103 年公務人員高考三等 衛生行政試題

等別：三等考試
類科：衛生行政
科目：生物統計學

一、某研究團隊想評估運動、飲食控制對減肥的功效，他們收集 250 名肥胖個案，並將 250 名個案分配到四組，第一組個案接受每天 40 分鐘運動，第二組個案接受飲食控制，第三組個案同時接受每天 40 分鐘運動及飲食控制，第四組個案沒有接受運動或飲食控制，經過一年後，評估這四組個案減重的成效，得到下列資料：

組別	介入前身體質量指數		介入後身體質量指數		介入後身體質量指數改變量	
	平均值	標準差	平均值	標準差	平均值	標準差
運動組(60人)	28.3	2.3	25.4	2.6	2.9	2.5
飲食控制組(62人)	27.9	2.5	25.8	2.9	2.1	2.7
運動及飲食控制組(61人)	28.1	2.7	24.9	3.1	3.2	2.9
未接受運動或飲食控制(67人)	28.2	2.1	28.3	2.3	-0.1	2.2

㈠請用適當的統計方法檢定運動及飲食控制組其介入前、後的身體質量指數改變量平均值是否有統計上顯著不同？設第一型誤差 $\alpha = 0.05$。並計算其身體質量指數平均值差異的 95% 信賴區間。(15 分)

㈡請用適當的統計檢定方法比較經介入後四組個案的身體質量指數平均值是否有統計上顯著不同？請列出虛無假設及對立假設，設第一型誤差 $\alpha = 0.05$ 。(25 分)

擬答：

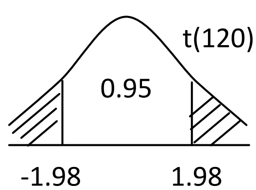
㈠設運動組 BMI 改變量為 X_i ，飲食控制組 BMI 改變量為 Y_i

$$H_0 : \mu_x = \mu_y \quad H_1 : \mu_x \neq \mu_y$$

$$\alpha = 0.05$$

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} = \frac{59 \cdot 2.5^2 + 61 \cdot 2.7^2}{60+62-2} = 6.7787$$

$$T^* = \frac{\overline{X} - \overline{Y}}{\sqrt{s_p^2\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{2.9-2.1}{\sqrt{6.7787\left(\frac{1}{60}+\frac{1}{62}\right)}} = 1.697 \notin C$$



不拒絕 H_0

沒有顯著的證據兩組 BMI 有差

運動組與飲食控制組 BMI 改變量差值 95%信賴區間為

$$(\bar{X}-\bar{Y}) \pm t_{\alpha/2}(n+m-2)\sqrt{s_p^2\left(\frac{1}{n}+\frac{1}{m}\right)}$$

$$\Rightarrow (2.9-2.1) \pm 1.98\sqrt{6.7787\left(\frac{1}{60}+\frac{1}{62}\right)}$$

$$\Rightarrow [-0.1336, 1.7336]$$

㈡ $SST = \sum\sum(\bar{X}_{i}.-\bar{X}..)^2 = \sum n_i(\bar{X}_{i}.-\bar{X}..)^2 = 446.36016$

$SSE = \sum\sum(X_{ij}-\bar{X}_{i}.)^2 = \sum(n_i-1)S_i^2$

$= (60-1)\times 2.6^2 + (62-1)\times 2.9^2 + (61-1)\times 3.1^2 + (67-1)\times 2.3^2$

$= 1837.59$

$SSTO = SST + SSE = 2283.95016$

ANOVA 表

	SS	df	MS	F
Treatment	446.36016	3	148.7867	19.918
Error	1837.59	246	7.4699	
Total	2283.95016	249		

H_0 :四組個案介入後 BMI 無差異

H_1 :四組個案介入後 BMI 有差異

$\alpha = 0.05$

$F^* = 19.918 \in C$ ，拒絕 H_0

$C = \{F^* > F_{0.05}(3,246) = 2.64\}$

有顯著證據說四組個案介入後 BMI 有差異

二、某研究想瞭解肥胖與糖尿病發生率的關係，收集 56 名肥胖個案及 95 名非肥胖個案，結果發現，肥胖個案中 22 名有糖尿病，非肥胖個案中 20 名有糖尿病，請以適當統計方法檢定肥胖組與非肥胖組的糖尿病發生率是否有統計上顯著差異？設第一型誤差 $\alpha = 0.05$；並計算肥胖組糖尿病發生率的 95%信賴區間。（15 分）

擬答：

㈠假設肥胖個案有糖尿病的發生率為 p_1

非肥胖個案有糖尿病的發生率為 p_2

$$\hat{p}_1 = \frac{\sum x_i}{n} = \frac{22}{56} \qquad \hat{p}_2 = \frac{\sum y_i}{m} = \frac{20}{95}$$

$$\hat{p} = \frac{\sum x_i + \sum y_i}{n+m} = \frac{22+20}{56+95} = \frac{42}{151}$$

$H_0: p_1 = p_2 \quad H_1: p_1 \neq p_2$

$\alpha = 0.05$

$$Z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{22}{56}-\frac{20}{95}}{\sqrt{\frac{42}{151}\times\frac{109}{151}\left(\frac{1}{56}+\frac{1}{95}\right)}} = 2.42 \in C$$

