

高點文教機構



高點建國課程折抵金

e-coupon 500元

- 持本券報名班內面授/VOD課程(2500元以上)；函授課程(8000元以上)可享優惠價後再折抵現金500元。
- 本券限親洽各點櫃台報名抵用，請於報名時主動出示本券，已報名者不得追溯使用亦不可兌換現金。
- 單一課程限抵用一張，並不得與其他行銷活動併用。
- 本班保有調整活動辦法之權利及最終解釋權。
- 使用期限：至105年8月31日止。

◎其他使用細節請詳洽各分班。

生物統計學 (含概要)

王凡老師提供

一、單一母體假設檢定

(1) 平均數 \bar{X} 前提：母體為常態分配	母體變異數 σ^2 已知	\bar{X} 的抽樣分配： $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
	母體變異數 σ^2 未知，以樣本變異數 S^2 估計 σ^2	大樣本： $\bar{X} \sim Z$ 分配 小樣本： $\bar{X} \sim t_{n-1}$
(2) 比例 \hat{p}	大樣本時， \hat{p} 的抽樣分配為： $\hat{p} \sim N\left(p, \frac{p \times (1-p)}{n}\right)$	若母體比例 p 未知，此時可用樣本比例 \hat{p} 估計 p
(3) 變異數 S^2 前提：母體為常態	$\chi^2 = \frac{(n-1) \times S^2}{\sigma^2} \sim \chi_{n-1}^2$	

二、兩母體假設檢定

(1)

平均數差值
 $\bar{X}_1 - \bar{X}_2$
 前提：兩母體均為常態分配

兩母體變異數 σ_1^2, σ_2^2 已知

兩母體變異數未知但假設為相同：

以「混合樣本變異數」 S_p^2 估計 σ^2

較準確， $\bar{X}_1 - \bar{X}_2$ 標準誤 $S.E. = \left(\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}\right)^{1/2}$

兩母體變異數未知且不同：

以 S_1^2, S_2^2 分別估計 σ_1^2, σ_2^2 ，

$\bar{X}_1 - \bar{X}_2$ 標準誤 $S.E. = \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^{1/2}$

$\bar{X}_1 - \bar{X}_2$ 的抽樣分配：

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

大樣本：

$\bar{X} \sim Z$ 分配

小樣本：

$$\bar{X} \sim t_{n-1}$$

大樣本：

$\bar{X} \sim Z$ 分配

小樣本：

$$\bar{X} \sim t_{n-1}$$

(2)

比例差
 $\hat{p}_1 - \hat{p}_2$

大樣本時， $\hat{p}_1 - \hat{p}_2$ 的抽樣分配為：

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 \times (1-p_1)}{n_1} + \frac{p_2 \times (1-p_2)}{n_2}\right)$$

若母體比例 p_1, p_2 未知，此時可用樣本比例 \hat{p}_1, \hat{p}_2 估計 p_1, p_2

若母體比例 p_1, p_2 未知但假設為相同，此時可用「混合樣本比例」 p^* 估計 p_1, p_2

(3)

兩變異數之比 S_1^2, S_2^2

前提：兩母體為常態

$$F = \frac{\frac{\chi_1^2 / v_1}{\chi_2^2 / v_2} = \frac{S_1^2 / S_2^2}{\sigma_1^2 / \sigma_2^2} \sim F_{v_1, v_2}$$

【版權所有，翻印必究】

****特殊情況****

1. 混合樣本變異數：

$$\text{混合樣本變異數 } S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

2. 混合樣本比例：

$$\text{混合樣本比例 } \bar{p} = \frac{n_1 \times \hat{p}_1 + n_2 \times \hat{p}_2}{n_1 + n_2}$$

【版權所有，翻印必究】

三、ANOVA

1. ANOVA 的變異來源分析：

$$\begin{aligned} \text{總變異} &= \text{組內變異} + \text{組間變異} \\ \text{SST} &= \text{SSW} + \text{SSB} \\ \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_i - \bar{Y})^2 \end{aligned}$$

2. ANOVA 的變異數分析表

變異來源	平方和(SS)	自由度(df)	平均平方和 (變異數)	F 值
組間 (因子)	SSB (SSF)	k-1	MSB=SSB/k-1	F=MSB/MSW 與 F 臨界值 比較?
組內 (隨機)	SSW (SSE)	n-k	MSW=SSW/n-k	
總和	SST	n-1		

