D} = \frac{2+2+2+1+2+2+2+2}{8} = \frac{15}{8} = 1.875$$

$$S_D^2 = \frac{\sum(D_i - \bar{D})^2}{7} = \frac{(0.125)^2 \times 7 + (0.875)^2}{7} = \frac{0.109 + 0.766}{7} = \frac{0.875}{7} = 0.125$$

$S_D = 0.354$

檢定統計量：

$$t = \frac{1.875 - 0}{0.354/\sqrt{8}} = \frac{1.875}{0.125} = 15.0$$

臨界值：$df = 7$，$t_{0.01,7} = 2.998$

結論：$t = 15.0 > 2.998$，拒絕 H_0。

有非常強的證據顯示新引擎顯著提高燃油效率（p-value < 0.001）。

6 比例檢定

6.1 單一比例檢定

定理 6.1 (單一比例 z 檢定)**.** 設 X 為 n 次獨立伯努利試驗中成功的次數，$X \sim B(n,p)$。

檢定 $H_0 : p = p_0$。

檢定統計量（n 夠大時）：

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{\text{approx}}{\sim} N(0,1) \quad \text{（在 } H_0 \text{ 下）}$$

其中 $\hat{p} = X/n$。

註. 注意：檢定統計量的分母使用 H_0 下的 p_0，而非樣本估計值 $\hat{p}$。這與信賴區間不同。

例題 6.1. 某候選人宣稱其支持率為 50%。隨機調查 400 位選民，有 220 人支持。在 $\alpha = 0.05$ 下，檢定支持率是否顯著高於 50%。

解答. 假設：$H_0 : p \leqslant 0.5$ vs $H_1 : p > 0.5$

樣本比例：$\hat{p} = 220/400 = 0.55$

檢定統計量：

$$z = \frac{0.55 - 0.50}{\sqrt{\frac{0.5 \times 0.5}{400}}} = \frac{0.05}{0.025} = 2.0$$

臨界值：$z_{0.05} = 1.645$

結論：$z = 2.0 > 1.645$，拒絕 H_0。

p-value $= \mathsf{P}(Z > 2.0) = 0.0228$

在 $\alpha = 0.05$ 下，有足夠證據顯示支持率高於 50%。

6.2 兩母體比例差檢定

定理 6.2 (兩母體比例差檢定)**.** 設兩獨立樣本的樣本比例為 $\hat{p}_1 = X_1/n_1$ 和 $\hat{p}_2 = X_2/n_2$。

檢定 $H_0 : p_1 = p_2$。

合併比例估計：

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

檢定統計量：

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \overset{\text{approx}}{\sim} N(0,1) \quad \text{（在 } H_0 \text{ 下）}$$

7 變異數檢定

7.1 單一變異數檢定

定理 7.1 (單一變異數的 χ^2 檢定). 設 $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$。檢定 $H_0 : \sigma^2 = \sigma_0^2$。

檢定統計量：

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1} \quad \text{（在 } H_0 \text{ 下）}$$

拒絕域：

- 雙尾（$H_1 : \sigma^2 \neq \sigma_0^2$）：$\chi^2 < \chi^2_{1-\alpha/2,n-1}$ 或 $\chi^2 > \chi^2_{\alpha/2,n-1}$
- 右尾（$H_1 : \sigma^2 > \sigma_0^2$）：$\chi^2 > \chi^2_{\alpha,n-1}$
- 左尾（$H_1 : \sigma^2 < \sigma_0^2$）：$\chi^2 < \chi^2_{1-\alpha,n-1}$

說明 (為何使用卡方分配？). 由第三部分定理，對常態母體：$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$。在 $H_0 : \sigma^2 = \sigma_0^2$ 下，以 σ_0^2 代入即得檢定統計量。

7.2 兩變異數比檢定

定理 7.2 (兩變異數比的 F 檢定). 設兩獨立常態樣本的樣本變異數為 S_1^2 和 S_2^2。檢定 $H_0 : \sigma_1^2 = \sigma_2^2$。

檢定統計量：

$$F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1} \quad \text{（在 } H_0 \text{ 下）}$$

拒絕域（雙尾，α）：$F < F_{1-\alpha/2,n_1-1,n_2-1}$ 或 $F > F_{\alpha/2,n_1-1,n_2-1}$

利用 $F_{1-\alpha/2,k_1,k_2} = \dfrac{1}{F_{\alpha/2,k_2,k_1}}$ 簡化查表。

說明 (為何使用 F 分配？). 由 F 分配定義：若 $U \sim \chi^2_{k_1}$ 與 $V \sim \chi^2_{k_2}$ 獨立，則 $(U/k_1)/(V/k_2) \sim F_{k_1,k_2}$。

