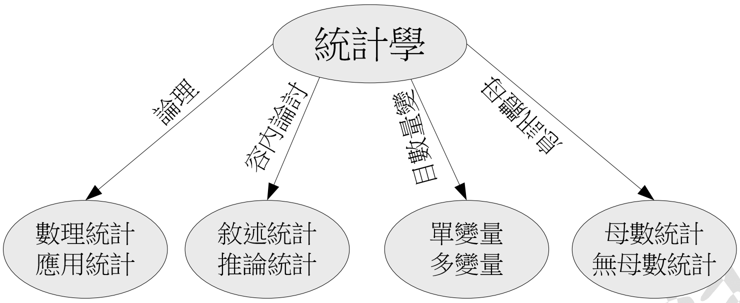
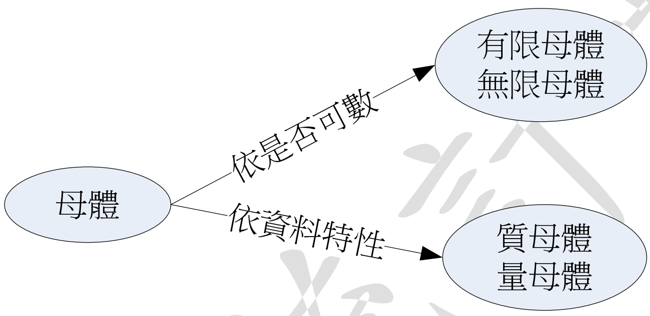
各單元重點公式整理

壹、統計概論

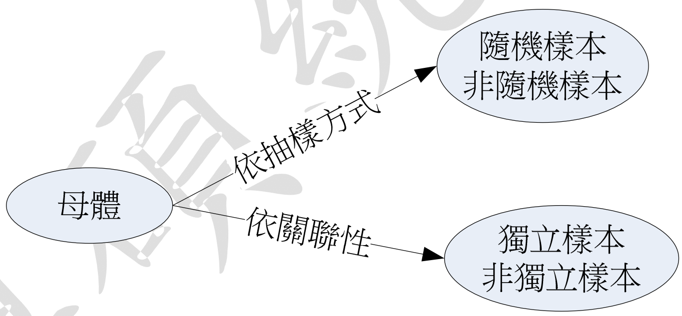
■ 統計學的分類



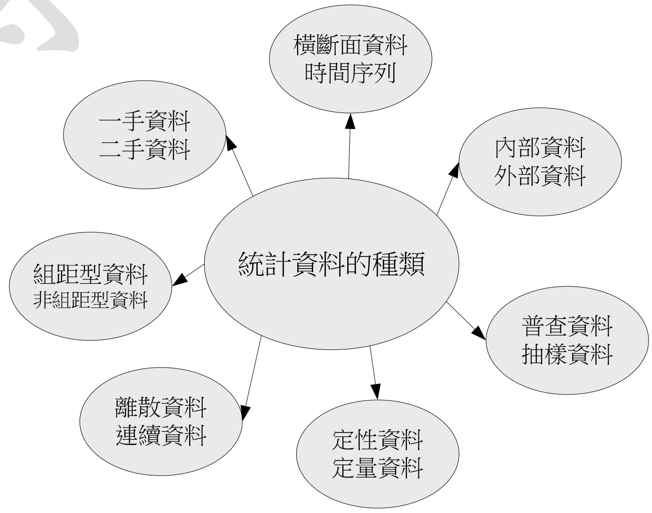
■ 母體的分類



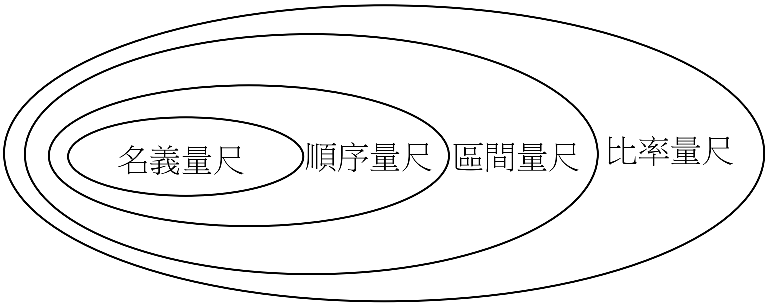
■ 樣本的分類



■ 資料的分類



■ 四種量尺之關係

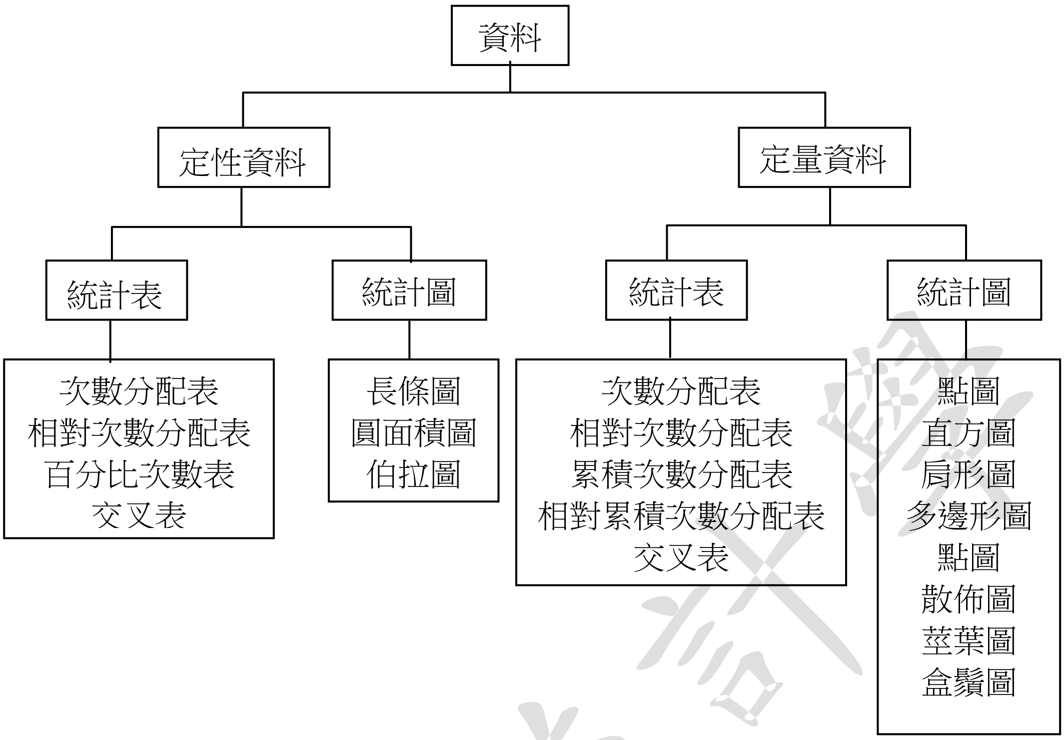


■ 各種量尺資料資與運算法之整理

種類	四則運算	統計量數	具固定零點
名義量尺	×	眾數	×
順序量尺	×	眾數、中位數	×
區間量尺	加減	眾數、中位數、平均數、標準差	×
比率量尺	加減乘除	眾數、中位數、平均數、標準差	✓

貳、常用的統計圖表

常用的統計圖與表總整理

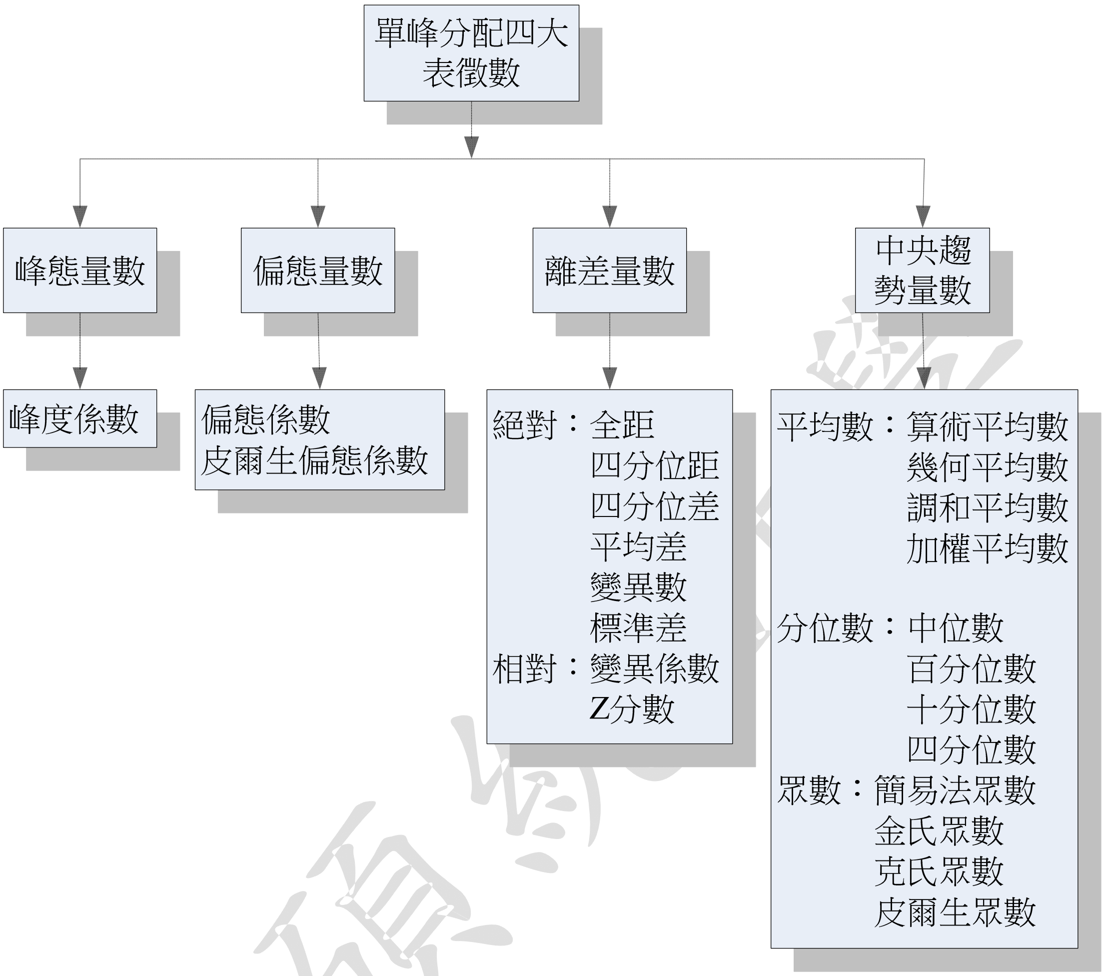


■ 史塔基法則 STURGE'S rule：

組數：$k = 1 + 3.322 \log n$

參、常用的統計量數

■ 單峰分配的四大表徵數



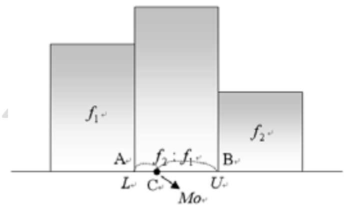
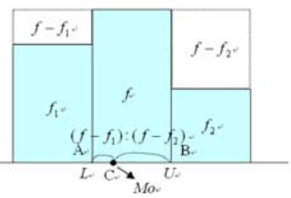
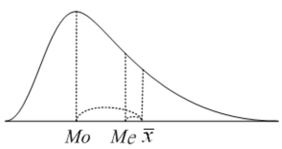
◆ 集中趨勢量數

名稱	資料型態	公式
算術平均數	非組距型	$\mu=\frac{1}{N}\sum_{i=1}^{N}x_i$ ， $\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$
	組距型	$\mu=\frac{1}{N}\sum_{i=1}^{k}f_i m_i$ ， $\overline{x}=\frac{1}{n}\sum_{i=1}^{k}f_i m_i$
加權平均數		$\frac{\sum_{i=1}^{k}w_i x_i}{\sum_{i=1}^{k}w_i}$
幾何平均數	非組距型	$G=\sqrt[n]{\prod_{i=1}^{n}x_i}$
	組距型	$G=\sqrt[n]{\prod_{i=1}^{n}x_i}$
調和平均數	非組距型	$H=\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$
	組距型	$H=\frac{n}{\sum_{i=1}^{k}\frac{f_k}{m_k}}$

◆ k 分位數

	非組距型資料		組距型資料
中位數 (Me)	$\frac{n}{2}\notin Z$	$Me=x_{\frac{1+n}{2}}$	利用內插公式
	$\frac{n}{2}\in Z$	$Me=\frac{x_{\frac{n}{2}}+x_{\frac{n}{2}+1}}{2}$	
四分位數 (Q_1,Q_2,Q_3)	$\frac{in}{4}\notin Z$	$Q_i=x_{\left(\left[\frac{in}{4}\right]+1\right)}$	
	$\frac{in}{4}\in Z$	$Q_i=\frac{x_{(\frac{in}{4})}+x_{(\frac{in}{4}+1)}}{2}$	
十分位數 $(D_1,D_2,\cdots,D_9)$	$\frac{in}{10}\notin Z$	$D_i=x_{\left(\left[\frac{in}{10}\right]+1\right)}$	
	$\frac{in}{10}\in Z$	$D_i=\frac{x_{(\frac{in}{10})}+x_{(\frac{in}{10}+1)}}{2}$	
百分位數 $(P_1,P_2,\cdots,P_{99})$	$\frac{in}{100}\notin Z$	$P_i=x_{\left(\left[\frac{in}{100}\right]+1\right)}$	
	$\frac{in}{100}\in Z$	$P_i=\frac{x_{(\frac{in}{100})}+x_{(\frac{in}{100}+1)}}{2}$	

◆ 眾數(mode)

金氏法		$(Mo-L):(U-Mo)=f_2:f_1$ $\Rightarrow Mo=L+\frac{U-L}{f_1+f_2}\times f_2$
克氏法		$(Mo-L):(U-Mo)$ $=(f-f_1):(f-f_2)$ $\Rightarrow Mo=L+\frac{U-L}{2f-f_1-f_2}\times(f-f_1)$
皮爾生經驗法則		$\bar{x}-Mo=3(\bar{x}-Me)$ $\Rightarrow Mo=\bar{x}-3(\bar{x}-Me)$

