統計

《統計學概要》

試題評析

(一)考題多為基本題型，考生欲拿高分不難。
(二)二項分配的機率求算，試題較為靈活，考生若不熟悉二項分配之定義及應用，可能無法正確地作答。
(三)有關迴歸直線與其它直線之優劣比較，坊間書籍較少提及，考生拿分較為不易。
(四)綜合來說，一般考生應可拿70分以上，程度較佳，考生拿90分以上應不成問題。

一、在某捐血中心之血庫中，有$\frac{1}{3}$是血型O^+，有$\frac{1}{15}$是血型O^-。今在此捐血中心之捐血人中隨機抽樣三位。設隨機變數X為三人中血型是O^+之人數，Y為三人中血型是O^-之人數。試求：

(一)隨機變數X及隨機變數Y之機率分配。（八分）
(二)隨機變數X＋Y之機率分配。（六分）
(三)隨機變數X＋Y之平均數及變異數。（四分）

答：

(一)1. $X \sim B\left(3,\frac{1}{3}\right)$為二項分配具參數3及$\frac{1}{3}$

$$\therefore f_X(x)=\binom{3}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{3-x} \quad x=0,1,2,3$$

2. $Y \sim B\left(3,\frac{1}{15}\right)$為二項分配具參數3及$\frac{1}{15}$

$$\therefore f_Y(y)=\binom{3}{y}\left(\frac{1}{15}\right)^y\left(\frac{14}{15}\right)^{3-y} \quad y=0,1,2,3$$

(二)X＋Y表示三人中血型為O之人數

$$\Rightarrow X+Y \sim B\left(3,\frac{6}{15}\right)$$

令$S=X+Y$

則S的pmf為 $fs(r)=\binom{3}{r}\left(\frac{6}{15}\right)^r\left(\frac{9}{15}\right)^{3-r}$， $r=0,1,2,3$

(三) $E(X+Y)=E(S)=3\cdot\frac{6}{15}=1.2$

$$Var(X+Y)=Var(S)=3\left(\frac{6}{15}\right)\left(\frac{9}{15}\right)=0.72$$

二、設在某大醫院看病之等待時間是具有落在區間40分鐘至3小時之均勻分配（Uniform Distribution）。

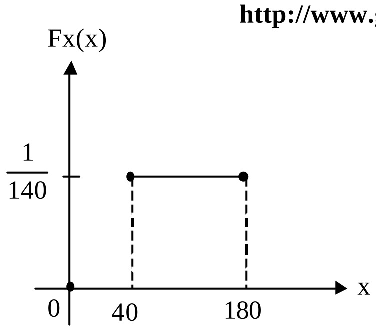
(一)試求此等待時間之機率密度函數，並繪出其圖形。（六分）
(二)試問任選一病人需等待50分鐘至2小時才能看病之機率。（四分）
(三)試問任選一病人需等待恰好1小時才能看病之機率。（四分）
(四)試求此等待時間之期望值。（四分）

答：

令X表示病人等候看病時間之隨機變數（單位：分鐘）

(一)$X \sim U(40, 80)$

$$\therefore fx(x)=\frac{1}{140}, \quad 40 \le x \le 180$$



(二)欲求機率為

$$P(50 \le X \le 120) = \int_{50}^{120} \frac{1}{140} dx = \frac{1}{140} \cdot x \Big|_{50}^{120} = \frac{1}{140} \cdot (120 - 50)$$
$$= \frac{1}{2}$$

(三)欲求機率為

$P(X = 60) = 0$（$\because$ X**為連續隨機變數**）

(四) $E(X) = \int_{40}^{180} x \cdot fx(x)dx$

$$= \int_{40}^{180} x \cdot \frac{1}{140} dx$$
$$= \frac{1}{140} \cdot \frac{1}{2} x^2 \Big|_{40}^{180}$$
$$= \frac{1}{280}(180^2 - 40^2)$$
$$= 110$$

$\therefore$**病人平均等候看病時間為**1**小時又**50**分鐘**

三、在某大城市隨機抽樣250戶家庭，顯示平均每戶擁有之電視機數是2.70，標準差為1.8。（參考附表）

(一)試以$\alpha = 0.05$檢定此城市平均每戶擁有之電視機數是否為2.5？（十分）

(二)試求(一)小題中之p–值。（五分）

(三)試此城市平均每戶擁有之電視機數為3，試求(一)小題中犯第二型錯誤之機率。（五分）

答：

設此城市平均每戶擁有之電視機數為μ

(一)1. $\begin{cases} H_0 : \mu = 2.5 \\ H_1 : \mu \ne 2.5 \end{cases}$

2. $\alpha = 0.05$

3. $C = \{ Z \,|\, |Z| > Z_{0.025} = 1.96\}$ **其中**$\alpha = P(Z > Z_\alpha)$

4. $Z = \dfrac{\overline{X} - 2.5}{\sqrt{\dfrac{S^2}{n}}} = \dfrac{2.7 - 2.5}{\sqrt{\dfrac{(1.8)^2}{250}}} \fallingdotseq 1.76 \notin C$