因為 $(n_1-1)S_1^2/\sigma_1^2 \sim \chi^2_{n_1-1}$ 且 $(n_2-1)S_2^2/\sigma_2^2 \sim \chi^2_{n_2-1}$ 獨立，在 $H_0 : \sigma_1^2 = \sigma_2^2$ 下：

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$$

例題 7.1. 兩組獨立常態樣本：$n_1 = 10$，$s_1^2 = 25$；$n_2 = 8$，$s_2^2 = 10$。在 $\alpha = 0.10$ 下，檢定兩母體變異數是否相等。

解答. 假設：$H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$

檢定統計量：

$$F = \frac{s_1^2}{s_2^2} = \frac{25}{10} = 2.5$$

臨界值：$df_1 = 9$，$df_2 = 7$

$F_{0.05,9,7} = 3.68$（右尾）

$F_{0.95,9,7} = 1/F_{0.05,7,9} = 1/3.29 = 0.304$（左尾）

拒絕域：$F < 0.304$ 或 $F > 3.68$

結論：$F = 2.5$ 不在拒絕域內，不拒絕 H_0。

在 $\alpha = 0.10$ 下，沒有足夠證據顯示兩母體變異數不相等。

8　卡方獨立性檢定

8.1　列聯表與獨立性

定義 8.1 (列聯表). **列聯表**（contingency table）用於呈現兩個類別變數的交叉分類次數。

設有 r 個列類別和 c 個行類別，O_{ij} 表示第 i 列第 j 行的觀察次數。

定義 8.2 (獨立性). 若兩類別變數獨立，則聯合機率等於邊際機率的乘積：

$$\mathsf{P}(X=i, Y=j) = \mathsf{P}(X=i)\cdot\mathsf{P}(Y=j)$$

8.2　卡方獨立性檢定

定理 8.1 (卡方獨立性檢定). 檢定 H_0：兩變數獨立 vs H_1：兩變數不獨立。

期望次數：在 H_0 下，

$$E_{ij} = \frac{(\text{第}i\text{ 列總和})\times(\text{第}j\text{ 行總和})}{n} = \frac{R_i\times C_j}{n}$$

檢定統計量：

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \overset{\text{approx}}{\sim} \chi^2_{(r-1)(c-1)} \quad \text{（在 } H_0 \text{ 下）}$$

拒絕域：$\chi^2 > \chi^2_{\alpha,(r-1)(c-1)}$

證明 (證明概要). 在 H_0（獨立）下，$E_{ij} = n\cdot p_{i\cdot}\cdot p_{\cdot j}$，其中 $p_{i\cdot}$ 和 $p_{\cdot j}$ 是邊際機率。

估計這些機率後（用 R_i/n 和 C_j/n），自由度減少為 $(r-1)(c-1)$：

- 原有 rc 個格子
- 估計 $(r-1)$ 個列邊際機率和 $(c-1)$ 個行邊際機率
- 加上總和 $=1$ 的限制
- $df = rc-1-(r-1)-(c-1) = (r-1)(c-1)$

註. 卡方檢定的適用條件：所有期望次數 $E_{ij}\geqslant 5$。若不滿足，可合併類別或使用 Fisher 精確檢定。

例題 8.1. 某公司調查員工工作級別與年終獎金等級的關係：

	低獎金	中獎金	高獎金	列總和
初級	30	40	10	80
中級	20	50	30	100
高級	10	30	80	120
行總和	60	120	120	300

在 $\alpha = 0.05$ 下，檢定工作級別與年終獎金是否獨立。

解答. **假設**：H_0：獨立 vs H_1：不獨立

計算期望次數：$E_{ij} = R_i\times C_j/300$

	低獎金	中獎金	高獎金
初級	$80\times 60/300 = 16$	$80\times 120/300 = 32$	$80\times 120/300 = 32$
中級	$100\times 60/300 = 20$	$100\times 120/300 = 40$	$100\times 120/300 = 40$
高級	$120\times 60/300 = 24$	$120\times 120/300 = 48$	$120\times 120/300 = 48$