◆ 絕對離差量數

名稱	資料型態	公式
全距	非組距型	$R = x_{\max} - x_{\min}$
	組距型	$R = U_{\max} - L_{\min}$
四分位距		$IQR = Q_3 - Q_1$
四分位差		$Q.D. = \dfrac{Q_3 - Q_1}{2}$
平均差	非組距型	$\text{M.A.D} = \dfrac{1}{n}\sum_{i=1}^{n}\left\|x_i - \bar{x}\right\|$
	組距型	$\text{MAD} = \dfrac{1}{n}\sum_{i=1}^{k} f_i\left\|m_i - \bar{x}\right\|$
變異數	非組距型	$\sigma^2 = \dfrac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \dfrac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu^2$ $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}x_i^2 - \dfrac{n}{n-1}\bar{x}^2$
	組距型	$\sigma^2 = \dfrac{1}{N}\sum_{i=1}^{k} f_i(m_i - \mu)^2 = \dfrac{1}{N}\sum_{i=1}^{k} f_i m_i^2 - \mu^2$ $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{k} f_i(m_i - \bar{x})^2 = \dfrac{1}{n-1}\sum_{i=1}^{k} f_i m_i^2 - \dfrac{n}{n-1}\bar{x}^2$
標準差		$\sigma = \sqrt{\sigma^2}, s = \sqrt{s^2}$

◆ 相對離差量數

變異係數	$\text{CV} = \dfrac{\sigma}{\mu}\times 100\%$ 或 $\text{CV} = \dfrac{s}{\bar{x}}\times 100\%$
Z 分數	$Z_i = \dfrac{x_i - \bar{x}}{s}$ 或 $Z_i = \dfrac{x_i - \mu}{\sigma}$

◆ 偏態量數(skewness)

- 偏態係數

$$\beta_1 = \frac{M_3}{\sigma^3}$$

- Pearson 偏態係數

$$SK = \frac{\mu - Mo}{\sigma} = \frac{3(\mu - Me)}{\sigma}$$

◆ 峰態量數(kurtosis)

- 峰態係數

$$\beta_2 = \frac{m_4}{s^4}$$

■ 動差(moment)

◆ r 階原動差

$$m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,r} \qquad , r = 1,2,3,\cdots$$

◆ r 階主動差

$$M_r = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^r \qquad , r = 1,2,3,\cdots$$

◆ 主動差與原動差之關係

$$M_1 = 0$$
$$M_2 = m_2 - (m_1)^2$$
$$M_3 = m_3 - 3m_2m_1 + 2(m_1)^3$$
$$M_4 = m_4 - 4m_3m_1 + 6m_2(m_1)^2 - 3(m_1)^4$$

■ Chebyshev 不等式

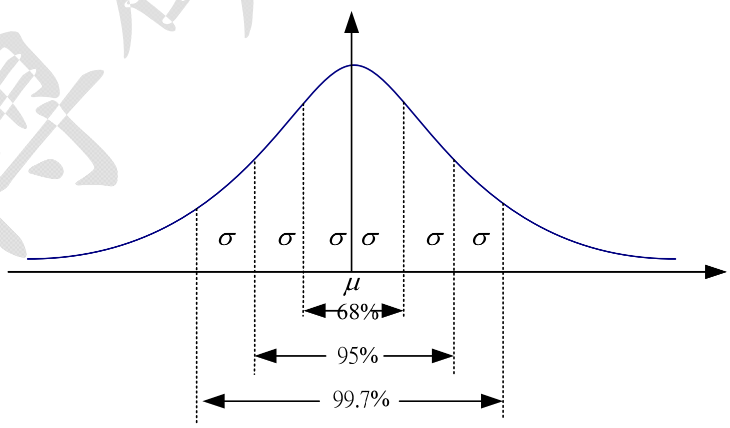
$$P(|X-\mu| \le k\sigma) \ge 1 - \frac{1}{k^2}, k > 1$$

◆ Chebyshev 之變形式

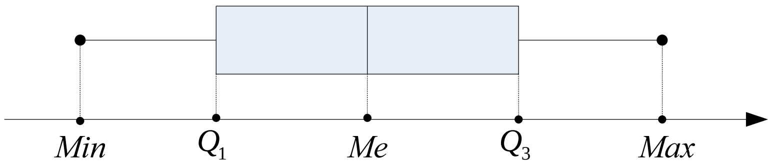
$$P(|\bar{x}-\mu| \le k\sqrt{\frac{\sigma^2}{n}}) \ge 1 - \frac{1}{k^2}, k > 1$$
$$P(|\hat{p}-p| \le k\sqrt{\frac{pq}{n}}) \ge 1 - \frac{1}{k^2}, k > 1$$

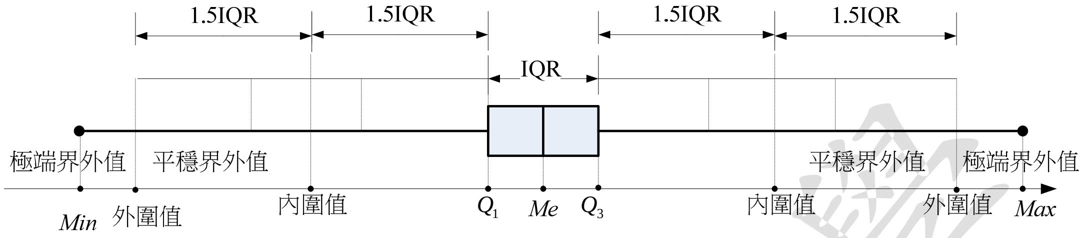
■ 經驗法則

約有 68%的觀測值落於$(\bar{x}-s, \bar{x}+s)$的區間內。
約有 95%的觀測值落於$(\bar{x}-2s, \bar{x}+2s)$的區間內。
約有 99.7%的觀測值落於$(\bar{x}-3s, \bar{x}+3s)$的區間內。



■ 盒鬚圖





■ 共變異數

$$\sigma_{xy}=\frac{\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)}{N}=\frac{\sum_{i=1}^{N}x_iy_i-N\mu_x\cdot\mu_y}{N}$$

$$s_{xy}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{n-1}=\frac{\sum_{i=1}^{n}x_iy_i-n\overline{x}\cdot\overline{y}}{n-1}$$

■ 皮爾森相關係數

$$\rho_{xy}=\frac{\sigma_{xy}}{\sigma_x\sigma_y}\text{，}r_{xy}=\frac{s_{xy}}{s_xs_y}$$

可以衡量兩變數間的直線相關程度

肆、機率論

■ 集合的基本概念

名稱	定義
聯集	$A\cup B=\{x \mid x\in A\vee x\in B\}$
交集	$A\cap B=\{x \mid x\in A\wedge x\in B\}$
餘集合	$A^C=\{x \mid x\in S\wedge x\notin A\}=S-A$
差集	$A-B=\{x \mid x\in A\wedge x\notin B\}$ $=A-(A\cap B)=(A\cup B)-B$

■ 集合基本運算法則

交換律	$A\cup B=B\cup A$，$A\cap B=B\cap A$
結合律	$(A\cup B)\cup C=A\cup(B\cup C)$ $(A\cap B)\cap C=A\cap(B\cap C)$
分配律	$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$ $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$
排容原理	$A'=S-A$
互補律	$(A')'=A$
De'Morgan's Law	$A'\cap B'=(A\cup B)'=S-(A\cup B)$ $A'\cup B'=(A\cap B)'=S-(A\cap B)$

■ 集合元素的計數

◆ $n(A\cup B)=n(A)+n(B)-n(A\cap B)$

◆ $n(A\cup B\cup C)=n(A)+n(B)+n(C)-n(A\cap B)-n(B\cap C)-n(C\cap A)$
$+n(A\cap B\cap C)$

◆ $n(A-B)=n(A)-n(A\cap B)$

◆ $n(A')=n(S)-n(A)$

◆ De'Morgan's Law

$$n(A'\cap B')=n(A\cup B)'=n(S)-n(A\cup B)$$
$$n(A'\cup B')=n(A\cap B)'=n(S)-n(A\cap B)$$

■ 條件機率

◆ 定義

$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$

◆ 性質

$P(A'|C)=1-P(A|C)$

$P(A\cup B|C)=P(A|C)+P(B|C)-P(A\cap B|C)$）

$P(A\cap B)=P(A)P(B|A)$）

$P(A\cap B\cap C)=P(A)P(B|A)P(C|A\cap B)$）

■ 乘法原理

$P(A\cap B)=P(A)P(B|A)$

$P(A\cap B\cap C)=P(A)P(B|A)P(C|A\cap B)$

$P(A\cap B\cap C\cap\cdots\cap Y\cap Z)=P(A)P(B|A)P(C|A\cap B)\cdots P(Z|A\cap B\cap\cdots\cap Y)$

■ 獨立與互斥

◆ 兩事件獨立

$$P(A|B)=P(A) \text{ 或 } P(A\cap B)=P(A)\cdot P(B)$$

◆ 三事件獨立

A,B,C 兩兩獨立

$P(A\cap B\cap C)=P(A)\cdot P(B)\cdot P(C)$

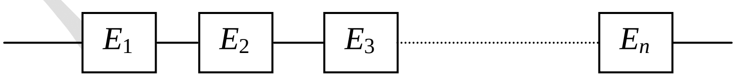
◆ 互斥事件

$$P(A\cap B)=0$$

若 A、B 獨立則 A'與 B'，A'與 B，A 與 B'皆獨立

■ 串聯與並聯系統

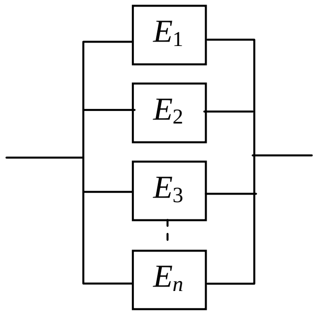
◆ 串聯系統(series)



$$P(E_1\cap E_2\cap\cdots\cap E_n)=\prod_{i=1}^{n}P(E_i)$$

◆ 並聯系統

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = 1 - \prod_{i=1}^{n} P(E_i')$$



■ 聯合機率

$$P(A_i \cap B_j) \geq 0, \forall i = 1,2,\ldots,r;\ j = 1,2,\ldots,c$$

$$\sum_{j=1}^{c}\sum_{i=1}^{r} P(A_i \cap B_j) = 1$$

■ 邊際機率

$$P(A_i) = \sum_{j=1}^{c} P(A_i \cap B_j) \quad ,i = 1,2,\ldots,r$$

$$P(B_j) = \sum_{i=1}^{r} P(A_i \cap B_j) \quad ,j = 1,2,\ldots,c$$

A \ B	B_1	B_2	…	B_r	A 的邊際機率
A_1	$P(A_1 \cap B_1)$	$P(A_1 \cap B_2)$	…	$P(A_1 \cap B_c)$	$P(A_1)$
A_2	$P(A_2 \cap B_1)$	$P(A_2 \cap B_2)$	…	$P(A_2 \cap B_c)$	$P(A_2)$
⋮	⋮	⋮	⋮	⋮	⋮
A_r	$P(A_r \cap B_1)$	$P(A_r \cap B_2)$	…	$P(A_r \cap B_c)$	$P(A_r)$
B 的邊際機率	$P(B_1)$	$P(B_2)$	…	$P(B_r)$	總和=1

A_2 的邊際機率

B_1 的邊際機率

■ 全機率定理

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)$$

■ 貝士定理

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k) \cdot P(B|A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)}$$

五、機率分配

- 機率分配
  - 離散型隨機變數

    $0 \le f(x), \quad -\infty < x < \infty$

    $$\int_{-\infty}^{\infty} f(x)dx = 1$$

  - 連續隨機變數的機率分配

    $0 \le f(x), \quad -\infty < x < \infty$

    $$\int_{-\infty}^{\infty} f(x)dx = 1$$

- 累積分配函數
  - 離散型隨機變數

    $$F(x) = f(X \le x) = \sum_{X \le x} f(X)$$

  - 連續隨機變數的機率分配

    $$F(x) = f(X \le x) = \int_{-\infty}^{x} f(x)dx$$

■ 機率分配的重要參數

名稱	離散機變數	連續機變數
期望值	$E(x)=\sum_x xf(x)$ $E(g(x))=\sum_x g(x)f(x)$	$E(x)=\int_x xf(x)dx$ $E(g(x))=\int_x g(x)f(x)dx$
中位數	$f(x\le\eta)=\sum_{x=-\infty}^{\eta} f(x)=\frac{1}{2}$ $F(\eta)=\frac{1}{2}$	$P(X\le\eta)=\int_{-\infty}^{\eta} f(x)dx=\frac{1}{2}$ $F(\eta)=\frac{1}{2}$
分位數	$P(X\le x_p)=\sum_{x=-\infty}^{x_p} f(x)=p$ $F(x_p)=p$	$P(X\le x_p)=\int_{-\infty}^{x_p} f(x)dx=p$ $F(x_p)=p$
眾數	兩面逼近法 $f(x)\ge f(x+1)$ 且 $f(x)\ge f(x-1)$	$f'(m)=0$且$f''(m)<0$
變異數	$V(x)=E[(x-E(x))^2]$ $=\sum_x (x-\mu)^2 f(x)$ $V(x)=E(x^2)-[E(x)]^2$	$V(x)=E\left[(x-E(x))^2\right]$ $=\int_x (x-\mu)^2 f(x)dx$ $V(x)=E(x^2)-[E(x)]^2$
偏態係數	$\beta_1=E[(\frac{x-\mu}{\sigma})^3]$	$\beta_1=E[(\frac{x-\mu}{\sigma})^3]$
峰態係數	$\beta_2=E[(\frac{x-\mu}{\sigma})^4]$	$\beta_2=E[(\frac{x-\mu}{\sigma})^4]$

■ 期望值的性質

- $E(c)=c$，c 為常數
- $E\left[c\cdot g(x)\right]=c\cdot E\left[g(x)\right]$
- $E\left[c+g(x)\right]=c+E\left[g(x)\right]$
- $E\left[c_1g_1(x)+c_2g_2(x)\right]=c_1E\left[g_1(x)\right]+c_2E\left[g_2(x)\right]$，此關係稱為線性運算關係。
- 若 $y=ax+b$，則 $E(y)=aE(x)+b$
- 期望值不一定存在(Petersbury 反論)

■ 變異數的性質

- $V(x)\ge 0$
- $V(C)=0$
- $V(ax\pm b)=a^2V(x)$

陸、二元隨機變數

■ 聯合機率函數

◆ 離散型隨機變數

$0 \le f(x_i, y_i) \le 1$

$\sum_x \sum_y f(x, y) = 1$

◆ 連續型隨機變數

$0 \le f(x_i, y_i)$

$$\int_x \int_y f(x, y) dy dx = 1$$

■ 邊際機率函數

X 的邊際機率分配	離散型	$f_X(x) = \sum_y f(x, y)$
	連續型	$f_X(x) = \int_y f(x, y) dy$
Y 的邊際機率分配	離散型	$f_Y(y) = \sum_x f(x, y)$
	連續型	$f_Y(y) = \int_x f(x, y) dx$