四、簡單直線迴歸分析與相關分析

1. 簡單直線迴歸估計式：

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i$$

2. 迴歸線斜率 \hat{b}_1 :

$$\hat{b}_1 = \frac{\sum (X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{Cov(X, Y)}{(S_x)^2}$$

\hat{b}_1 的抽樣分配為 :

$$\hat{b}_1 \sim N\left(b_1, \frac{\sigma^2}{\sum (X_i - \bar{X})^2}\right)$$

故得 \hat{b}_1 的檢定方法為 :

母體變異數已知，Z 檢定

母體變異數未知，t 檢定

$$\hat{b}_1 \frac{\hat{b}_1 - b_1}{\frac{\sigma}{\sqrt{\sum (X_i - \bar{X})^2}}} \sim Z \quad \hat{b}_1 \frac{\hat{b}_1 - b_1}{\frac{S_{Y|X}}{\sqrt{\sum (X_i - \bar{X})^2}}} \sim t_{n-2}$$

其中 $S_{Y|X}$ 為「樣本殘差變異數」，用以估計未知的母體變異數：

$$S_{Y|X}^2 = \frac{\sum \varepsilon^2}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}$$

3. 決定係數 R^2 : 有多少百分比的變異來自簡單直線迴歸模型。

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = (\hat{b}_1)^2 \cdot \frac{\sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2} = (\hat{b}_1)^2 \cdot \frac{S_X^2}{S_Y^2}$$

4. 簡單直線迴歸的變異來源分析：

總變異 = 由誤差項解釋的部分 + 由迴歸線解釋的部分

$$\begin{aligned} SST &= SSE + SSR \\ \sum_{i=1}^n (Y_i - \bar{Y})^2 &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \end{aligned}$$

5. 簡單直線迴歸模型的變異數分析表

變異來源	平方和(SS)	自由度(df)	平均平方和 (變異數)	F 值
迴歸模型	SSR	1	MSR=SSR/1	F=MSR/MSW 與 F 臨界值 比較?
誤差	SSW	n-2	MSW=SSW/n-1	
總和	SST	n-1		

6. 相關係數 r_{xy} :

$$r_{xy} = \frac{\sum (X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \times \sum (Y_i - \bar{Y})^2}} = \frac{Cov(X, Y)}{S_x \cdot S_y}$$

7. 迴歸估計式斜率 \hat{b}_1 、相關係數 r_{xy} 與決定係數 R^2 之間的關係：

$$\hat{b}_1 = r_{xy} \cdot \frac{S_y}{S_x}$$

$$R^2 = (r_{xy})^2$$

五、重要名詞解釋

中央極限定理	不論母體的分配型態為何，當抽樣樣本數夠大時，樣本統計量的分配型態會趨近常態分配。
標準誤	樣本統計量的抽樣分配的標準差。
95%信賴區間	一段由上界和下界包圍起來的區間，此區間包含母體參數的機率稱為信心水準。當信賴區間包含母體參數的機率為 95%時，謂之「95%信賴區間」。
虛無假設	某個關於母體的假設或主張，這個假設通常是研究者想要否定、推翻的對象，謂之虛無假設。
α 值（顯著水準）	在 H_0 為真的情形下，研究者依照其主觀判斷，而拒絕 H_0 的機率，亦即研究者甘願承受多少犯下型一錯誤的機會。 α 值可由研究者主觀訂定。

P 值	發生比某個觀察值還要極端的累加機率，叫 p 值。
β 值	在 H_1 為真的情形下，研究者依照其主觀判斷，而拒絕 H_1 的機率，亦即研究者甘願承受多少犯下型二錯誤的機會。
檢力	$1 - \beta$ 為檢力，亦即從罹患疾病的人當中，篩檢工具能正確找出多少比例罹患疾病的人。「檢力」就是篩檢工具的「敏感度」。
迴歸係數 β_1	自變項 X 變化一單位，依變項 Y 相對應會改變多少單位。
決定係數 R^2	總變異中有多少百分比可以由迴歸線來解釋。
相關係數	用來評判 X 與 Y 相關性的強度指標。

【版權所有，翻印必究】