統計、經行

《統計學》

試題評析
1.除One-Way ANOVA各組樣本數不等之試題不等之試題稍微偏難之外，其餘試題均為常見題型，取分不難。
2.指數分布的無記憶性特性應用，若考生不熟悉相關題型，此題雖易，但也可能失分。
3.綜合來說，一般同學拿60分以上不成問題，程度較佳同學應可拿80分以上。

一、直排輪運動傷害研究，以過去一年內全國直排輪運動人口為母體，隨機抽出250位做問卷調查。資料顯示，87位受訪者運動時帶護肘其中10位手肘受傷，163位運動時未帶護肘的受訪者中35位手肘受傷。欲檢定護肘是否較能保護直排輪運動者。
(一)請寫出適當的檢定虛無假設與對立假設。（六分）
(二)請求出(一)中檢定的p–值。（十四分）

答：

(一)令p_1表示帶護肘的人手肘受傷之比例
p_2表示未帶護討的人手肘受傷之比例

$$\Rightarrow \begin{cases} H_0: p_1 \ge p_2 \\ H_1: p_1 < p_2 \end{cases}$$

(二)1. $\begin{cases} H_0: p_1 \ge p_2 \\ H_1: p_1 < p_2 \end{cases}$

2.檢定統計量 $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$

$$= \frac{\dfrac{10}{87}-\dfrac{35}{163}}{\sqrt{\dfrac{10+35}{87+163}\left(1-\dfrac{10+35}{87+163}\right)\left(\dfrac{1}{87}+\dfrac{1}{163}\right)}}$$

≈ -1.96

3.p–值 $= p(z < -1.96) = 0.025$

二、投擲公正硬幣一萬次，試以柴比雪夫（Chebyshev）不等式，對「出現正面的比率介於0.49與0.51之間」這一事件的機率給一個下界。（二十分）

答：

(一)令X_i表丟第i次硬幣之結果

$$X_i = \begin{cases} 1, & \text{第次投擲結果為正面} \\ 0, & \text{第次投擲結果為負面} \end{cases}$$

$$X_i \overset{iid}{\sim} Ber(p) \quad i = 1,2,\cdots,10000$$

而 $\bar{X} = \dfrac{\sum X_i}{n} = \hat{p} \Rightarrow \begin{cases} E(\bar{X}) = p \\ Var(\bar{X}) = \dfrac{1}{n}p(1-p) \end{cases}$

(二)由柴原比雪夫不等式得知

$$p\left[|\bar{X}-p| \le k \cdot \frac{1}{n}p(1-p)\right] \ge 1 - \frac{1}{k^2}$$

其中 $|\bar{X}-p| \le k \cdot \dfrac{1}{n}p(1-p)$

$$\Rightarrow p - \frac{k}{n}p(1-p) \le \bar{X} \le p + \frac{k}{n}p(1-p)$$

$$\Rightarrow \begin{cases} p-\frac{k}{n}p(1-p)=0.49 \quad (n=10000) \\ p+\frac{k}{n}p(1-p)=0.51 \end{cases}$$

$$\Rightarrow \begin{cases} p=0.5 \\ k=400 \end{cases}$$

$$\therefore p(0.49\le \overline{X}\le 0.51)\ge 1-\frac{1}{25}=\frac{624}{625}=\frac{159,999}{160,000}$$

亦即 $p(0.49\le \hat{p}\le 0.51)\ge \frac{624}{625}=\frac{159,999}{160,000}$

三、考慮簡單線性迴歸模型$y_i=\alpha+\beta x_i+\varepsilon_i$，$i=1, \cdots, n$。

(一)已知$\hat{\alpha}=2$，**且**$(\overline{x},\overline{y})=(2,10)$。**請求出預測方程式，並請解釋**$\beta$**的意義。（十四分）**

(二)假定$n=5$，**且殘差**$\hat{\varepsilon}_1=5$，$\hat{\varepsilon}_2=-2$，$\hat{\varepsilon}_3=0$，$\hat{\varepsilon}_4=-1$，**試求**$\hat{\varepsilon}_5=$ **？（六分）**