檢定統計量：

$$\begin{aligned}
\chi^2 &= \frac{(30-16)^2}{16} + \frac{(40-32)^2}{32} + \frac{(10-32)^2}{32} \\
&\quad + \frac{(20-20)^2}{20} + \frac{(50-40)^2}{40} + \frac{(30-40)^2}{40} \\
&\quad + \frac{(10-24)^2}{24} + \frac{(30-48)^2}{48} + \frac{(80-48)^2}{48} \\
&= \frac{196}{16} + \frac{64}{32} + \frac{484}{32} + 0 + \frac{100}{40} + \frac{100}{40} \\
&\quad + \frac{196}{24} + \frac{324}{48} + \frac{1024}{48} \\
&= 12.25 + 2 + 15.125 + 0 + 2.5 + 2.5 + 8.17 + 6.75 + 21.33 \\
&= 70.625
\end{aligned}$$

臨界值：$df = (3-1)(3-1) = 4$，$\chi^2_{0.05,4} = 9.488$

結論：$\chi^2 = 70.625 > 9.488$，強烈拒絕 H_0。

有非常強的證據顯示工作級別與年終獎金不獨立（即有關聯）。

9 信賴區間與假設檢定的關係

定理 9.1 (雙尾檢定與信賴區間的等價性). 對於雙尾檢定 $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$：

在顯著水準 α 下拒絕 $H_0 \iff \theta_0$ 不在 $(1-\alpha)$ 信賴區間內

證明. 以 μ 的 t 檢定為例。

$(1-\alpha)$ 信賴區間為 $\bar{X} \pm t_{\alpha/2,n-1} \cdot S/\sqrt{n}$

θ_0 不在區間內 $\iff |\bar{X} - \theta_0| > t_{\alpha/2,n-1} \cdot S/\sqrt{n}$

$$\iff \left|\frac{\bar{X} - \theta_0}{S/\sqrt{n}}\right| > t_{\alpha/2,n-1}$$

$\iff$ 拒絕 H_0

例題 9.1. 例 4.1 中，95% 信賴區間為 $485 \pm 2.064 \times 6 = (472.62, 497.38)$。

$\mu_0 = 500$ 不在此區間內，故在 $\alpha = 0.05$ 下拒絕 $H_0 : \mu = 500$。

但本題是左尾檢定，所以需要計算單側信賴區間或直接用 p 值判斷。

10 本章練習題

練習題 10.1. 設 $X_1, \dots, X_{36} \overset{\text{i.i.d.}}{\sim} N(\mu, 36)$（即 $\sigma = 6$）。檢定 $H_0 : \mu = 20$ vs $H_1 : \mu < 20$，$\alpha = 0.05$。

(a) 求拒絕域
(b) 若真實 $\mu = 18$，求型二錯誤機率
(c) 若要使檢定力達到 0.80（當 $\mu = 18$），需要多大樣本？

解答. (a) 檢定統計量：$Z = \dfrac{\bar{X} - 20}{6/\sqrt{36}} = \dfrac{\bar{X} - 20}{1}$

左尾檢定，$z_{0.05} = 1.645$

拒絕域：$Z < -1.645$，即 $\bar{X} < 20 - 1.645 = 18.355$

(b) 當 $\mu = 18$ 時，$\bar{X} \sim N(18, 1)$
$\beta = \mathsf{P}(\bar{X} \geqslant 18.355 \mid \mu = 18) = \mathsf{P}(Z \geqslant 0.355) = 1 - 0.639 = 0.361$
(c) 設樣本量 n，標準誤 $= 6/\sqrt{n}$
臨界值：$c = 20 - 1.645 \times 6/\sqrt{n}$
檢定力 $= \mathsf{P}(\bar{X} < c \mid \mu = 18) = 0.80$
$P\left(Z < \dfrac{c-18}{6/\sqrt{n}}\right) = 0.80$
$\dfrac{c-18}{6/\sqrt{n}} = z_{0.20} = 0.842$
代入 c：$\dfrac{20 - 1.645 \times 6/\sqrt{n} - 18}{6/\sqrt{n}} = 0.842$
$\dfrac{2 - 9.87/\sqrt{n}}{6/\sqrt{n}} = 0.842$
$\dfrac{2\sqrt{n} - 9.87}{6} = 0.842$
$2\sqrt{n} = 5.052 + 9.87 = 14.922$
$\sqrt{n} = 7.46$，$n = 55.7$
需要至少 56 個樣本。

練習題 10.2. 某研究比較兩種藥物的效果。若採用配對設計（每位病人同時接受兩種藥物），樣本量為 n；若採用獨立設計，兩組各 n 人。假設兩設計的母體標準差相同。