■ 邊際機率與聯合機率分配之關係

◆ 離散型隨機變數

$$\sum_x \sum_y f(x, y) = \sum_x f_X(x) = \sum_y f_Y(y)$$

◆ 連續型隨機變數

$$\int_x \int_y f(x, y) dy dx = \int_x f_X(x) dx = \int_y f_Y(y) dy$$

■ 累積分配函數

◆ 離散型隨機變數

$$F(x,y)=f(X\le x,Y\le y)=\sum_{x=-\infty}^{x}\sum_{y=-\infty}^{y}f(x,y)$$

◆ 連續型隨機變數

$$F(x,y)=f(X\le x,Y\le y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(x,y)dxdy$$

■ 條件機率分配函數與獨立性

$$f(x|y)=\frac{f(x,y)}{f_Y(y)},\quad f(y)\ne 0$$

$$f(y|x)=\frac{f(x,y)}{f_X(x)},\quad f(x)\ne 0$$

隨機變數 X 與 Y 互為獨立

$$f(x|y)=f_X(x)\text{ 或 }f(y|x)=f_Y(y)$$

$$\Rightarrow f(x,y)=f_X(x)f_Y(y)$$

■ 聯合機率分配函數的重要參數

期望值	離散型	$E(x)=\sum_x\sum_y xf(x,y)=\sum_x xf_X(x)$ $E(Y)=\sum_x\sum_y yf(x,y)=\sum_y yf_Y(y)$
	連續型	$E(x)=\int_x\int_y xf(x,y)dxdy=\int_x xf_X(x)dx$ $E(y)=\int_x\int_y yf(x,y)dxdy=\int_y yf_Y(y)dy$
變異數	$V(x)=E(x^2)-[E(x)]^2$ $V(y)=E(y^2)-[E(y)]^2$	
條件期望值	離散型	$E(x|y)=\sum_x xf(x|y)$ $E(y|x)=\sum_y yf(y|x)$
	連續型	$E(x|y)=\int_x xf(x|y)dx$ $E(y|x)=\int_y yf(y|x)dy$
條件變異數	$V(x|y)=E[x-E(x|y)]^2=E(x^2|y)-[E(x|y)]^2$ $V(y|x)=E[y-E(y|x)]^2=E(y^2|x)-[E(y|x)]^2$	

■ 期望值的性質

- $E(c \times g(x,y)) = c \times E(g(x,y)),\quad c$ 為常數。
- $E(c_1 g_1(x,y) + c_2 g_2(x,y)) = c_1 E(g_1(x,y)) + c_2 E(g_2(x,y))$
- $E(ax+by+c) = aE(x) + bE(y) + c$
- 若 X、Y 為兩獨立隨機變數，則 $E(xy) = E(x)E(y); E(\frac{x}{y}) = E(x)E(\frac{1}{y})$

■ 條件期望值的性質

- $E\left[(ax+by)\middle|z\right] = aE(x|z) + bE(y|z)$
- $E[E(yx)] = E(y), E\left[E(x|y)\right] = E(x)$

■ 共變異數

$$\begin{aligned} Cov(x,y) &= E\left[(x-E(x))(y-E(y))\right] \\ &= E(xy) - E(x)E(y) \end{aligned}$$

◆ 共變異數的性質

- $Cov(x,a) = 0, Cov(a,a) = 0$
- $Cov(x,x) = V(x)$
- $Cov(x \pm y, x) = Cov(x,z) \pm Cov(y,z)$
- $Cov(x+y, z-w) = Cov(x,z) - Cov(x,w) + Cov(y,z) - Cov(y,w)$
- $Cov(ax,y) = a \cdot Cov(x,y)$
- $Cov(ax,by) = ab \cdot Cov(x,y)$
- $V(ax \pm by) = a^2 V(x) \pm 2abCov(x,y) + b^2 V(y)$
- 若兩隨機變數 X,Y 獨立，則 $Cov(x,y) = 0$

■ 相關係數

$$\rho_{xy} = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}}$$

■ 獨立之判斷法則

- 兩事件獨立： $P(A|B) = A$ 或 $P(A \cap B) = P(A)P(B)$
- 兩隨機變數之獨立： $f(x|y) = f(x)$ 或 $f(x,y) = f(x)f(y)$
- 期望值之獨立： $E(x|y) = E(x)$ 或 $E(xy) = E(x)E(y)$
- 母體分配未知之兩變數獨立檢定：卡方獨立檢定

■ 馬可夫不等式

$$P(x \ge c\mu) \le \frac{1}{c} \text{ 或 } P(x \le c\mu) \ge 1 - \frac{1}{c}$$

■ 柴比雪夫不等式

$$P(|x-\mu| \le k\sigma) \ge 1-\frac{1}{k^2} \text{ 或 } P(|x-\mu| \ge k\sigma) \le \frac{1}{k^2}$$

◆ 柴比雪夫不等式的變形

- $P(|x-\mu| \le c) \ge 1-\frac{\sigma^2}{c^2}$
- $P(|\overline{x}-\mu| \le k\sqrt{\frac{\sigma^2}{n}}) \ge 1-\frac{1}{k^2}$

柒、(略)

捌、動差與動差母函數

- r 階原動差
  - 離散型隨機變數

$$E(x^r)=\sum_x x^r f(x)$$

  - 連續隨機變數

$$E(x^n)=\int_x x^n f(x)dx$$

- r 階主動差
  - 離散型隨機變數

$$E\left[(x-\mu)^r\right]=\sum_x (x-\mu)^r f(x)$$

  - 連續隨機變數

$$E\left[(x-\mu)^r\right]=\int_x (x-\mu)^r f(x)dx$$

- 階乘動差

$$E\left[x(x-1)(x-2)\cdots(x-r+1)\right]$$

- 動差母函數
  - 用途
    - 可用來求期望值、變異數、偏態與峰態係數。
    - 可判斷母體的機率函數為何種分配。
    - 隨機變數的轉換
  - 定義

$$M(t)=E(e^{tx})$$

  - 性質

$$E(x^r)=\frac{d^r}{dt^r}M(t)\Big|_{t=0}$$

  - 與期望值變異數之關係

$$\mu=E(x)=\frac{d}{dt}M(t)\Big|_{t=0}$$

$$V(x)=E(x^2)-\left[E(x)\right]^2\Rightarrow V(x)=\frac{d^2}{dt^2}M(t)\Big|_{t=0}-\left(\frac{d}{dt}M(t)\Big|_{t=0}\right)^2$$

    - 對數運算法

$$E(x)=\frac{d}{dt}\ln\left(M(t)\right)\Big|_{t=0}\text{，}V(x)=\frac{d^2}{dt^2}\ln\left(M(t)\right)\Big|_{t=0}$$

■ 機率母函數

$$P_X(t) = E[t^x]$$

◆ 性質

$$\left.\frac{d^n P_X(t)}{dt^n}\right|_{t=1} = E[x(x-1)\cdots(x-n+1)]$$

九、常用的機率分配

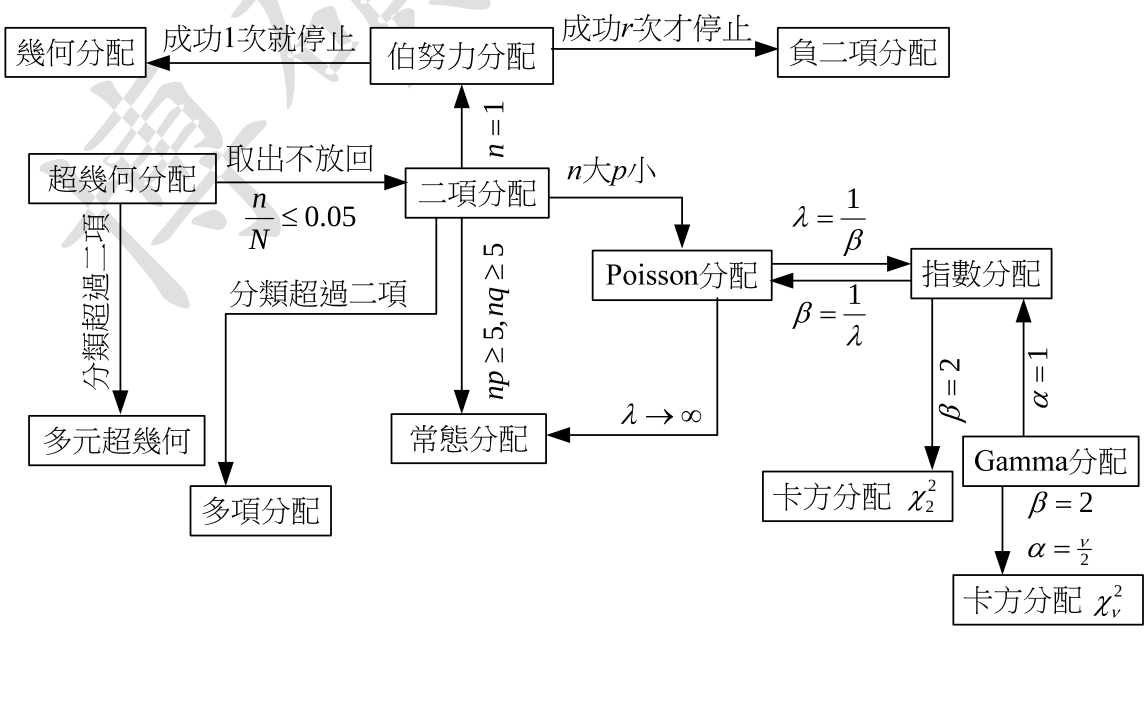
分配	機率函數	期望值	變異數	動差母函數	再生性	無記憶性
均勻	$f(x)=\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{e^t(1-e^{nt})}{n(1-e^t)}$	×	×
伯努力	$f(x)=p^xq^{1-x}$	p	pq	(pe^t+q)	×	×
二項	$f(x)=C_x^n p^x q^{n-x}$	np	npq	$(pe^t+q)^n$	○	×
多項	$f(x)=C^n_{x_1,x_2,\cdots,x_k} p_1^{x_1} p_2^{x_2}\cdots p_k^{x_k}$	np_i	np_iq_i	$(p_1e^{t_1}+p_2e^{t_2}+ \ldots + p_ke^{t_k})^n$	×	×
超幾何	$f(x)=\frac{C_x^S C_{n-x}^{N-S}}{C_n^N}$	np	$npq\cdot\frac{N-n}{N-1}$		×	×
幾何	$f(x)=q^{x-1}p$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^t}{1-qe^t}$	×	○
負二項	$f(x)=C_{r-1}^{x-1}p^rq^{x-r}$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$(\frac{pe^t}{1-qe^t})^r$	○	×
Poisson	$f(x)=\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	$e^{\lambda(e^t-1)}$	○	×
多維超幾何	$f(x_1,x_2,\cdots,x_k)=\frac{C_{x_1}^{N_1}C_{x_2}^{N_2}\cdots C_{x_k}^{N_k}}{C_n^N}$	np_i	$np_iq_i\cdot\frac{N-n}{N-1}$		×	×

十、常見的連續型機率分配

■　連續型機率函數的重要參數

分配	機率函數	期望值	變異數	動差母函數
均勻	$f(x)=\dfrac{1}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{bt}-e^{at}}{t(b-a)},t\neq 0$
常態	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$	μ	σ^2	$e^{\mu t+\frac{\sigma^2}{2}t^2}$
標準常態	$f(z)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$	0	1	$e^{\frac{t^2}{2}}$
指數	$f(x)=\dfrac{1}{\mu}e^{-\frac{x}{\mu}}$	μ	μ^2	$\dfrac{1}{1-\mu t}$
Gamma	$f(x)=\dfrac{1}{\beta\Gamma(\alpha)}(\dfrac{x}{\beta})^{\alpha-1}e^{-\frac{x}{\beta}}$	$\alpha\beta$	$\alpha\beta^2$	$(\dfrac{1}{1-\beta t})^{\alpha}$
卡方	$f(x)=\dfrac{1}{2\Gamma(\frac{\nu}{2})}(\dfrac{x}{2})^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}$	ν	2ν	$(\dfrac{1}{1-2t})^{\frac{\nu}{2}}$

■　各種分配之關係



■ 各種分配之加法性

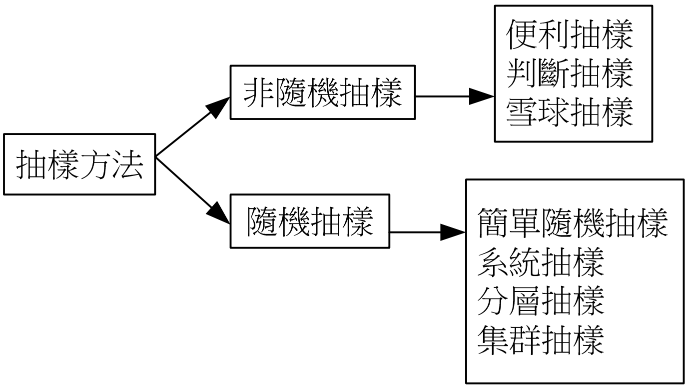
X_i 之分配	Y 之分配	限制條件	加法性
伯努力 $Ber(1,p)$	$B(n,p)$		無
二項 $B(n_i,p)$	$B(\sum_{i=1}^{n} n_i, p)$	p 相同	有
幾何 $G(p)$	$NB(n,p)$		無
負二項 $NB(n_i,p)$	$NB(\sum_{i=1}^{n} n_i, p)$	p 相同	有
Poisson $Pio(\lambda_i)$	$Pio(\sum_{i=1}^{n} \lambda_i)$	無限制	有
常態 $N(\mu_i,\sigma_i^2)$	$N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$	無限制	有
指數 $Exp(\beta)$	$Gamma(n,\beta)$		無
Gamma $G(\alpha_i,\beta)$	$G(\sum_{i=1}^{n} \alpha_i, \beta)$	β 相同	有
卡方 $\chi^2_{\nu_i}$	$\chi^2_{\sum_{i=1}^{n} \nu_i}$	無限制	有