答：

(一)1. $\hat{\alpha}=\overline{y}-\hat{\beta}\overline{x}$

$\Rightarrow 2=10-\hat{\beta}\cdot 2\Rightarrow \beta=4$

$\therefore$預測方程式為$y=2+4x$

2. $\hat{\beta}$代表x每增加一單位，y隨之變化的變化量

(二)$\because \sum\hat{\varepsilon}_i=0$

$\therefore 5+2+0-1+\hat{\varepsilon}_5=0$

$\Rightarrow \hat{\varepsilon}_5=-6$

四、某路公車班車之間的時間間隔具有指數分布（Exponential Distribution），其望值10分鐘。假設你到站時，前班車於5分鐘前離去，試求你等車時間的期望值（cxpectiation）。（二十分）

答：

令X代表等車時間

依題意，$X\rightarrow \sum_{XP}(\lambda=\frac{1}{10})$；且E(X)=10

則 $\int(X \mid X_{25})=\frac{f(x)}{P\{X_{25}\}}$

$$=\frac{\frac{1}{10}e^{-\frac{x}{10}}}{e^{-\frac{5}{10}}}$$

$$=\frac{1}{10}e^{-\frac{(x-5)}{10}}\text{；}x>5$$

$$\therefore E(X \mid X_{25})=\int_5^\infty X\cdot\frac{1}{10}e^{-\frac{(x-5)}{10}}dx$$

$=15$

五、設一因子設計的統計模型：$Y_{ij}-\mu+\alpha_i+\delta_{ij}$，$\delta_{ij}\sim N(0,\sigma^2)$，$i=1,2,\cdots,\alpha$，$j=1,2,\cdots,n$**獨立，及**$\sum_{i=1}^{n}n_i\alpha_i=0$。**由此一因子實驗所收集到的資料中**$n_1=3$，$n_2=4$，$n_3=3$，$\alpha=3$**及**

樣本1	**樣本2**	**樣本3**
15.4	11.2	14.9
15.2	11.4	10.8
18.6	15.8	12.7
	10.4	
$\overline{y}_1=16.4$	$\overline{y}_2=12.2$	$\overline{y}_3=12.8$
$s_1^2=3.64$	$s_2^2=3.96$	$s_3^2=4.21$

(一)求α_i的最小平方估計量。（七分）
(二)試在顯著水準$\alpha = 0.05$下，檢定H_0：$\alpha_1 - \alpha_2 - \alpha_1 - 0$。（七分）
(三)求$\alpha_1 - \alpha_2$的95%信賴區間。（六分）
（已知：$F_{0.05}(3, 7) - 4.35, F_{0.05}(2, 7) = 4.7374, t_{0.025}(3) = 3.182, t_{0.025}(2) - 4.303, t_{0.025}(7) - 2.365$）

答：

(一)1.令 $E = \sum_i \sum_j (Y_{ij} - \mu - \alpha_i)^2$

$$\Rightarrow \begin{cases} 0 = \dfrac{\partial E}{\partial \alpha_i} = -2\sum_j (Y_{ij} - \mu - \alpha_i) & \cdots\cdots\cdots\cdots(a) \\ 0 = \dfrac{\partial E}{\partial \mu} = -2\sum_j (Y_{ij} - \mu - \alpha_i) \end{cases}$$

2.(1) $\because \sum_i n_i \alpha_i = 0$

$$\therefore 0 = \sum_i \sum_j (Y_{ij} - \mu - \alpha_i) = \sum_i \left[\sum_j Y_{ij} - \sum_j \mu - \sum_j \alpha_i\right]$$
$$= \sum_i \left[\sum_j Y_{ij} - n_i\mu - n_i\alpha_i\right]$$
$$= \sum_i \sum_j Y_{ij} - \mu\sum_i n_i - \sum_i n_i\alpha_i$$
$$= \sum_i \sum_j Y_{ij} - \mu\sum_i n_i$$

