《統計學》

試題評析	這次的地特三等經建行政統計學第一和第四大題為基本題，這二大題要先做，拿下基本分。第二大題有點數理統計的範圍，不過對於考經建行政的同學只要觀念清楚花點時間拿到分數是沒問題的。第三題考的是實驗設計的巢形設計，考在統計行政類算合理，但考在經建行政有點超出範圍，不過這題26分不用太在意。這份考卷最高分應該是80分，第三大題把它當一因子來做，應該也有給一點分數。
考點命中	第一題：《高點・高上統計學講義》第三回，4-1常用分配，蘇建郎編撰，頁11、13。 第二題：《高點・高上統計學講義》第六回，6-1點估計充分性，蘇建郎編撰，頁26。 《高點・高上統計學講義》第二回，3-6變數變換，蘇建郎編撰，頁41。 《高點・高上統計學講義》第二回，3-5條件分配，蘇建郎編撰，頁34。 第四題：《高點・高上迴歸分析講義》第一回，9-2簡單線性迴歸與相關分析，蘇建郎編撰，頁16。

一、已知兩個獨立且相同分配的隨機變數分別為 X 和 Y，而且 $X+Y$ 的動差母函數（moment generating function）為 $m_{X+Y}(t)=e^{4(e^t-1)}$。

請回答下列問題：

(一)求隨機變數 X 的動差母函數及機率分配函數。（12分）

(二)計算機率 $P(X\geq 2)$。（5分）

(三)利用題(一)的動差母函數，求母體變異數（variance）。（8分）

答：

(一) $m_{X+Y}(t)=e^{4(e^t-1)} \Rightarrow X+Y\sim poi(4)$，因為 X,Y 相同分配，故 $X,Y\sim poi(2)$

$$m_X(t)=e^{2(e^t-1)},t\in \mathrm{R}\ ,\ f_X(x)=e^{-2}\frac{2^x}{x!}\quad ,x=0,1,.....$$

(二) $P(X\geq 2)=1-f(0)-f(1)=1-\dfrac{e^{-2}2^0}{0!}-\dfrac{e^{-2}2^1}{1!}=1-3e^{-2}=0.594$

(三) $m_X(t)=e^{2(e^t-1)},t\in \mathrm{R}$

$$m_X(t)=e^{2(e^t-1)}(2e^t)\qquad \Rightarrow E(X)=m'_X(t=0)=e^{2(e^0-1)}(2e^0)=2e^0=2$$

$$m''_X(t)=e^{2(e^t-1)}(2e^t)^2+e^{2(e^t-1)}(2e^t)\ \Rightarrow E(X^2)=m''_X(t=0)=e^{\lambda(e^0-1)}(\lambda e^0)^2+e^{\lambda(e^0-1)}(\lambda e^0)=\lambda^2+\lambda$$

$$Var(X)=E(X^2)-[E(X)]^2=2^2+2-2^2=2$$

二、設 X_1 和 X_2 是兩個獨立且有相同分配的隨機變數，其機率密度函數如下所示：

$$f(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}}，0<x<\infty，0<\theta<\infty。$$

請回答下列問題：（每小題6分，共24分）

(一)證明 X_1+X_2 是 θ 的充分統計量（sufficient statistic）。

(二)求 $Y_1=X_1+X_2$ 和 $Y_2=X_2$ 的聯合機率密度函數。

(三) Y_1 和 Y_2 如題(二)所定義，計算已知 Y_1 下，Y_2 的期望值，即 $E(Y_2\mid Y_1)$。

(四) Y_1 和 Y_2 如題(二)所定義，求 $E(Y_2\mid Y_1)$ 的變異數，即 $Var\left(E(Y_2\mid Y_1)\right)$。

答：

(一)統計量之分配(抽樣分配) $T = X_1 + X_2 \sim Gamma(2,\theta)$，即 $f_T(t) = \frac{t}{\theta^2} e^{-\frac{t}{\theta}}$,$t > 0$

聯合分配 $f_{X_1,X_2}(x_1,x_2;\theta) = f_{X_1}(x_1;\theta) f_{X_2}(x_2;\theta) = \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \frac{1}{\theta} e^{-\frac{x_2}{\theta}} = \frac{1}{\theta^2} e^{-\frac{x_1+x_2}{\theta}}$