■ 無記憶性

只有幾何分配與指數分配具有無記憶性的特質，其餘分配皆不具無記憶性

拾壹、抽樣與抽樣分配

■ 常用的抽樣方法



■ 影響抽樣分配的因素

- 母體本身的分配形狀。
- 所採用的樣本統計量
- 樣本大小

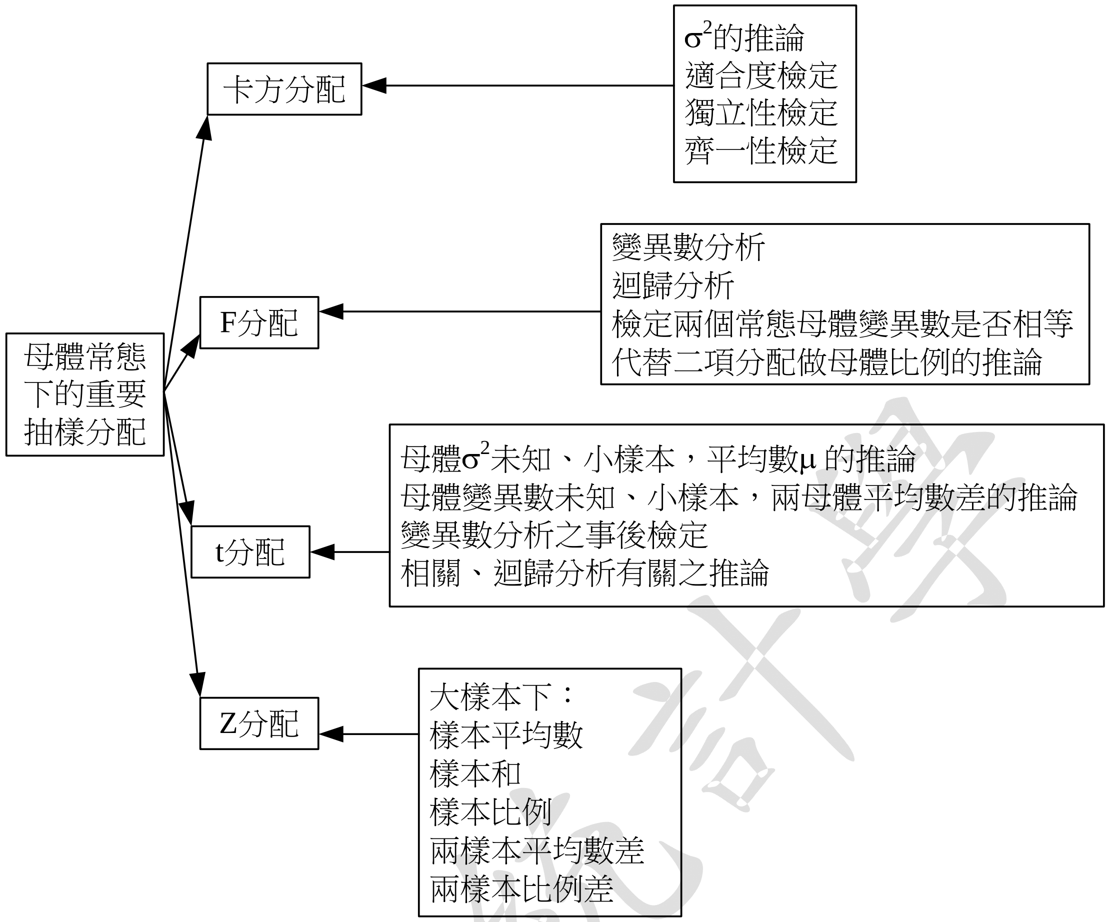
■ 大數法則

從一母體平均數為 μ 的母體中隨機抽取 n 個樣本，當樣本數 $n \to \infty$ 時，則樣本平均數會趨近於母體平均數

■ 中央極限定理

有一母體的平均數為 μ，變異數為 σ^2，自該母體隨機抽取 n 個樣本，當樣本數 n 夠大時，則樣本平均數 $\bar{x}$ 的抽樣分配會近似常態分配

■ 常態母體的四大抽樣分配



■ 常見的抽樣分配題型之期望值與變異數(大樣本情況)

名稱	平均數期望值	變異數	
		無限母體 $\frac{n}{N} \le 0.05$ 或 有限母體且取出歸還	有限母體且取出不歸還
$\bar{x}$ 的抽樣分配	μ	$\frac{\sigma^2}{n}$	$\frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}$
樣本和的抽樣分配	$n\mu$	$n\sigma^2$	$n\sigma^2\cdot\frac{N-n}{N-1}$
$\hat{p}$ 的抽樣分配	p	$\frac{pq}{n}$	$\frac{pq}{n}\cdot\frac{N-n}{N-1}$
$\bar{x}_1-\bar{x}_2$ 的抽樣分配	$\mu_1-\mu_2$	$\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$	$\frac{\sigma_1^2}{n_1}\frac{N_1-n_1}{N_1-1}+\frac{\sigma_2^2}{n_2}\frac{N_2-n_2}{N_2-1}$
$\hat{p}_1-\hat{p}_2$ 的抽樣分配	p_1-p_2	$\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}$	$\frac{p_1q_1}{n_1}\frac{N_1-n_1}{N_1-1}+\frac{p_2q_2}{n_2}\frac{N_2-n_2}{N_2-1}$

【若母體變異數未知，則以樣本變異數取代】

■ 卡方、F、t 統計量的公式

統計量	定義公式	實用公式
卡方	$\chi_\nu^2 = \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma})^2$	$\chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma^2}$
F	$F_{\nu_1,\nu_2} = \frac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2}$	$F_{n_1-1,n_2-1} = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$
t	$t_\nu = \frac{z}{\sqrt{\frac{\chi_\nu^2}{\nu}}}$	$t_{n-1} = \frac{\overline{x} - \mu}{\sqrt{\frac{s^2}{n}}}$

■ Z、卡方、F、t 分配的期望值與變異數

統計量	期望值	變異數
Z	0	1
卡方	ν	2ν
F	$\frac{\nu_2}{\nu_2 - 1}, \nu_2 > 2$	$\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \nu_2 > 4$
t	0	$\frac{\nu}{\nu - 2}, \nu > 2$

■ 卡方、t、F 分配與常態分配之關係

◆ $n \to \infty, \sqrt{2\chi^2} \sim N(\sqrt{2n-1}, 1)$

◆ $n \to \infty, t \sim N(0,1)$

◆ $n_1 = 1, n_2 \to \infty, \sqrt{F} \sim N(0,1)$

■ 重要查表性質

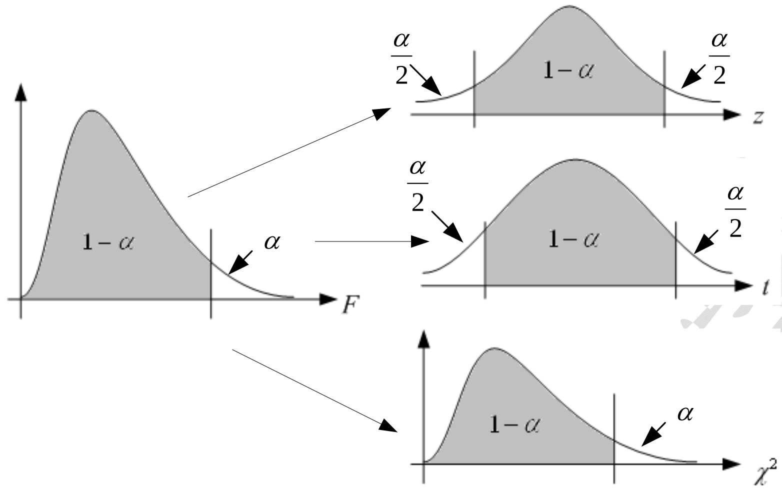
$$F_{\alpha,\nu_1,\nu_2} = \frac{1}{F_{1-\alpha,\nu_2,\nu_1}}$$

■ F 分配與另外三個分配間的轉換關係

分子 ν_1 / 分母 ν_2	1	ν_1	∞
1	$\frac{z_1^2}{z_2^2}$	$\frac{1}{t^2(\nu_1)}$	$\frac{1}{z^2}$
ν_2	$t^2(\nu_2)$	$F(\nu_1, \nu_2)$	$\frac{\nu_2}{\chi^2(\nu_2)}$
∞	z^2	$\frac{\chi^2(\nu_1)}{\nu_1}$	1

■ 以 F 統計量為主，查表轉換公式

- $z=\sqrt{F(1,\infty)}$
- $t(\nu)=\sqrt{F(1,\nu)}$
- $\chi^2(\nu)=\nu\cdot F(\nu,\infty)$



■ 常見抽樣分配題型的標準化公式

$\bar{x}$ 的抽樣分配 $\xrightarrow[\text{小樣本母體常態}]{\text{大樣本 } n \ge 30}$

無限母體 $\frac{n}{N} \le 0.05$
$\dfrac{\bar{x}-\mu}{\sqrt{\dfrac{\sigma^2}{n}}}$
有限母體且 取出不放回
$\dfrac{\bar{x}-\mu}{\sqrt{\dfrac{\sigma^2}{n}\dfrac{N-n}{N-1}}}$

S 的抽樣分配 $\xrightarrow[\text{小樣本母體常態}]{\text{大樣本 } n \ge 30}$

無限母體 $\frac{n}{N} \le 0.05$
$\dfrac{S-n\mu}{\sqrt{n\sigma^2}}$
有限母體且 取出不放回
$\dfrac{S-n\mu}{\sqrt{n\sigma^2\dfrac{N-n}{N-1}}}$

$\Rightarrow$ **Z**

$\hat{p}$ 的抽樣分配 $\xrightarrow{np \ge 5 \text{ 且 } nq \ge 5}$

無限母體 $\frac{n}{N} \le 0.05$
$\dfrac{\hat{p}-p}{\sqrt{\dfrac{pq}{n}}}$
有限母體且 取出不放回
$\dfrac{\hat{p}-p}{\sqrt{\dfrac{pq}{n}\dfrac{N-n}{N-1}}}$

$\bar{x}_1-\bar{x}_2$ 的抽樣分配 $\xrightarrow[\text{小樣本母體常態}]{\text{大樣本}\ n_1,n_2\ge 30}$

無限母體 $\frac{n}{N}\le 0.05$
$\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$
有限母體且 取出不放回
$\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}\dfrac{N_1-n_1}{N_1-1}+\dfrac{\sigma_2^2}{n_2}\dfrac{N_2-n_2}{N_2-1}}}$

$\Rightarrow$ **Z**

$\hat{p}_1-\hat{p}_2$ 的抽樣分 $\xrightarrow[n_2p_2\ge 5\ \text{且}\ n_2q_2\ge 5]{n_1p_1\ge 5\ \text{且}\ n_1q_1\ge 5}$

無限母體 $\frac{n}{N}\le 0.05$
$\dfrac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\dfrac{p_1q_1}{n_1}+\dfrac{p_2q_2}{n_2}}}$
有限母體且 取出不放回
$\dfrac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\dfrac{p_1q_1}{n_1}\dfrac{N_1-n_1}{N_1-1}+\dfrac{p_2q_2}{n_2}\dfrac{N_2-n_2}{N_2-1}}}$

拾貳、估計

■ 估計式

用來估計母體參數的樣本統計量，通式以 $\hat{\theta}$ 表示

■ 估計式的評斷標準

- 不偏性
- 有效性
  - 絕對有效性
  - 相對有效性
  - 最小變異不偏性
  - 漸進不偏性
- 一致性
- 充分性

■ 評斷標準的定義

評斷標準		定　　義
不偏性		$E(\hat{\theta})=\theta$
有效性	絕對	$MSE(\hat{\theta})=E[(\hat{\theta}-\theta)^2]$，為所有估計式中最小者
	相對	$\dfrac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}<1$ 表 $\hat{\theta}_1$ 相對 $\hat{\theta}_2$ 具相對有效性
最小變異不偏性		$E(\hat{\theta})=\theta$ 且 $V(\hat{\theta})$ 小於其他不偏估計式的變異數
漸進不偏性		$\lim\limits_{n\to\infty}E(\hat{\theta}_n)=\theta$ 或 $\lim\limits_{n\to\infty}\left[E(\hat{\theta}_n)-\theta\right]=0$
一致性	強則	$P(\lim\limits_{n\to\infty}\hat{\theta}_n=\theta)=1$
	弱則	$\lim\limits_{n\to\infty}P(\left\|\hat{\theta}_n-\theta\right\|<\varepsilon)=1,\forall\varepsilon>0$
充分性		$x_1,x_2,\cdots,x_n$ 的分配與母體參數無關，由這組樣本所求得的統計量稱為充分統計量

■ 常用的估計式的偏誤情況

	樣本取出放回	樣本取出不放回
$\bar{x}$ 估計 μ	不偏	不偏
s^2 估計 σ^2	不偏	偏誤
s 估計 σ	偏誤	偏誤
$\hat{p}$ 估計 p	不偏	不偏
$\hat{p}\hat{q}$ 估計 pq	偏誤	偏誤
$\hat{p}_1-\hat{p}_2$ 估計 p_1-p_2	不偏	不偏

■ 點估計的方法

名　　稱	方　　　　　　　　法
最小平方法	$SSE=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(x_i-\hat{\theta})^2=Min\sum_{i=1}^{n}(x_i-E(x\|\theta))^2$
動差法	$\mu_r=E(x^r)\quad ,r=1,2,3,\cdots$ ， $m_r=\frac{1}{n}\sum_{i=1}^{n}x_i^r\quad ,r=1,2,3,\cdots$ $m_r=\mu_r\quad ,r=1,2,3,\cdots,p$
最大概似法	$L(x_1,x_2,\cdots,x_n;\theta_1,\theta_2,\cdots,\theta_l)=\prod_{i=1}^{n}f(x_i;\theta_1,\theta_2,\cdots,\theta_l)$ $\frac{\partial}{\partial\theta_i}\ln L=0$

■ 點估計式尋找方法的性質

	不偏性	最小變異不偏性	漸進不偏性	一致性	有效性	充分性	需已知母體分配
最小平方法	✓	✓(線性)		✓			✗
動差估計式				✓			✓
最大概似法			✓	✓	✓(漸進)		✓

拾參、區間估計

- **信賴水準**

  區間包含母體參數的信心或可靠度，信賴水準的大小等於$1-\alpha$。

- **信賴區間**

  在一個給定的信賴水準$(1-\alpha)$下所構成的一個區間。它是由樣本統計量以及抽樣誤差所構成的一個包含上限、下限的區間

- **母數的信賴區間**

  - 第一類

種類	使用時機			雙尾信賴區間
μ	大樣本	σ^2已知		$\mu=\bar{x}\pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$
		σ^2未知		$\mu=\bar{x}\pm z_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}$
	小樣本母體常態	σ^2已知		$\mu=\bar{x}\pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$
		σ^2未知		$\mu=\bar{x}\pm t_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}$
	小樣本母體非常態	σ^2已知		$\mu=\bar{x}\pm\sqrt{\frac{1}{\alpha}}\sqrt{\frac{\sigma^2}{n}}$
p	大樣本			$p=\hat{p}\pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$
$\mu_1-\mu_2$	大樣本	σ^2已知		$\mu_1-\mu_2=(\bar{x}_1-\bar{x}_2)\pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
		σ^2未知		$\mu_1-\mu_2=(\bar{x}_1-\bar{x}_2)\pm z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
	小樣本母體常態	σ^2已知		$\mu_1-\mu_2=(\bar{x}_1-\bar{x}_2)\pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
		σ^2未知	相等	$\mu_1-\mu_2=(\bar{x}_1-\bar{x}_2)\pm t_{\frac{\alpha}{2},n_1+n_2-2}\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}$ $s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$
			不等	$\mu_1-\mu_2=(\bar{x}_1-\bar{x}_2)\pm t_{\frac{\alpha}{2},\nu}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ $\nu=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$