(a) 配對設計的自由度為何？
(b) 獨立設計的自由度為何？
(c) 哪種設計通常有較高的檢定力？為什麼？

解答. (a) 配對設計：$df = n - 1$
(b) 獨立設計：$df = 2n - 2$
(c) **配對設計通常有較高檢定力**，原因：
- 配對設計消除了個體間差異的變異
- 差異 $D_i = X_i - Y_i$ 的變異數通常小於 $\mathrm{var}(\bar{X} - \bar{Y})$
- 雖然自由度較小，但標準誤的減少通常更顯著

定量比較：若 $\mathrm{corr}(X, Y) = \rho$，則

$$\mathrm{var}(D) = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$$

當 $\rho > 0$ 時，$\mathrm{var}(D) < \sigma_X^2 + \sigma_Y^2$，配對設計更有效。

練習題 10.3. 某製造商宣稱產品不良率不超過 5%。檢驗 200 件產品，發現 15 件不良品。在 $\alpha = 0.05$ 下，是否有足夠證據拒絕製造商的宣稱？

解答. 假設：$H_0 : p \leqslant 0.05$ vs $H_1 : p > 0.05$

樣本比例：$\hat{p} = 15/200 = 0.075$

檢定統計量：

$$z = \frac{0.075 - 0.05}{\sqrt{\frac{0.05 \times 0.95}{200}}} = \frac{0.025}{\sqrt{0.0002375}} = \frac{0.025}{0.0154} = 1.62$$

臨界值：$z_{0.05} = 1.645$

結論：$z = 1.62 < 1.645$，不拒絕 H_0。

p-value $= \mathsf{P}(Z > 1.62) = 0.0526 > 0.05$

在 $\alpha = 0.05$ 下，沒有足夠證據拒絕製造商的宣稱（但結果接近顯著邊界）。

練習題 10.4. 擲一枚骰子 120 次，各點數出現次數如下：

點數	1	2	3	4	5	6
次數	25	17	15	23	24	16

在 $\alpha = 0.05$ 下，檢定骰子是否公正。

解答. 假設：H_0：骰子公正（各點機率 $= 1/6$）vs H_1：骰子不公正

期望次數：$E_i = 120 \times 1/6 = 20$（每個點數）

檢定統計量：

$$\begin{aligned}
\chi^2 &= \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} \\
&= \frac{(25-20)^2}{20} + \frac{(17-20)^2}{20} + \frac{(15-20)^2}{20} \\
&\quad + \frac{(23-20)^2}{20} + \frac{(24-20)^2}{20} + \frac{(16-20)^2}{20} \\
&= \frac{25+9+25+9+16+16}{20} = \frac{100}{20} = 5
\end{aligned}$$

臨界值：$df = 6 - 1 = 5$，$\chi^2_{0.05,5} = 11.07$

結論：$\chi^2 = 5 < 11.07$，不拒絕 H_0。

沒有足夠證據顯示骰子不公正。

練習題 10.5. 設母體服從 $N(\mu, 25)$。以 $n = 100$ 的樣本檢定 $H_0 : \mu = 10$ vs $H_1 : \mu \neq 10$，$\alpha = 0.05$。求當真實 $\mu = 11$ 時的檢定力。

解答. 拒絕域：$|Z| > 1.96$，其中 $Z = \dfrac{\bar{X} - 10}{5/\sqrt{100}} = \dfrac{\bar{X} - 10}{0.5}$

拒絕 H_0 當 $\bar{X} < 10 - 1.96 \times 0.5 = 9.02$ 或 $\bar{X} > 10 + 1.96 \times 0.5 = 10.98$

當 $\mu = 11$ **時**，$\bar{X} \sim N(11, 0.25)$

$$\begin{aligned}
\text{Power} &= \mathsf{P}(\bar{X} < 9.02 \mid \mu = 11) + \mathsf{P}(\bar{X} > 10.98 \mid \mu = 11) \\
&= P\left(Z < \frac{9.02 - 11}{0.5}\right) + P\left(Z > \frac{10.98 - 11}{0.5}\right) \\
&= \mathsf{P}(Z < -3.96) + \mathsf{P}(Z > -0.04) \\
&\approx 0 + 0.516 = 0.516
\end{aligned}$$

當 $\mu = 11$ 時，檢定力約為 51.6%。

（若只考慮右尾，Power $\approx \mathsf{P}(Z > -0.04) \approx 0.516$）