《統計學》

試題評析

前四題均屬基本題型，考生大多都可以拿到分數。惟獨第四題的第(三)小題，題意有誤；當抽樣分配未知，或抽樣分配已知但不易求機率值時，可以在樣本數n夠大之下，利用中央極限定理求機率近似值，而不是求樣本平均數之期望值和變異數。至於第五題爲迴歸模型設定，不易討論，因此，本題爲是否高分關鍵。

一、假設 $\mathrm{Var}(X)=16$ ， $\mathrm{Var}(Y)=9$ ，and $\mathrm{Cov}(X, Y)=-6$ 。

(一)X和Y兩變數之相關係數。(5分)

(二) $4X+1$ 和 $2Y-4$ 之相關係數。(5分)

答：

(一) $\rho_{XY}=\dfrac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y}=\dfrac{-6}{\sqrt{16}\sqrt{9}}=-\dfrac{1}{2}$

(二) $\mathrm{cov}(4X+1\,,\,2Y-4)=8\,\mathrm{cov}(X,Y)=8\times(-6)=-48$

$V(4X+1)=4^2\cdot V(X)=16\times 16=256$

$V(2Y-4)=2^2\cdot V(Y)=4\times 9=36$

$$\rho=\frac{\mathrm{cov}(4X+1\,,\,2Y-4)}{\sqrt{V(4X+1)}\sqrt{V(2Y-4)}}=\frac{-48}{\sqrt{256}\sqrt{36}}=-\frac{1}{2}$$

二、設 X_1 和 X_2 為兩隨機變數，其p. d. f. 為

$$f(X_1, X_2)=\begin{cases}\dfrac{X_1X_2}{k}, & X_1=1,2,3;\ X_2=1,2;\ k>0\\ 0, & \text{其他}\end{cases}$$

(一)分別求出 X_1 和 X_2 之機率函數。(10分)

(二)計算 X_1 和 X_2 乘積等於3之機率。(5分)

(三)請寫出兩隨機變數之聯合機率表，並計算出 k 值。(10分)

答：

$$\sum_{X_1=1}^{3}\sum_{X_2=1}^{2}\frac{X_1X_2}{k}=\frac{1}{k}(1\times1+1\times2+2\times1+2\times2+3\times1+3\times2)=\frac{18}{k}=1$$

$k=18$

(一) (1) $f(X_1)=\displaystyle\sum_{X_2=1}^{2}\frac{X_1X_2}{18}=\frac{X_1}{6}$ ， $X_1=1,2,3$

(2) $f(X_2)=\displaystyle\sum_{X_1=1}^{3}\frac{X_1X_2}{18}=\frac{X_2}{3}$ ， $X_2=1,2$

(二) $P(X_1X_2=3)=P(X_1=3\,,X_2=1)=\dfrac{3\times1}{18}=\dfrac{1}{6}$

(三) $\because\ \displaystyle\sum_{X_1=1}^{3}\sum_{X_2=1}^{2}f(X_1X_2)=\sum_{X_1=1}^{3}\sum_{X_2=1}^{2}\frac{X_1X_2}{k}=1$

$\therefore k=18$

X_1 \ X_2	$X_2=1$	$X_2=2$
$X_1=1$	$\frac{1}{18}$	$\frac{1}{9}$
$X_1=2$	$\frac{1}{9}$	$\frac{2}{9}$
$X_1=3$	$\frac{1}{6}$	$\frac{1}{3}$

三、設X為一隨機變數，其期望值 $E(X)=2$，而 $E(X^2)=9$，應用Chebyshev inequality（不等式）決定 $p(-2<X<6)$ 機率之下限。（10分）

答：

$$Var(X)=EX^2-(EX)^2=9-2^2=5$$

$$P(-2<X<6)=P(-4<X-EX<4)$$

$$=P(|X-EX|<4)\geq 1-\frac{Var(X)}{4^2}=1-\frac{5}{16}=\frac{11}{16}$$

四、若XYZ公司之每日股價變動（X）為一隨機變數其隨機分配如下：

$X=x$	1	2	3
$P(X=x)$	0.4	0.4	0.2

（一）試求隨機取2個樣本之平均數的抽樣分配（Sampling Distribution）（只考慮放還抽樣）。（10分）

（二）根據（一）式，計算樣本平均數之期望值（Expected Value）和標準差（Standard Deviation）。（10分）

（三）利用中央極限定理計算出樣本平均數之期望值和變異數（Variance）。（10分）

答：

(x_1,x_2)	$f(x_1,x_2)$	$\bar{x}=\frac{x_1+x_2}{2}$
(1,1)	$0.4\times0.4=0.16$	1
(1,2)	$0.4\times0.4=0.16$	1.5
(1,3)	$0.4\times0.2=0.08$	2
(2,1)	$0.4\times0.4=0.16$	1.5
(2,2)	$0.4\times0.4=0.16$	2
(2,3)	$0.4\times0.2=0.08$	2.5
(3,1)	$0.2\times0.4=0.08$	2
(3,2)	$0.2\times0.4=0.08$	2.5
(3,3)	$0.2\times0.2=0.04$	3

$\bar{x}$	$f(\bar{x})$
1	0.16
1.5	$0.16\times2=0.32$
2	$0.08\times2+0.16=0.32$
2.5	$0.08\times2=0.16$
3	0.04