種　類	使用時機		雙　尾　信　賴　區　間
$p_1 - p_2$	大樣本	母體變異數相等	$p_1 - p_2 = (\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$
		母體變異數不相等	$p_1 - p_2 = (\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}}\sqrt{(\frac{\overline{pq}}{n_1} + \frac{\overline{pq}}{n_2})}$ $\overline{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$
μ_d	母體常態大樣本	母體變異數未知	$\mu_d = \overline{d} \pm t_{\frac{\alpha}{2},n-1}\sqrt{\frac{s_d^2}{n}}$

註：上述若為有限母體，必須補上有限母體修正因子

◆　第二類

種　類	使用時機	雙　尾　信　賴　區　間
σ^2	μ 已知	$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$
	μ 未知	$\frac{n\hat{\sigma}^2}{\chi^2_{\frac{\alpha}{2},n}} \le \sigma^2 \le \frac{n\hat{\sigma}^2}{\chi^2_{1-\frac{\alpha}{2},n}}$
$\frac{\sigma_1^2}{\sigma_2^2}$		$\frac{s_1^2/s_2^2}{F_{n_1-1,n_2-1,\alpha/2}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2/s_2^2}{F_{n_1-1,n_2-1,1-\alpha/2}}$

■　影響信賴區間長的因素

信賴水準

樣本數

估計式的選擇

拾肆、假設檢定 I、II

■ 兩種錯誤的機率

◆ 型 I 的錯誤

當 H_0 為真(H_1 為偽)，拒絕 H_0 所發生的錯誤，稱為型 I 的錯誤。型 I 錯誤中最大者稱為 α 錯誤，也稱為 α 風險或顯著水準。

$$\alpha = Max\,P(\text{拒絕 } H_0 | H_0 \text{ 為真})$$

◆ 型 II 的錯誤

當 H_1 為真(H_0 為偽)，拒絕 H_1 所發生的錯誤，稱為型 II 的錯誤，又稱為 β 錯誤、β 風險。

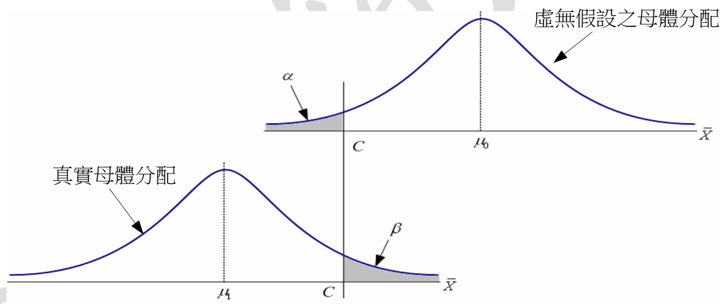
$$\beta = P(\text{拒絕 } H_1 | H_1 \text{ 為真})$$

◆ 整理

真實狀況	決策	
	不拒絕 H_0	拒絕 H_0
H_0 為真	正確　機率：$1-\alpha$	型 I 錯誤　機率：α
H_0 為僞	型 II 錯誤　機率：β	正確　機率：$1-\beta$

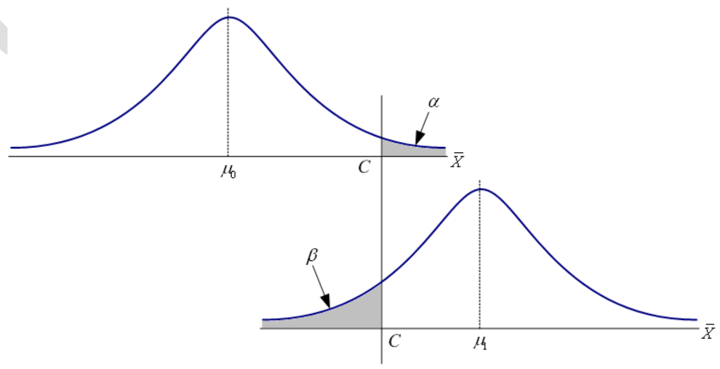
■ 型 II 錯誤 β 的推導

◆ 左尾檢定



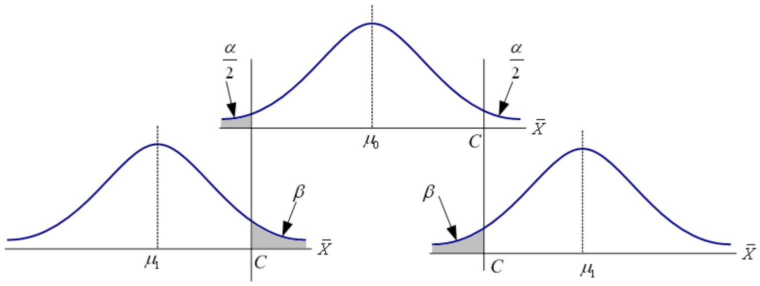
$$C = \mu_0 - z_\alpha\sqrt{\frac{\sigma^2}{n}} = \mu_1 + z_\beta\sqrt{\frac{\sigma^2}{n}}$$

◆ 右尾檢定



$$C = \mu_0 + z_\alpha\sqrt{\frac{\sigma^2}{n}} = \mu_1 - z_\beta\sqrt{\frac{\sigma^2}{n}}$$

◆ 雙尾檢定



$$C=\mu_0+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}=\mu_1-z_\beta\sqrt{\frac{\sigma^2}{n}} \text{ 或 } C=\mu_0-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}=\mu_1+z_\beta\sqrt{\frac{\sigma^2}{n}}$$

■ 給定誤差α、β的條件下，所需樣本數

◆ 單尾檢定

$$n=\frac{\sigma^2(z_\alpha+z_\beta)^2}{(\mu_1-\mu_0)^2}$$

◆ 雙尾檢定

$$n=\frac{\sigma^2(z_{\alpha/2}+z_\beta)^2}{(\mu_1-\mu_0)^2}$$

■ 檢定力

$$\text{檢定力}=1-\beta$$

■ 檢定力函數

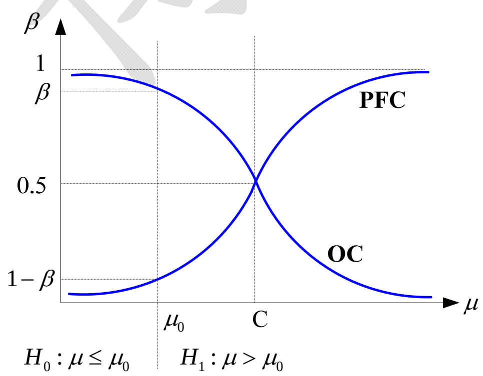
$$f(\theta)=1-\beta(\theta)$$

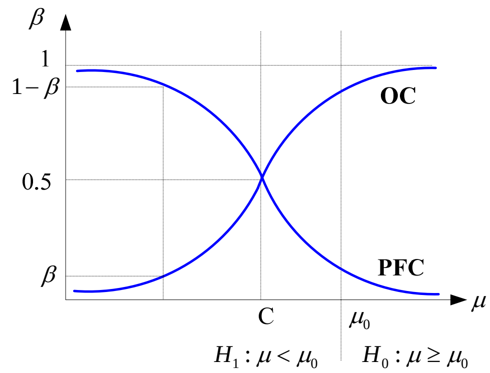
■ 影響檢定力的因素

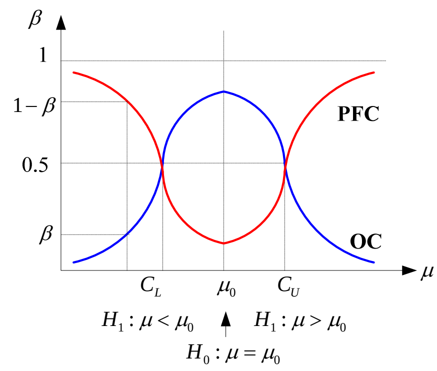
- 樣本大小
- 顯著水準
- 檢定統計量的選擇
- 決策法則之決定

■ 作業特性曲線

在所有可能的母體平均數之下，將其對應所犯型 II 錯誤的機率 β 繪成的曲線



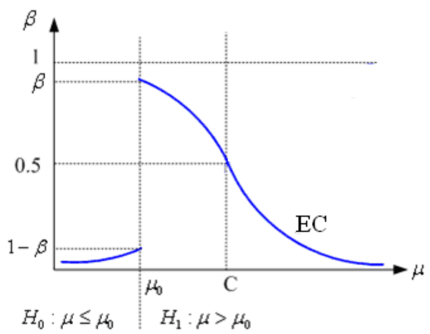


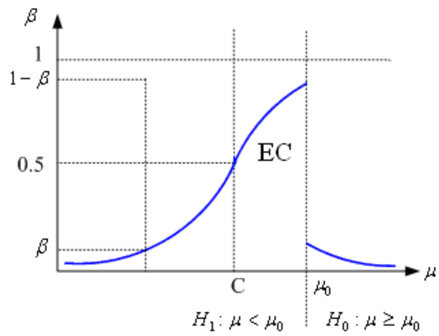


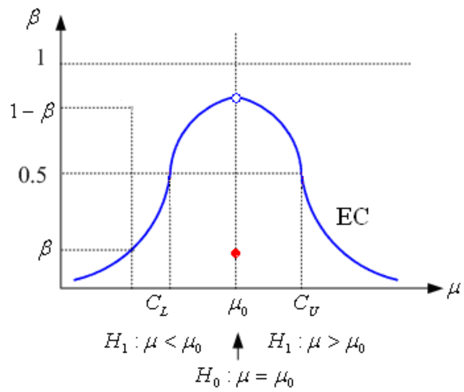
■ 錯誤曲線

在所有可能母體參數θ情況下，可能犯錯機率的曲線

$$\text{EC}=\begin{cases} PFC, & \theta \in H_0 \\ OCC, & \theta \in H_1 \end{cases}$$







■ 標準檢定法

◆ 單母體平均數的檢定

- 大樣本
  - σ^2已知：$z^* = \dfrac{\overline{x}-\mu_0}{\sqrt{\dfrac{\sigma^2}{n}}}$
  - σ^2未知：$z^* = \dfrac{\overline{x}-\mu_0}{\sqrt{\dfrac{s^2}{n}}}$
- 小樣本
  - 常態
    - σ^2已知：$z^* = \dfrac{\overline{x}-\mu_0}{\sqrt{\dfrac{\sigma^2}{n}}}$
    - σ^2未知：$t^*_{n-1} = \dfrac{\overline{x}-\mu_0}{\sqrt{\dfrac{s^2}{n}}}$
  - 非常態
    - σ^2已知：$k^* = \dfrac{\overline{x}-\mu_0}{\sqrt{\dfrac{\sigma^2}{n}}}$
    - σ^2未知：無母數統計
      - Wilcoxon 符號檢定
      - 符號檢定 Sign test

若為有限母體，且取出不放回，標準誤部分需加有限母體修正因子$\sqrt{\dfrac{N-n}{N-1}}$

◆ 單母體比例的檢定

- 大樣本：$z^* = \dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$
- 小樣本：以二項分配求出P-value 使用P值法

若為有限母體，且取出不放回，標準誤部分需加有限母體修正因子$\sqrt{\dfrac{N-n}{N-1}}$

◆ 單母體變異數的檢定

- 常態
  - μ已知：$\chi^{2*} = \dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma_0^2}$
  - μ未知：$\chi^{2*} = \dfrac{(n-1)s^2}{\sigma_0^2}$

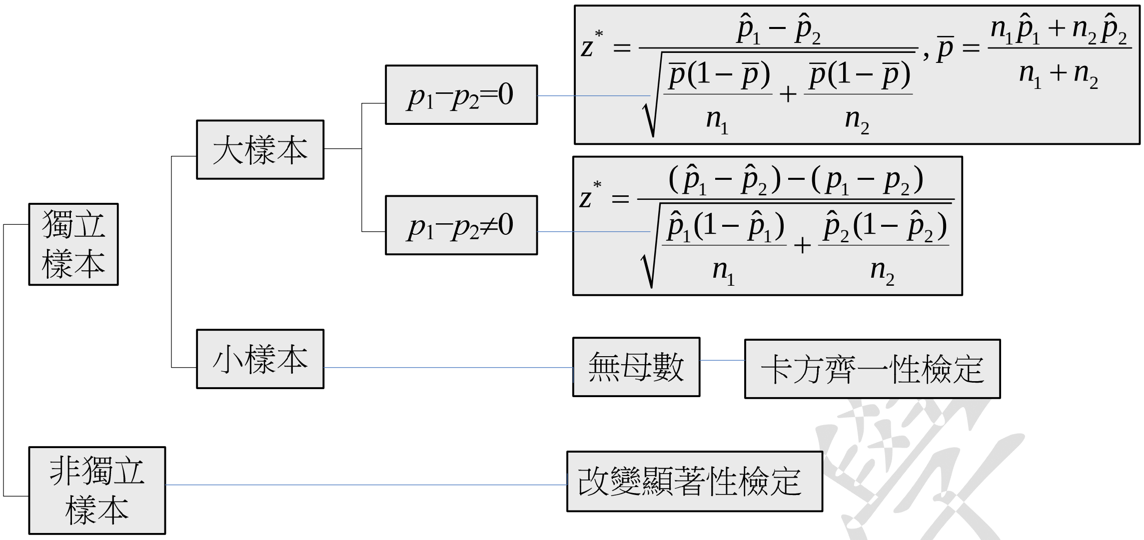
◆ 兩母體平均數差的檢定-獨立樣本

- 大樣本
  - σ^2已知：$z^* = \dfrac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$
  - σ^2未知：$z^* = \dfrac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$
- 小樣本
  - 常態
    - σ^2已知：$z^* = \dfrac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$
    - σ^2未知但不等：$t^* = \dfrac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}, \nu = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$
    - σ^2未知但相等：$t^* = \dfrac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}, s_p = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$
  - 非常態：無母數
    - Wilcoxon等級和檢定
    - Mann-Whitney U檢定