5.**在**$\alpha = 0.05$**下，我們無法拒絕**H_0**，亦即此城市平均每戶擁有之電視機數為**25**台**

(二)p-value = $P(Z > 1.76) = 0.0392$

(三) $C = \{ Z \,|\, |Z| > Z_{0.025} = 1.96\} = \left\{ \overline{X} \,\middle|\, \left| \dfrac{\overline{X} - 2.5}{\sqrt{\dfrac{(1.8)^2}{250}}} \right| > 1.96 \right\}$

$$= \{\overline{X} \,|\, \overline{X} < 2.2769 \text{或} \overline{X} > 2.7231\}$$

$$\beta = P(2.2769 < \overline{X} < 2.7231 \quad | \mu = 3)$$
$$= P\left(\frac{2.2769 - 3}{\sqrt{\frac{(1.8)^2}{250}}} < Z < \frac{2.723 - 3}{\sqrt{\frac{(1.8)^2}{250}}} \right)$$
$$= P(-0.16 < Z < -0.06)$$
$$= 0.0397$$

四、在一項有關某電視節目之收視情形調查，顯示200位男性觀眾中有60位不喜歡此節目，而300位女性觀眾中有75位不喜歡此節目。在$\alpha = 0.10$之下，試以兩種不同之統計方法推論男性觀眾與女性觀眾對此節目之喜好是否有所不同？並比較其結果。（參考附表）（二十分）

答：

設$P_{男}$為男性觀眾中不喜歡此節目之比例

設$P_{女}$為女性觀眾中不喜歡此節目之比例

1. $\begin{cases} H_0: P_{男} = P_{女} \\ H_1: P_{男} \neq P_{女} \end{cases}$

2. $\alpha = 0.10$

3. $C = \{ Z \mid |Z| > Z_{0.05} = 1.645 \}$ **其中**$\alpha = P(Z > Z_\alpha)$

4. $\hat{P} = \dfrac{60+75}{200+300} = 0.27$

$$Z = \frac{\hat{P}_{男} - \hat{P}_{女}}{\sqrt{\hat{P}(1-\hat{P})\left(\dfrac{1}{n_{男}} + \dfrac{1}{n_{女}}\right)}} = \frac{\dfrac{60}{200} - \dfrac{75}{300}}{\sqrt{(0.27)(0.73)\left(\dfrac{1}{200} + \dfrac{1}{300}\right)}}$$

$= 1.23 \notin C$

5. $\hat{P}_{男} = \dfrac{60}{200} = 0.3$

$\hat{P}_{女} = \dfrac{75}{300} = 0.25$

在顯著水準$\alpha = 0.10$**之下，我們無法拒絕虛無假設。**

亦即性別差異對此節目的喜好度並無顯著的影響。

五、下表為5位電腦推銷員之年資（年）及最近三個月內之銷售量（台）：

年資（x）	4	9	6	10	7
銷售量（y）	19	28	31	39	21

根據上述資料，請回答下列問題：

(一)設年資為自變數，銷售量為應變數。試用最小平方法求出迴歸直線之方程式：$\hat{y} = b_0 + b_1 x$**。（十分）**

(二)試解釋(一)小題中b1之意義。（三分）

(三)與其他直線比較，此迴歸直線有何優點？（五分）

(四)何種測度（measure）可以判定此迴歸直線是否適用？並請計算及說明之？（六分）

答：

(一)**令**$E = \sum (y_i - b_0 - b_1 x_i)^2$

令$\begin{cases} 0 = \dfrac{\partial E}{\partial b_0} = -2\sum (y_i - b_0 - b_1 x_i) \\ 0 = \dfrac{\partial E}{\partial b_1} = -2\sum x_i (y_i - b_0 - b_1 x_i) \end{cases}$

$\Rightarrow \begin{cases} \sum y_i = n b_0 + b_1 \sum x_i \\ \sum x_i y_i = b_0 \sum x_i + b_1 \sum x_i^2 \end{cases}$

$\Rightarrow \begin{cases} \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = \dfrac{138}{5} - (2.52)\left(\dfrac{36}{5}\right) = 9.456 \\ \hat{b}_1 = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = 2.52 \end{cases}$

$\therefore$**迴歸直線為** $\hat{y} = 9.456 + 2.52x$

(二)b_1**為迴歸直線之斜率，表示推銷員年資每增加一年，可添增之電腦銷售量（**2.52**台）**

(三)1.**因為此迴歸線由最小平方法估得，故估計值** $\hat{y}_i$ **與真實值**y_i**間誤差的平方和會較其它直線為短**

2.**會通過** $(\bar{x}, \bar{y})$

3.b_1**具解釋意義**

(四)可以R^2**（判定係數）判定此迴歸線是否適用**

$$SST = \sum(y_i - \overline{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 4068 - \frac{(138)^2}{5} = 259.2$$

$$SSR = b_1^2 \sum(x_i - \overline{x})^2 = b_1^2 \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] = (2.52)^2 \left[282 - \frac{(36)^2}{5} \right]$$

$$= 144.7891$$

$\therefore R^2 = \frac{SSR}{SST} = \frac{144.7891}{259.2} = 0.5586$ **此迴歸線的解釋能力達**55.86%