(一)抽樣分配

$$f(\bar{x})=\begin{cases}0.16 & ,\ \bar{x}=1,2.5\\ 0.32 & ,\ \bar{x}=1.5,2\\ 0.04 & ,\ \bar{x}=3\\ 0 & ,\ \text{o.w.}\end{cases}$$

(二) (1)期望值：

$$E\overline{X}=\sum_{\bar{x}}\bar{x}\cdot f(\bar{x})$$
$$=(1\times0.16)+(2.5\times0.16)+(1.5\times0.32)+(2\times0.32)+(3\times0.04)=1.8$$

(2) $E\overline{X}^{2}=\bar{x}^{2}\cdot f(\bar{x})$

$$=(1^2\times0.16)+(2.5^2\times0.16)+(1.5^2\times0.32)+(2^2\times0.32)+(3^2\times0.04)=3.52$$

$$V(\overline{X})=E\overline{X}^{2}-(E\overline{X})^2=3.52-1.8^2=0.28$$

$\therefore$ 標準差 $\sigma_{\bar{x}}=\sqrt{0.28}=0.5291$

(三) $EX=\sum_{x=1}^{3}x\cdot f(x)=(1\times0.4)+(2\times0.4)+(3\times0.2)=1.8$

$$EX^{2}=\sum_{x=1}^{3}x^2\cdot f(x)=(1^2\times0.4)+(2^2\times0.4)+(3^2\times0.2)=3.8$$

$$V(x)=EX^{2}-(EX)^2=3.8-1.8^2=0.56$$

$\therefore$ $E\overline{X}=EX=1.8$

$$Var(\overline{X})=\frac{\sigma^2}{n}=\frac{0.56}{n}$$

根據中央極限定理

$$\overline{X}\xrightarrow[n\to\infty]{d}N\left(1.8,\frac{0.56}{n}\right)$$

五、假設以下為臺灣對茶葉需求之模型：

模型一St. dev.	$\hat{Q}=3.8 + 0.15\log P_c + 0.76\log Y + 0.06\log P_t$ (1.9) (0.15) (0.25) (0.42)
模型二St. dev.	$\hat{Q}=3.2 + 0.67\log Y + 0.03\log P_t$ (2.1) (0.26) (0.41)
模型三St. dev.	$\hat{Q}=3.7 + 0.15\log P_c + 0.73\log Y$ (0.75) (0.15) (0.15)
模型四St. dev.	$\hat{Q}=2.8 + 0.19\log P_c + 0.26\log Y - 1.48\log P_t + 1.18\log P_I$ (1.9) (0.11) (0.32) (0.9) (0.56)

其中Q＝臺灣對茶葉需求量（取對數）

Y＝可支配所得　P_c＝進口咖啡價格　P_t＝臺灣茶葉價格

P_I＝進口茶葉價格

請根據統計和經濟理由，逐一說明每一模型之適切性，並選擇一最佳模型來解釋臺灣對茶葉需求。（25分）

答：

(一)適切性：

1.模型一：適切性不佳

(1)統計理由：

$\log\ P_t$ 檢定 t 值 $=\dfrac{0.06}{0.42}=0.14$ 很不顯著

$\log\ P_C$ 檢定 t 值 $=\dfrac{0.15}{0.15}=1$ 亦不顯著

因此模型含有2個不重要自變數可能造成

①自變數之間的共線性，使得Y（可支配所得）的效性降低，估計需求量 $\hat{Q}$ 的偏差變大。

②調整判定係數 R_a^2 將會變小，使得模型解釋能力降低。

(2)經濟理由：

當茶葉價格 P_t 愈高，則茶葉需求量 Q 愈低，

因此 $\log P_t$ 之迴歸係數估計值不應爲“正”。

2.模型二：適切性不佳

(1)統計理由：

$\log\ P_t$ 檢定 t 值 $=\dfrac{0.03}{0.41}=0.07 \fallingdotseq 0$

顯然在此模型中是爲極不重要的自變數，

因此將與模型一產生相同的問題。

(2)經濟理由：

與模型一相同。

3.模型三：適切性較佳

(1)統計理由：

雖然 $\log\ P_C$ 檢定 t 值 $=\dfrac{0.15}{0.15}=1$ 不顯著，

但不至於過小，而 $\log$ Y之檢定 t 值 $=\dfrac{0.73}{0.15}=4.87$ 顯著，表示 P_C 對Y的效性與對 R_a^2（模型解釋能力）影響不大。

(2)經濟理由：

可支配所得Y愈大，則茶葉需求量將會愈大，因此 $\log$ Y係數 $0.73>0$ 是爲合理；再者，在經濟學上，對某小部份族群而言，咖啡與茶葉有其替代作用，當咖啡價格愈高時，將有愈多的人轉向茶葉需求。

4.模型四：適切性不佳

(1)統計理由：

可支配所得Y對茶葉需求量Q而言，是重要關鍵的自變數，但是檢定 t 值 $=\dfrac{0.26}{0.32}=0.81$ 並不顯著，顯然模型存在嚴重的共線性。

(2)經濟理由：

進口茶葉價格 P_I 愈高時，則茶葉需求量Q將會愈低，因此 P_I 係數不應爲“正”。

(二)綜合上述理由，以模型三來解釋台灣對茶葉需求最佳。