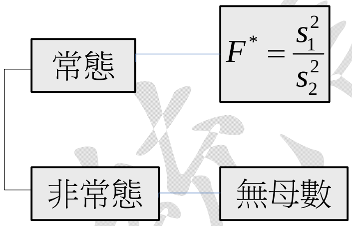
◆ 兩母體平均數差的檢定-成對樣本

- 大樣本：$z^* = \dfrac{\overline{d} - \mu_d}{\sqrt{\dfrac{s_d^2}{n}}}$
- 小樣本
  - 常態：$t^* = \dfrac{\overline{d} - \mu_d}{\sqrt{\dfrac{s_d^2}{n}}}$；變異數分析之隨機集區
  - 非常態：無母數
    - 符號檢定
    - Wilcoxon符號檢定

◆ 兩母體比例差的檢定



◆ 兩母體變異數比的檢定



■ 四種檢定法

◆ 信賴區間法

◆ 臨界值法

◆ 標準檢定法

◆ P 值法

拾陸、變異數分析

■ 單因子變異數分析（完全隨機）

◆ 變異之計算

$SST = SSA + SSE$

$$SST=\sum_{i=1}^{n_i}\sum_{j=1}^{k}(x_{ij}-\bar{\bar{x}})^2=\sum_{i=1}^{n_i}\sum_{j=1}^{k}x_{ij}^2-n_T\bar{\bar{x}}^2=n_T\sigma^2$$

$$SSA=\sum_{i=1}^{n_i}\sum_{j=1}^{k}(\bar{x}_j-\bar{\bar{x}})^2=\sum_{j=1}^{k}n_j\bar{x}_j^2-n_T\bar{\bar{x}}^2\overset{n_j\text{相同}}{=}n_T\sigma_{\bar{x}}^2$$

$$SSE=\sum_{i=1}^{n_i}\sum_{j=1}^{k}(x_{ij}-\bar{x}_j)^2=SST-SSA=\sum_{j=1}^{k}(n_j-1)s_j^2$$

◆ 變異數分析表

變異來源	平方和(SS)	自由度(DF)	平均平方和(MS)	(檢定統計量)F
因子	SSA	$k-1$	$MSA=\dfrac{SSA}{k-1}$	$F^*=\dfrac{MSA}{MSE}$
隨機	SSE	n_T-k	$MSE=\dfrac{SSE}{n_T-k}$	
總和	SST	n_T-1		

若 $F^*>F_{\alpha,k-1,n_T-k}$ 或 P-value$=P(F_{k-1,n_T-k}>F^*)<\alpha$ $\Rightarrow$拒絕 H_0

■ 單因子變異數分析隨機集區設計、雙因子未重複

◆ 變異之計算

$SST=SSA+SSB+SSE$

$$SST=\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-\bar{\bar{x}})^2=\sum_{i=1}^{r}\sum_{j=1}^{c}x_{ij}^2-n_T\bar{\bar{x}}^2=n_T\sigma^2$$

$$SSA=\sum_{i=1}^{r}\sum_{j=1}^{c}(\bar{A}_j-\bar{\bar{x}})^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\bar{A}_j^2-n_T\bar{\bar{x}}^2=r\sum_{i=1}^{c}\bar{A}_j^2-n_T\bar{\bar{x}}^2=n_T\sigma_{\bar{A}_j}^2$$

$$SSB=\sum_{i=1}^{r}\sum_{j=1}^{c}(\bar{B}_i-\bar{\bar{x}})^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\bar{B}_i^2-n_T\bar{\bar{x}}^2=c\sum_{j=1}^{c}\bar{B}_i^2-n_T\bar{\bar{x}}^2=n_T\sigma_{\bar{B}_i}^2$$

$$SSE=\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij}-\bar{A}_j-\bar{B}_i+\bar{\bar{x}})^2=SST-SSA-SSB$$

◆ 變異數分析表

變異來源	平方和	自由度	平均平方和	F 值
A 因子	SSA	$c-1$	$MSA=\dfrac{SSA}{c-1}$	$F_A^*=\dfrac{MSA}{MSE}$
B 因子	SSB	$r-1$	$MSB=\dfrac{SSB}{r-1}$	$F_B^*=\dfrac{MSB}{MSE}$
隨機	SSE	$(c-1)(r-1)$	$MSE=\dfrac{SSE}{(c-1)(r-1)}$	
總變異	SST	n_T-1		

檢定 A 因子是否有影響，若 $F_A^* > F_{\alpha,(c-1),(r-1)(c-1)}$ $\Rightarrow$拒絕 H_0

檢定 B 因子是否有影響，若 $F_B^* > F_{\alpha,(r-1),(r-1)(c-1)}$ $\Rightarrow$拒絕 H_0

■ 雙因子重複

◆ 變異之計算

$SST=SSA+SSB+SSAB+SSE$

$$SST=\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}(x_{ijk}-\bar{\bar{x}})^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}x_{ijk}^2-n_T\bar{\bar{x}}^2=n_T\sigma_T^2$$

$$SSA=rn\sum_{j=1}^{c}(\bar{A}_j-\bar{\bar{x}})^2=rn\sum_{j=1}^{c}\bar{A}_j^2-n_T\bar{\bar{x}}=n_T\sigma_{\bar{A}_j}^2$$

$$SSB=cn\sum_{i=1}^{r}(\bar{B}_i-\bar{\bar{x}})^2=cn\sum_{i=1}^{r}\bar{B}_i^2-n_T\bar{\bar{x}}=n_T\sigma_{\bar{B}_j}^2$$

$$SSE=\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}(x_{ijk}-\overline{A_jB_i})^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}x_{ijk}^2-n\sum_{i=1}^{r}\sum_{j=1}^{c}\overline{A_jB_i}^2$$

$$SSAB=\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}(\overline{A_jB_i}-\bar{A}_j-\bar{B}_j+\bar{\bar{x}})^2=SST-SSA-SSB-SSE$$

◆ 變異數分析表

變異來源	平方和	自由度	平均平方和	F 值
A 因子	SSA	$c-1$	$MSA=\dfrac{SSA}{c-1}$	$F_A^*=\dfrac{MSA}{MSE}$
B 因子	SSB	$r-1$	$MSB=\dfrac{SSB}{r-1}$	$F_B^*=\dfrac{MSB}{MSE}$
交互作用	$SSAB$	$(c-1)(r-1)$	$MSAB=\dfrac{SSAB}{(c-1)(r-1)}$	$F_{AB}^*=\dfrac{MSAB}{MSE}$
隨機	SSE	$rc(n-1)$	$MSE=\dfrac{SSE}{rc(n-1)}$	
總變異	SST	n_T-1		

- 檢定 A 因子是否有影響

  若 $F_A^* > F_{\alpha,(c-1),rc(n-1)}$ $\Rightarrow$拒絕 H_0，表 A 因子對依變數具有影響。

- 檢定 B 因子是否有影響

  若 $F_B^* > F_{\alpha,(r-1),rc(n-1)}$ $\Rightarrow$拒絕 H_0，表 B 因子對依變數具有影響。

- 檢定 AB 因子是否有交互影響

  若 $F_{AB}^* > F_{\alpha,(r-1)(c-1),rc(n-1)}$ $\Rightarrow$拒絕 H_0，表 A、B 因子對依變數具有交

互作用。

■ 多重比較程序

◆ Fisher 最小顯著差異法

$$LSD = t_{\frac{\alpha}{2}, n_T - k}\sqrt{\frac{MSE}{n_i} + \frac{MSE}{n_j}}$$

若 $\left|\bar{x}_i - \bar{x}_j\right| \geq LSD \Rightarrow$ 表示 μ_i 與 μ_j 有顯著差異

◆ Bonferroni 法

$$\mu_i - \mu_j = (\bar{x}_i - \bar{x}_j) \pm t_{n_T - k, \frac{\alpha}{2m}}\sqrt{\frac{MSE}{n_i} + \frac{MSE}{n_j}}$$

若 $\mu_i - \mu_j$ 之聯合信賴區間包含 0，則接受虛無假設。

◆ Scheffe 法

$$\mu_i - \mu_j = (\bar{x}_i - \bar{x}_j) \pm \sqrt{(k-1)F_{\alpha, k-1, n_T - k}}\sqrt{\frac{MSE}{n_i} + \frac{MSE}{n_j}}$$

若 $\mu_i - \mu_j$ 之聯合信賴區間包含 0，則接受虛無假設

◆ Tukey 公正顯著差異法

$$HSD = q_{\alpha, k, n_T - k}\sqrt{\frac{MSE}{n} + \frac{MSE}{n}}$$

若 $\left|\bar{x}_i - \bar{x}_j\right| > HSD$ ，則拒絕虛無假設

◆ Tukey-Karamer 檢定程序

$$\omega = \frac{q_{\alpha, k, n_T - k}}{\sqrt{2}}\sqrt{\frac{MSE}{n_i} + \frac{MSE}{n_j}}$$

當 $\left|\bar{x}_i - \bar{x}_j\right| > \omega$ 則拒絕虛無假設

■ 共同母體變異數的信賴區間

$$\frac{(n_T - k)MSE}{\chi^2_{\frac{\alpha}{2}, n_T - k}} \leq \sigma^2 \leq \frac{(n_T - k)MSE}{\chi^2_{1-\frac{\alpha}{2}, n_T - k}}$$

■ 多個母體變異數之檢定

◆ Hartley 檢驗法

$$H^* = \frac{\max s_i^2}{\min s_i^2}$$

當 $H^* > H_{\alpha, k, n}$ $\Rightarrow$ 拒絕 H_0

◆ Bartlett 檢定法

$$B = \frac{2.3026}{C}(n_T - k)(\log\frac{MSE}{GMSE}), C = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{n_T - k}\right)$$

若 $B > \chi^2_{\alpha, k-1}$ $\Rightarrow$ 拒絕 H_0

拾柒、簡單線性迴歸與相關分析

■ 簡單線性迴歸的五個重要式子

- 迴歸模型：$y = E(y|x) + \varepsilon = \mu_{y|x} + \varepsilon = \alpha + \beta x + \varepsilon$
- 母體迴歸線：$E(y|x) = \mu_{y|x} = \alpha + \beta x$
- 誤差：$\varepsilon = y - E(y|x) = y - \mu_{y|x} = y - \alpha - \beta x$
- 樣本迴歸線：$\hat{y} = \hat{\alpha} + \hat{\beta} x$
- 殘差：$e = y - \hat{y} = y - \hat{\alpha} - \hat{\beta} x$

■ 簡單線性迴歸模型的基本假設

- $E(\varepsilon_i) = 0$
- $V(\varepsilon_i) = \sigma^2$
- $Cov(\varepsilon_i, \varepsilon_j) = 0$

■ 樣本迴歸係數公式

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x}) y_i}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n} x_i (y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x}) x_i} = \frac{s_{xy}^2}{s_x^2}$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

■ 變異之計算

$$SST{=}SSR{+}SSE$$

$$SST = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 = (n-1)s_y^2 = n\hat{\sigma}_y^2$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 = \hat{\beta}^2 \sum_{i=1}^{n}(x_i - \overline{x})^2 = \hat{\beta}^2 \left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 \right) = \hat{\beta}^2 (n-1) s_x^2 = \hat{\beta}^2 \cdot n\hat{\sigma}_x^2$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} y_i^2 - \hat{\alpha}\sum_{i=1}^{n} y_i - \hat{\beta}\sum_{i=1}^{n} x_i y_i$$

■ 變異數分析表

變異來源	平方和	自由度	平均平方和	F
迴歸	SSR	1	$MSR = \frac{SSR}{1}$	$F^* = \frac{MSR}{MSE}$
誤差	SSE	$n-2$	$MSE = \frac{SSE}{n-2}$	
總和	SST	$n-1$		

若檢定統計量 $F^* > F_{1,n-2,\alpha}$ 則拒絕 H_0

■ 判定係數

$$R^2=\frac{SSR}{SST}=\frac{\sum(\hat{y}_i-\overline{y})^2}{\sum(y_i-\overline{y})^2}=\hat{\beta}^2\cdot\frac{s_x^2}{s_y^2}=\hat{\beta}^2\cdot\frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2}$$

■ 斜率項的檢定

$$\text{雙尾檢定}\begin{cases}H_0:\beta=0\\H_1:\beta\neq 0\end{cases}$$

$$\text{單尾檢定}\begin{cases}H_0:\beta\leq 0\\H_1:\beta>0\end{cases}\text{或}\begin{cases}H_0:\beta\geq 0\\H_1:\beta<0\end{cases}\Rightarrow t\text{ 檢定}$$

◆ 檢定統計量

$$t^*=\frac{\hat{\beta}-\beta_0}{s_{\hat{\beta}}}=\frac{\hat{\beta}}{\sqrt{\frac{MSE}{\sum(x_i-\overline{x})^2}}}$$

◆ β 的 $1-\alpha$ 信賴區間

$$\hat{\beta}-t_{\frac{\alpha}{2},n-2}\sqrt{\frac{MSE}{\sum(x_i-\overline{x})^2}}\leq\beta\leq\hat{\beta}+t_{\frac{\alpha}{2},n-2}\sqrt{\frac{MSE}{\sum(x_i-\overline{x})^2}}$$

■ 截距項的檢定

◆ 檢定統計量

$$t^*=\frac{\hat{\alpha}-\alpha_0}{s_{\hat{\alpha}}}=\frac{\hat{\alpha}}{\sqrt{\frac{\sum x_i^2}{n}\cdot\frac{MSE}{\sum(x_i-\overline{x})^2}}}$$

■ 截距項的 $1-\alpha$ 信賴區間

$$\hat{\alpha}-t_{\frac{\alpha}{2},n-2}\sqrt{\frac{\sum x_i^2}{n}\frac{MSE}{\sum(x_i-\overline{x})^2}}\leq\alpha\leq\hat{\alpha}+t_{\frac{\alpha}{2},n-2}\sqrt{\frac{\sum x_i^2}{n}\frac{MSE}{\sum(x_i-\overline{x})^2}}$$

■ $E(y|x_0)$ 的信賴區間

$$E(y|x_0)=\hat{y}_0\pm t_{\frac{\alpha}{2},n-2}\sqrt{MSE\left[\frac{1}{n}+\frac{(x_0-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right]}$$

■ y_0 的預測區間

$$y_0=\hat{y}_0\pm t_{\frac{\alpha}{2},n-2}\sqrt{\left[1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right]MSE}$$

■ 影響信賴區間與預測區間長度的因素

- 顯著水準越大 t 值越小，信賴區間越短。
- MSE 越大則信賴區間越長。
- x_0 離 $\overline{x}$ 越遠，信賴區間越長。
- 樣本數 n 越大，信賴區間越短。