$C = \{|Z^*| > Z_{0.025} = 1.96\}$

拒絕 H_0 ，有顯著的證據說

肥胖個案與非肥胖個案的糖尿病發生率有差異

㈡肥胖組糖尿病發生率的 95%信賴區間為

$$\hat{p} \pm z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\Rightarrow \frac{22}{56} \pm 1.96 \cdot \sqrt{\frac{\frac{22}{56}\times\frac{34}{56}}{56}}$$

$$\Rightarrow [0.2649, 0.5208]$$

三、某研究團隊想研究國小男女學童的齲齒率是否有不同，分別檢查 150 名男學童及 145 名女學童，結果發現，男學童 60 名有齲齒，女學童 40 名有齲齒情形，請以適當統計方法檢定男女學童的齲齒率是否有統計上顯著不同？設第一型誤差 $\alpha = 0.05$；並計算男女學童的齲齒率差異的 95%信賴區間。(15 分)

擬答：

㈠假設男學童的齲齒率為 p_1 ，女學童的齲齒率為 p_2

$$\hat{p}_1 = \frac{\sum x_i}{n} = \frac{60}{150} \qquad \hat{p}_2 = \frac{\sum y_i}{m} = \frac{40}{145}$$

$$\hat{p} = \frac{\sum x_i + \sum y_i}{n+m} = \frac{60+40}{150+145} = \frac{100}{295}$$

$H_0: p_1 = p_2 \quad H_1: p_1 \neq p_2$

$\alpha = 0.05$

$$Z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}} = \frac{\frac{60}{150}-\frac{40}{145}}{\sqrt{\frac{100}{295}\times\frac{195}{295}\left(\frac{1}{150}+\frac{1}{145}\right)}} = 2.252 \in C$$

$C = \{|Z^*| > Z_{0.025} = 1.96\}$

拒絕 H_0 ，有顯著的證據說

男學童與女學童的齲齒率有差異

㈡男女學童的齲齒率差異的 95%信賴區間為

$$(\hat{p}_1-\hat{p}_2)\pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}+\frac{\hat{p}_2(1-\hat{p}_2)}{m}}$$

$$\Rightarrow (\frac{60}{150}-\frac{40}{145})\pm 1.96\sqrt{\left(\frac{\frac{60}{150}\times\frac{90}{150}}{150}+\frac{\frac{40}{145}\times\frac{105}{145}}{145}\right)}$$

$$\Rightarrow [0.0172, 0.2311]$$

四、某研究想要評估腰圍(公分)與空腹血糖(mg/dL)的關係，研究結果得到腰圍與空腹血糖間的皮爾森相關係數為 0.447；另外也得到簡單線性迴歸模式及變異數分析表如下：

$$\hat{Y}=-20.04+1.54X$$

變異數分析表

變異來源	平方和	自由度	均方	F值	P值
迴歸模型	42744.6	1	42744.6	37.1	<0.001
殘差	171627.8	149	1151.9		
總和	214372.3	150			

請回答下列問題：

㈠請說明皮爾森相關係數與迴歸係數的關係及兩者間的差異。(6 分)

㈡請用適當的統計方法檢定腰圍與空腹血糖間的皮爾森相關係數是否有統計上的顯著性？設第一型誤差 $\alpha=0.05$ (8 分)

㈢請解釋上述簡單線性迴歸模式？並說明此迴歸模式是否具統計上的顯著性？(設第一型誤差 $\alpha=0.05$)，並求出當腰圍為 78 時，其預期的空腹血糖值為何？(10 分)

㈣請求出上述簡單線性迴歸模式的決定係數，並解釋該決定係數的意義。(6 分)

擬答：

㈠皮爾森相關係數可用來描述兩屬量變項之間正負變動關係與關聯性強弱，當相關係數的絕對值越大，兩變項之間的關聯性越強。迴歸係數代表以自變數預測應變數的斜率項，代表每增加一單位的自變數 X，可增加 β_1 單位的應變數 Y。

而兩者之間的關係為 $\hat{\beta}_1=r_{X,Y}\dfrac{S_Y}{S_X}$

可看出迴歸係數 $\hat{\beta}_1$ 與相關係數 $r_{X,Y}$ 同號。

㈡ $H_0\ \rho=0\quad H_1\ \rho\neq 0$

$\alpha=0.05$

$$T^* = r\sqrt{\frac{n-2}{1-r^2}} = 0.447 \times \sqrt{\frac{151-2}{1-0.447^2}} = 6.1 \in C$$

$$C = \{|T^*| > t_{0.025}(149) \approx 1.96\}$$

拒絕 H_0，有顯著證據相關係數不為 0

註：事實上不需要作檢定，因為此檢定與迴歸分析之 F 統計量所對應之 P 值等價，故此處之 p-value<0.001，相關係數顯著不為 0

(三) $\hat{Y} = -20.04 + 1.54X$

代表以腰圍預測空腹血糖值，每增加一公分的腰圍，會增加 1.54 mg/dL 的空腹血糖值。所以當 $X = 78$，空腹血糖預測值為

$\hat{Y}_{X=78} = -20.04 + 1.54 \times 78 = 100.08$ mg/dL

$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$

$\alpha = 0.05$

$F^* = (T^*)^2 = 37.1 \in C$

且 p-value<0.001，代表迴歸係數顯著不為 0，即迴歸模式模式顯著。

(四) $R^2 = r^2 = 0.447^2 = 19.9809\%$

R^2 判定係數代表迴歸模式的解釋度

即以腰圍來解釋空腹血糖值時，有 19.9809%的解釋度