$$\Rightarrow \hat{\mu} = \frac{\sum_i \sum_j Y_{ij}}{\sum n_i} = \bar{Y}_{..}$$

(2)將$\hat{\mu}$代入(a)

$$0 = \sum_j (Y_{ij} - \bar{Y}_{..} - \alpha_i) = \sum_i \sum_j Y_{ij} - \sum_i \sum_j \bar{Y}_{11} - \sum_i \sum_j \alpha_i$$
$$= \left[\sum_j Y_{i.} - n_i\bar{Y}_{..} - n_i\alpha_i\right]$$
$$= \left[n_1\bar{Y}_{i.} - n_i\bar{Y}_{..} - n_i\alpha_i\right]$$
$$= n_i(\bar{Y}_{i.} - \bar{Y}_{..} - \alpha_i)$$
$$\Rightarrow \hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..}$$

(二)1. $\begin{cases} H_0: \alpha_1 = \alpha_3 = \alpha_3 = 0 \\ H_1: \alpha_i\text{不為全}0 \end{cases}$

2.$\alpha = 0.05$

3.$C = \{F \mid F > F_{0.05}(2.7) = 4.7374\}$

4.(1)$SSE = 3(3.64) + 4(3.96) + 3(4.21) = 39.39$

(2)$SSTR = 3(16.4 - 13.64)^2 + 4(12.2 - 13.64)^2 + 3(12.8 - 13.64)^2$
$= 33.264$

(3) $F = \dfrac{MSTR}{MSE} = \dfrac{\frac{33.264}{2}}{\frac{39.39}{7}} = 2.9557 \notin C$

5.在$\alpha = 0.05$下，我們無法拒絕H_0

(三)1. $\begin{cases} \hat{\alpha}_1 = \bar{y}_{1.} - \bar{y}_{..} \\ \hat{\alpha}_2 = \bar{y}_{2.} - \bar{y}_{..} \end{cases} \Rightarrow \widehat{\alpha_1 - \alpha_2} = \bar{y}_{1.} - \bar{y}_{2.} \sim N\left(\alpha_1 - \alpha_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$

2. $$\begin{cases} \dfrac{(n_1-1)s_1^2}{\sigma^2} \sim x^2(n_1-1) \\ \dfrac{(n_2-1)s_2^2}{\sigma^2} \sim x^2(n_2-1) \\ \dfrac{(n_3-1)s_3^2}{\sigma^2} \sim x^2(n_3-1) \end{cases} \Rightarrow \frac{(n_1-1)s_1^2}{\sigma^2}+\frac{(n_2-1)s_2^2}{\sigma^2}+\frac{(n_3-1)s_3^2}{\sigma^2} \sim x^2(n_1+n_2+n_3-3)$$

3.**令** $$t=\frac{\dfrac{\overline{y}_{1.}-\overline{y}_{2.}-(\alpha_1-\alpha_2)}{\sqrt{\sigma^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}}{\sqrt{\dfrac{\dfrac{(n_1-1)s_1^2}{\sigma^2}+\dfrac{(n_2-1)s_2^2}{\sigma^2}+\dfrac{(n_3-1)s_3^2}{\sigma^2}}{n_1+n_2-3}}}$$

$$=\frac{\overline{y}_{1.}-\overline{y}_{2.}-(\alpha_1-\alpha_2)}{\sqrt{\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2+(n_3-1)s_3^2}{n_1+n_2+n_3-3}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} \sim t(n_1+n_2+n_3-3)$$

4.$\alpha_1-\alpha_2$**的**95%C.I**為**

$$(\overline{y}_{1.}-\overline{y}_{2.}) \pm t_{0.025}(7)\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2+(n_3-1)s_3^2}{n_1+n_2+n_3-3}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\Rightarrow (16.4-12.2) \pm 2.365\sqrt{\frac{(3-1)(3.64)+(4-1)(3.96+(3-1)(4.21)}{3+4+3-3}\left(\frac{1}{3}+\frac{1}{4}\right)}$$

$$\Rightarrow (0.6146,\ 7.7854)$$