■ 率項 $\hat{\beta}$ 與樣本相關係數 r_{xy} 之關係

$$r_{xy} = \hat{\beta}\frac{s_x}{s_y}$$

■ 相關係數與判定係數間的關係

$$r_{xy} = \pm\sqrt{R^2}$$

正負號與斜率項相同

■ 虛無假設 $\rho_{xy} = 0$ 情況下的統計推論

$$t^* = \frac{r_{xy}}{\sqrt{\dfrac{1-r_{xy}^2}{n-2}}}$$

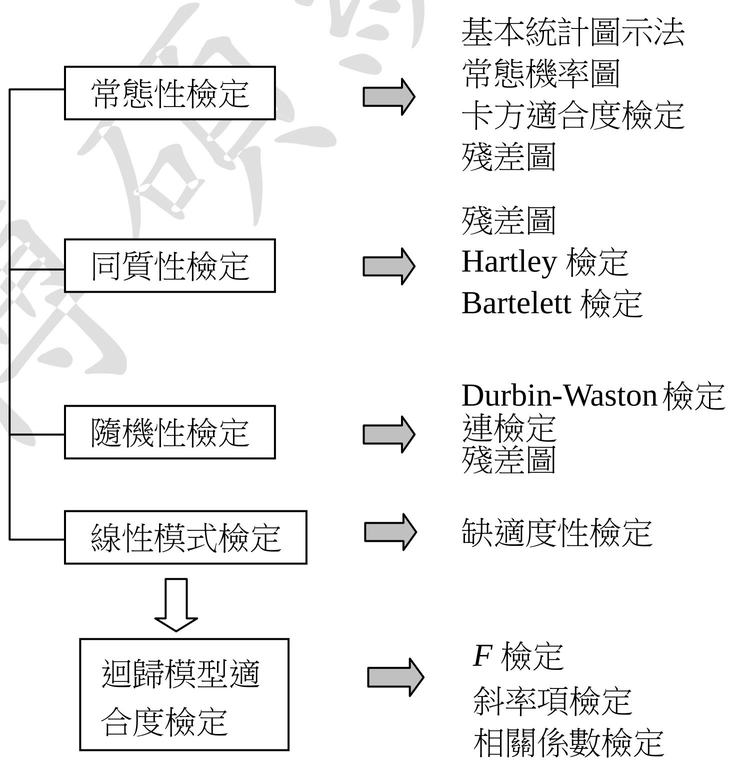
■ 虛無假設 $\rho_{xy} = \rho_0(\rho_0 \neq 0)$ 情況下的統計推論

$$z^* = \frac{\frac{1}{2}\ln(\frac{1+r_{xy}}{1-r_{xy}}) - \frac{1}{2}\ln(\frac{1+\rho_0}{1-\rho_0})}{\sqrt{\dfrac{1}{n-3}}}$$

◆ 母體相關係數的信賴區間

$$\frac{1}{2}\ln(\frac{1+\rho_{xy}}{1-\rho_{xy}}) = \frac{1}{2}\ln(\frac{1+r_{xy}}{1-r_{xy}}) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n-3}} \Rightarrow \rho_{xy}$$

◆ 簡單線性迴歸模型的建購流程



■ 線性模式適合度的檢定

◆ 變異的分解

$$SSE = SSL + SSPE$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_i)^2, SSL = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\hat{y}_i - \overline{y}_i)^2, SSPE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2$$

◆ 變異數分析表

變異來源	平方和	自由度	均方和	F
缺適性	SSL	$k-2$	$MSL = \dfrac{SSL}{k-2}$	$F^* = \dfrac{MSL}{MSPE}$
純誤差	$SSPE$	$n-k$	$MSPE = \dfrac{SSPE}{n-k}$	
誤差	SSE	$n-2$		

拾捌、多元迴歸

■ 多元迴歸模型的基本假設

◆ 依變數為常態隨機變數，而自變數為選定之控制變數。

◆ 對不同的 x 值，誤差的期望值為 0，即 $E(\varepsilon_i)=0$

◆ 具變異數齊一性，即 $E(\varepsilon_i)=\sigma^2$ 或 $\sigma^2_{y|x_1,x_2,\cdots,x_k}=\sigma^2$ 。

◆ ε 相互獨立，即 $Cov(\varepsilon_i,\varepsilon_j)=0, i\neq j$ 。

◆ ε 與自變數 x 無關，即 $Cov(\varepsilon_i,x_j)=0$ 。

■ 二元線性迴歸方程係數的推導

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}=\begin{bmatrix} S_{1y} \\ S_{2y} \end{bmatrix}$$

$$\hat{\alpha}=\overline{y}-\hat{\beta}_1\overline{x}_1-\hat{\beta}_2\overline{x}_2$$

$$S_{11}=\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2=\sum_{i=1}^{n}x_{1i}^2-n\overline{x}_1^2=n\sigma^2_{x_1}$$

$$S_{22}=\sum_{i=1}^{n}(x_{2i}-\overline{x}_2)^2=\sum_{i=1}^{n}x_{2i}^2-n\overline{x}_2^2=n\sigma^2_{x_2}$$

$$S_{12}=S_{21}=\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)(x_{2i}-\overline{x}_2)=\sum_{i=1}^{n}x_{1i}x_{2i}-n\overline{x}_1\overline{x}_2=Cov(x_1,x_2)$$

$$S_{1y}=\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)(y_i-\overline{y})=\sum_{i=1}^{n}x_{1i}y_i-n\overline{x}_1\overline{y}=Cov(x_1,y)$$

$$S_{2y}=\sum_{i=1}^{n}(x_{2i}-\overline{x}_2)(y_i-\overline{y})=\sum_{i=1}^{n}x_{2i}y_i-n\overline{x}_2\overline{y}=Cov(x_2,y)$$

■ 二元迴歸的統計推論

◆ 變異的分解

$$SST=SSR+SSE$$

$$SST=\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}y_i^2-n\overline{y}^2=n\sigma_y^2=(n-1)s_y^2$$

$$SSR=\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2=\hat{\beta}_1S_{1y}+\hat{\beta}_2S_{2y}=SST-SSE$$

$$SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}y_i^2-\hat{\alpha}\sum_{i=1}^{n}y_i-\hat{\beta}_1\sum_{i=1}^{n}y_ix_{1i}-\hat{\beta}_2\sum_{i=1}^{n}y_ix_{2i}=SST-SSR$$

◆ 變異數分析表

變異來源	平方和	自由度	均方和	F
迴歸	SSR	2	$MSR=\dfrac{SSR}{2}$	$F^*=\dfrac{MSR}{MSE}$
誤差	SSE	$n-3$	$MSE=\dfrac{SSE}{n-3}$	
總和	SST	$n-1$		

$$F^* > F_{\alpha,2,n-3}\text{時}\Rightarrow\text{拒絕虛無假設}$$

■ 二元迴歸的迴歸係數統計推論

◆ β_1 的檢定

$$t^* = \frac{\hat{\beta}_1}{\sqrt{\frac{S_{22}}{S_{11}S_{22}-S_{12}^2}MSE}}$$

◆ β_1 的信賴區間

$$\beta_1 = \hat{\beta}_1 \pm t_{\frac{\alpha}{2},n-3}\sqrt{\frac{S_{22}}{S_{11}S_{22}-S_{12}^2}MSE}$$

◆ β_2 的檢定

$$t^* = \frac{\hat{\beta}_2}{\sqrt{\frac{S_{11}}{S_{11}S_{22}-S_{12}^2}MSE}}$$

◆ β_2 的信賴區間

$$\beta_2 = \hat{\beta}_2 \pm t_{\frac{\alpha}{2},n-3}\sqrt{\frac{S_{11}}{S_{11}S_{22}-S_{12}^2}MSE}$$

◆ α 的檢定

$$t^* = \frac{\hat{\alpha}}{\sqrt{(\frac{\bar{x}_1^2S_{22}-2\bar{x}_1\bar{x}_2S_{12}+\bar{x}_2^2S_{11}}{S_{11}S_{22}-S_{12}^2}+\frac{1}{n})MSE}}$$

◆ α 的信賴區間

$$\alpha = \hat{\alpha} \pm t_{\frac{\alpha}{2},n-3}\sqrt{(\frac{\bar{x}_1^2S_{22}-2\bar{x}_1\bar{x}_2S_{12}+\bar{x}_2^2S_{11}}{S_{11}S_{22}-S_{12}^2}+\frac{1}{n})MSE}$$

■ 二元迴歸依變數的 $1-\alpha$ 信賴區間

$$\mu_{y_0|x_{10},x_{20}} = \hat{y}_0 \pm t_{\frac{\alpha}{2},n-3}\sqrt{(\frac{x_{10}^2S_{22}-2x_{10}x_{20}S_{12}+x_{20}^2S_{11}}{S_{11}S_{22}-S_{12}^2}+\frac{1}{n})MSE}$$

■ 二元迴歸依變數的 $1-\alpha$ 預測區間區間

$$y_0 = \hat{y}_0 \pm t_{\frac{\alpha}{2},n-3}\sqrt{(\frac{x_{10}^2S_{22}-2x_{10}x_{20}S_{12}+x_{20}^2S_{11}}{S_{11}S_{22}-S_{12}^2}+\frac{1}{n}+1)MSE}$$

■ k 個自變數的多元迴歸

驗證性迴歸

- 整體迴歸適合度的檢定
  - 使用ANOVA表 → F檢定 $\begin{cases} H_0:\beta_1=\beta_2=\cdots=\beta_k=0 \\ H_1:\beta_1,\beta_2,\cdots,\beta_k\text{不全為}0 \end{cases}$
  - 使用複相關係數 → t檢定 $\begin{cases} H_0:\rho_{y\hat{y}}=0 \\ H_1:\rho_{y\hat{y}}\neq 0 \end{cases}$

拒絕虛無假設 ↓

- 迴歸係數的檢定
  - 使用檢定統計量 → $t^*=\dfrac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$ $\begin{cases} H_0:\beta_i=0 \\ H_1:\beta_i\neq 0 \end{cases}$
  - 等效檢定
  - 利用偏相關係數檢定
    - → $F^*=\dfrac{MSR(x_2|x_1)}{MSE(x_1,x_2)}$ $\begin{cases} H_0:\rho_{y2.1}=0 \\ H_1:\rho_{y2.1}\neq 0 \end{cases}$
    - → $t^*=\dfrac{r_{y2.1}}{\sqrt{\dfrac{1-r_{y2.1}^2}{n-3}}}$ $\begin{cases} H_0:\beta_2=0 \\ H_1:\beta_2\neq 0 \end{cases}$

探索式迴歸

- 部分項的檢定 — 遞減模式ANOVA表 → F檢定 $\begin{cases} H_0:\beta_1=\beta_2=\cdots=\beta_p=0 \\ H_1:\beta_1,\beta_2,\cdots,\beta_p\text{不全為}0 \end{cases}$

■ 整體迴歸適合度的檢定

◆ 使用 ANOVA 表

● 變異的分解

$$SST=SSR+SSE$$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

● ANOVA 表

變異來源	平方和	自由度	均方和	F
迴歸	SSR	k	$MSR=\dfrac{SSR}{k}$	$F^*=\dfrac{MSR}{MSE}$
誤差	SSE	$n-k-1$	$MSE=\dfrac{SSE}{n-k-1}$	
總和	SST	$n-1$		

$F^*>F_{\alpha,k,n-k-1}$時⇒拒絕虛無假設

◆ 使用複相關係數

$$t^*=\frac{r_{y\hat{y}}}{\sqrt{\dfrac{1-r_{y\hat{y}}^2}{n-k-1}}},\; r_{y\hat{y}}=\frac{\sum(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\sum(y_i-\bar{y})^2}\sqrt{\sum(\hat{y}_i-\bar{\hat{y}})^2}}$$

■ 調整判定係數

$$R_a^2 = 1-(1-R^2)\frac{n-1}{n-k-1}$$

■ 部分項的檢定

◆ 變異數分析表

完整模式 ANOVA 表

變異來源	平方和	自由度
迴歸	SSR_F	k
誤差	SSE_F	$n-k-1$
總和	SST	$n-1$

遞減模式 ANOVA 表

變異來源	平方和	自由度
迴歸	SSR_R	q
誤差	SSE_R	$n-q-1$
總和	SST	$n-1$

$$F_p^* = \frac{\dfrac{SSR_F - SSR_R}{q}}{\dfrac{SSE_F}{n-k-1}}$$

$F_p^* > F_{\alpha,q,n-k-q-1}$時，拒絕虛無假設，表示遞減模式並未顯著的較完整模式好

◆ 迴歸係數的統計推論

$$t^* = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$$

$\left|t^*\right| > t_{\frac{\alpha}{2},n-k-1}$時，拒絕虛無假設，表示該自變數對依變數$y$具解釋力

◆ 迴歸係數的$1-\alpha$信賴區間

$$\hat{\beta}_i - t_{\frac{\alpha}{2},n-k-1}s_{\hat{\beta}_i} \le \beta_i \le \hat{\beta}_i + t_{\frac{\alpha}{2},n-k-1}s_{\hat{\beta}_i}$$

■ 偏判定係數

引進自變數x_2後，x_2對y的偏判定係數

$$R_{y2.1}^2 = \frac{SSR(x_2|x_1)}{SSE(x_1)}$$

引進自變數x_3後，x_3對y的偏判定係數

$$R_{y3.12}^2 = \frac{SSR(x_3|x_1,x_2)}{SSE(x_1,x_2)}$$

■ 偏相關係數

在x_2固定的條件下，x_1與y的偏相關係數

$$r_{y1.2} = \frac{r_{yx_1} - r_{yx_2}\cdot r_{x_1x_2}}{\sqrt{1-r_{yx_2}^2}\sqrt{1-r_{x_1x_2}^2}}$$

■ 偏判定係數與偏相關係數之關係

$$r_{y1.2} = \pm\sqrt{R^2_{y1.2}}\ \text{(正負號與 } \hat{\beta}_1 \text{ 相同)}$$

$$r_{y3.21} = \pm\sqrt{R^2_{y3.21}}\ \text{(正負號與 } \hat{\beta}_3 \text{ 相同)}$$

- 偏相關係數的檢定
  - 引進自變數 x_2
    - F 檢定

$$F^* = \frac{MSR(x_2 | x_1)}{MSE(x_1, x_2)}$$

當 $F^* > F_{\alpha,1,n-3}$ 時拒絕虛無假設

    - t 檢定

$$t^* = \frac{r_{y2.1}}{\sqrt{\dfrac{1-r^2_{y2.1}}{n-3}}} \text{ 或 } t^* = \frac{\hat{\beta}_2}{s_{\hat{\beta}_2}}$$

當 $|t^*| > r_{\frac{\alpha}{2},n-3}$ 時拒絕虛無假設

  - 引進自變數 x_3
    - F 檢定

$$F^* = \frac{MSR(x_3 | x_1, x_2)}{MSE(x_1, x_2, x_3)}$$

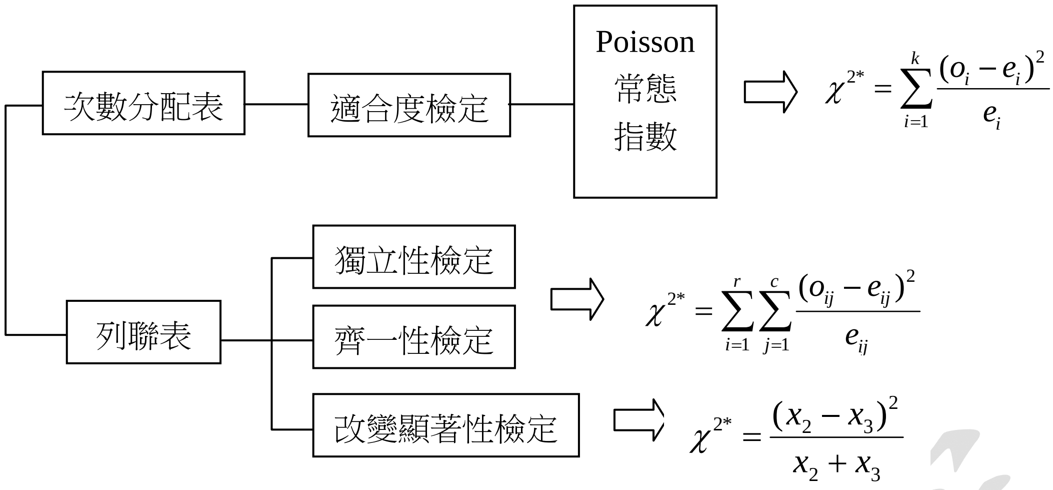
$F^* > F_{\alpha,1,n-4}$ 時拒絕虛無假設。

    - t 檢定

$$t^* = \frac{r_{y3.12}}{\sqrt{\dfrac{1-r^2_{y3.12}}{n-4}}} \text{ 或 } t^* = \frac{\hat{\beta}_3}{s_{\hat{\beta}_3}}$$

當 $|t^*| > t_{\frac{\alpha}{2},n-3}$ 時拒絕虛無假設

拾玖、類別資料的分析



- 卡方檢定用在類別型態資料的檢定
  - 母體適合度檢定

    檢定母體是否服從某種分配，一般而言，卡方檢定可適用於任何分配形狀的母體。
  - 獨立性檢定

    檢定兩分類型態的變數是否具相關性，並且衡量相關程度。
  - 卡方齊一性檢定

    檢定兩個或兩個以上隨機樣本，是否來自同一個多項分配，或百分比一致。
  - **改變顯著性檢定**

    以列聯表的方式，檢定兩母體比例差是否相等、大於或等於 0

- 關聯性的衡量
  - ϕ 相關係數

    列聯表為 2×2 時

$$\phi=\sqrt{\frac{\chi^{2*}}{n}}$$

  - 列聯係數：

    列聯表為 $3\times 3, 4\times 4, 5\times 5, \cdots$，呈方陣形式者

$$C=\sqrt{\frac{\chi^{2*}}{\chi^{2*}+n}}$$

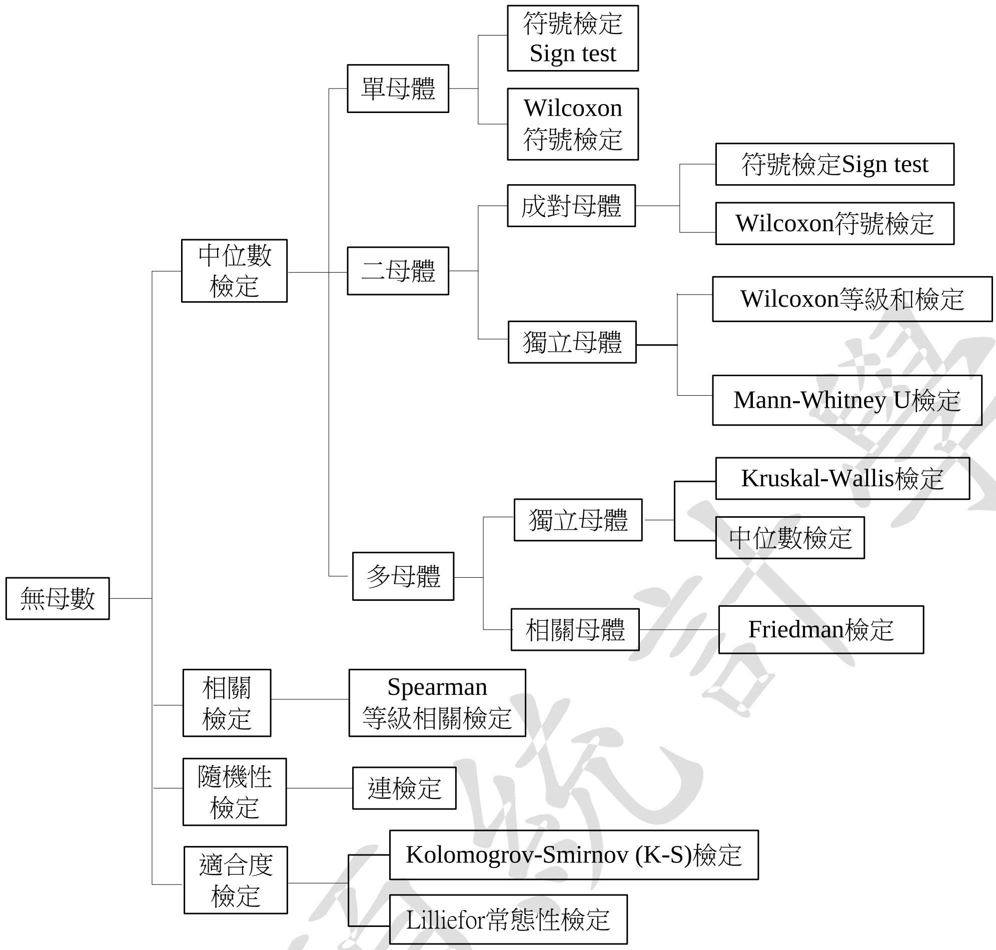
  - Crame's V_c：

    列聯表的行數與列數不相等時

$$V_c=\sqrt{\frac{\phi^2}{\min(i-1, j-1)}}$$

貳拾、無母數統計

■ 整理


無母數
中位數檢定
單母體
符號檢定 Sign test
Wilcoxon 符號檢定
二母體
成對母體
符號檢定Sign test
Wilcoxon符號檢定
獨立母體
Wilcoxon等級和檢定
Mann-Whitney U檢定
多母體
獨立母體
Kruskal-Wallis檢定
中位數檢定
相關母體
Friedman檢定
相關檢定
Spearman 等級相關檢定
隨機性檢定
連檢定
適合度檢定
Kolomogrov-Smirnov (K-S)檢定
Lilliefor常態性檢定

■ 母數與無母數統計之方法對照表

檢定類型		母數統計	無母數統計
單母體	中央趨勢	$H_0:\mu$ (≤或=或≥) μ_0 z 檢定 t 檢定	$H_0:\eta$ (≤或=或≥) η_0 符號檢定 Wilcoxon 符號檢定
	母體比例	$H_0:p$ (≤或=或≥) p_0 z 檢定 以二項分配求 P_value	$H_0:p$ (≤或=或≥) p_0 F 檢定
	母體變異數	$H_0:\sigma^2$ (≤或=或≥) σ_0^2 卡方檢定	
	隨機性檢定		H_0 : 樣本為隨機樣本 連檢定
二母體	中央趨勢	$H_0:\mu_1$ (≤或=或≥) μ_2 $H_0:\mu_d$ (≤或=或≥)0 z 檢定 t 檢定	$H_0:\eta_1$ (≤或=或≥) η_2 成對： 符號檢定 Wilcoxon 符號檢定 獨立： Wilcoxon 等級和檢定 Mann-Whitney U 檢定
	母體比例	$H_0:p_1$ (≤或=或≥) p_2 z 檢定	$H_0:p_1=p_2$ 獨立： 卡方齊一性檢定 不獨立： 改變顯著性檢定
	母體變異數	$H_0:\sigma_1^2$ (≤或=或≥) σ_2^2 F 檢定	
	合度檢定		H_0 : 樣本分配與特定分配相同 卡方適合度檢定 K-S 檢定 Lilliefor 常態性檢定
	相關性檢定	$H_0:\rho_{xy}$ (≤或=或≥)0 Person 相關性檢定	H_0 : A類別與B類別無關 卡方獨立性檢定(類別變項) $H_0:\rho_s$ (≤或=或≥)0 等級相關檢定(順序量尺)
多母體	中央趨勢	$H_0:\mu_1=\mu_2=\mu_3=...$ 變異數分析⇒F 檢定	$H_0:\eta_1=\eta_2=\eta_3=...$ 中位數檢定 K-W 檢定 Friedman 檢定
	母體比例		$H_0:p_{1j}=p_{2j}=\cdots$ 卡方齊一性檢定

	母體變異數	Hartley 檢定 Bartelett 檢定	

■ 單母體中位數的檢定

◆ 符號檢定

● 小樣本的符號檢定

◎ 右尾檢定

$$P-value = f(x \ge k) = \sum_{x=k}^{n} C_x^n (0.5)^x (0.5)^{n-x}$$

$P-value < \alpha \Rightarrow$ 拒絕虛無假設。

◎ 左尾檢定

$$P-value = f(x \le k) = \sum_{x=0}^{k} C_x^n (0.5)^x (0.5)^{n-x}$$

$P-value < \alpha \Rightarrow$ 拒絕虛無假設。

◎ 雙尾檢定

1. $k > \frac{n}{2}$時

$$P-value = 2f(x \ge k) = 2\sum_{x=k}^{n} C_x^n (0.5)^x (0.5)^{n-x}$$

$P-value < \alpha \Rightarrow$ 拒絕虛無假設。

2. $k < \frac{n}{2}$時

$$P-value = 2f(x \le k) = 2\sum_{x=0}^{k} C_x^n (0.5)^x (0.5)^{n-x}$$

$P-value < \alpha \Rightarrow$ 拒絕虛無假設

● 大樣本的符號檢定

$$z^* = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5 \times 0.5}{n}}}$$

◆ Wilcoxon 符號等級檢定

● 小樣本檢定

$$R^* = \min(R^+, R^-)$$

$$R^+ = \sum_{D_i > 0} Rank(|D_i|),\ R^- = \sum_{D_i < 0} Rank(|D_i|)$$

● 大樣本檢定

$$z^* = \frac{R^* - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

■ 成對母體中位數檢定

與單母體中位數的檢定法相同

■ 兩獨立母體中位數的檢定

◆ Wilcoxon 等級和檢定

● 小樣本檢定

W^* =樣本數較小的那組等級和

● 大樣本檢定

$$z^* = \frac{W^* - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}}$$

◆ Mann-Whitney U 檢定

● 小樣本檢定

$$U^* = \min(U_1, U_2)$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W_1 . U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - W_2$$

雙尾： $P\_value = P(U < U^*) < \frac{\alpha}{2}$ ，則拒絕虛無假設

單尾： $P\_value < \alpha$ ，則拒絕虛無假設

● 大樣本檢定

$$z^* = \frac{U^* - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}}$$

■ 多母體檢定中位數中位數的檢定

◆ 中位數檢定

$$\chi^{2*} = \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

當 $\chi^{2*} > \chi^2_{\alpha,(r-1)(c-1)}$ 時，拒絕 H_0

◆ 、Kruskal-Wallis（K-W）檢定

$$K^* = \frac{SSA}{SST/n-1} = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

當 $K^* > \chi^2_{(k-1),\alpha} \Rightarrow$ 拒絕 H_0

$$SST = \sum_{i=1}^{n_i} \sum_{j=1}^{k} R_{ij}^2 - n\bar{\bar{R}}^2 = \frac{1}{6} n(n+1)(2n+1) - n \times \frac{1}{4}(n+1)^2 = \frac{n(n+1)(n-1)}{12}$$

$$SSA = \sum_{j=1}^{k} n_j (\bar{R}_j - \bar{\bar{R}})^2 = \sum_{j=1}^{k} \frac{R_j^2}{n_j} - \frac{n(n+1)^2}{4}$$

◆ Friedman 檢定

$$F_r^* = \frac{12}{rc(c+1)} \sum_{j=1}^{c} R_j^2 - 3r(c+1)$$

當 $F_r^* > \chi^2_{\alpha,c-1}$ 時，拒絕虛無假設

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{c}(R_{ij} - \overline{\overline{R}})^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}R_{ij}{}^2 - rc\overline{\overline{R}}^2 = \frac{rc(c+1)(c-1)}{12}$$

$$SSA = \sum_{i=1}^{r}\sum_{j=1}^{c}(\overline{R}_j - \overline{\overline{R}})^2 = r\sum_{j=1}^{c}(\overline{R}_j - \overline{\overline{R}})^2 = \frac{1}{r}\sum_{j=1}^{c}R_j^2 - \frac{rc(c+1)^2}{4}$$

- 母體分配檢定
  - Kolomogrov-Smirnov (K-S)檢定

$$D^* = \max\left|F(x_i) - S(x_i)\right|$$

$$\text{若 } D^* > D_{\frac{\alpha}{2},n} \Rightarrow \text{拒絕 } H_0$$

  - Lilliefor 常態性檢定

$$D^* = \max\left|F(x_i) - S(x_i)\right|$$

$$\text{若 } D^* > D_{\frac{\alpha}{2},n} \Rightarrow \text{拒絕 } H_0$$

- 隨機性檢定-連檢定
  - 小樣本$(n_1 + n_2 < 20)$

R^* = 資料相鄰且符號相同的個數

(1) $R^* \ge \dfrac{2n_1n_2}{n_1+n_2} + 1$時

$P - value = 2 \times P(R \ge R^*)$

(2) $R^* < \dfrac{2n_1n_2}{n_1+n_2} + 1$

$P - value = 2 \times P(R \le R^*)$

$P - value < \alpha$ 時，拒絕虛無假設。

  - 大樣本$(n_1, n_2 \ge 10)$

$$z^* = \frac{R^* - \dfrac{2n_1n_2}{n_1+n_2} + 1}{\sqrt{\dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}}$$

- Spearman 等級相關檢定
  - $4 \le n \le 30$

$$r_s^* = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

  - $10 \le n < 30$

$$t^* = \frac{r_s}{\sqrt{\dfrac{1-r_s^2}{n-2}}}$$

  - $n \ge 30$

$$z^* = \frac{r_s}{\sqrt{\dfrac{1}{n-1}}}$$