計算 $f_{\underset{\sim}{X}|T}(\underset{\sim}{x}\,|\,t) = \frac{f_{\underset{\sim}{X},T}(\underset{\sim}{x},t)}{f_T(t)} = \frac{\frac{1}{\theta^2} e^{-\frac{x_1+x_2}{\theta}}}{\frac{t}{\theta^2} e^{-\frac{t}{\theta}}} = \frac{1}{t}$ 與參數 θ 無關。

其中 $t = x_1 + x_2$，故 $T = X_1 + X_2$ 為參數 θ 之充分統計量。

(二) $f_{X_1,X_2}(x_1,x_2) = \frac{1}{\theta^2} e^{-\frac{x_1+x_2}{\theta}}$,$x_1 > 0, x_2 > 0$

$$\begin{cases} Y_1 = X_1 + X_2 \\ Y_2 = X_2 \end{cases} \Rightarrow \begin{cases} x_1 = y_1 - y_2 \\ x_2 = y_2 \end{cases} \Rightarrow J = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{\theta^2} e^{-\frac{y_1}{\theta}}$，$y_1 > 0, y_2 > 0$

(三) $E(Y_2 | Y_1) = E(X_2 | X_1 + X_2)$

又 $E(X_1 | X_1 + X_2) = E(X_2 | X_1 + X_2)$ （X_1, X_2 相同分配）

且 $E(X_1 | X_1 + X_2) + E(X_2 | X_1 + X_2) = E(X_1 + X_2 | X_1 + X_2) = X_1 + X_2$

$\Rightarrow$ $E(X_2 | X_1 + X_2) = \frac{X_1 + X_2}{2}$，故 $E(Y_2 | Y_1) = \frac{X_1 + X_2}{2}$

(四) $Var[E(Y_2 | Y_1)] = Var(\frac{X_1 + X_2}{2}) = \frac{1}{4}[Var(X_1) + Var(X_2)] = \frac{\theta^2}{2}$

三、研究洋蔥在不同溫度下軟化的程度，隨機抽取6袋洋蔥，每袋有5顆洋蔥。3袋洋蔥存放於華氏60度的儲存室，另外3袋洋蔥存放於華氏80度的儲存室。應變數（dependent variable）Y 為軟化的程度。利用統計軟體（SAS）PROC MEANS得到下列結果：

溫度	袋	總和	平均數	修正的平方和（corrected sum squares）
60	1	50	10	52.58
60	2	55	11	42.82
60	3	75	15	19.30
80	4	70	14	22.78
80	5	90	18	51.12
80	6	80	16	57.40
		420	84	246.00

請回答下列問題：

(一)寫出計算適合此實驗設計的變異數分析表（ANOVA Table），包含來源（source）、自由度（degree of freedom）及平方和（sum of squares）。（18分）

(二)以 F 檢定（F test）檢定溫度的效果，列出虛無假設與對立假設，檢定統計量、棄卻域（rejection region）和結論。（8分）

（參考的 f 值：$f_{5,24,0.05} = 2.621$, $f_{1,24,0.05} = 4.260$, $f_{1,4,0.05} = 7.709$, $f_{1,23,0.05} = 4.279$）

答：

(一)二因子巢形設計，溫度A，60度A1，80度A2，袋B

$SSTO = 7380 - \frac{420^2}{30} = 1500$

$SSA = \frac{180^2 + 240^2}{15} - \frac{420^2}{30} = 6000 - 5880 = 120$

$SSB(A1) = \frac{50^2 + 55^2 + 75^2}{5} - \frac{180^2}{15} = 2230 - 2160 = 70$

$SSB(A2) = \frac{70^2 + 90^2 + 80^2}{5} - \frac{240^2}{15} = 3880 - 3840 = 40$

$SSB = SSB(A1) + SSB(A2) = 70 + 40 = 110$

$SSE = SSTO - SSA + SSB = 1500 - 120 - 110 = 1270$

變異來源	自由度df	平方和SS
溫度	1	120
袋(溫度) 袋(60) 袋(80)	4 2 2	110 70 40
誤差	24	1270
總合	29	1500

(二)

(1) H_0 : 不同溫度對軟化程度無差異

H_1 : 不同溫度對軟化程度有差異

(2) $F^* | = \frac{MSA}{MSE} = \frac{120/1}{1270/24} = 2.2677$

(3) $\alpha = 0.05$ ， $C = \{F^* \mid F^* >_{0.05} (1,24) = 4.26\}$

$F^* | = 2.2677 \notin C$ ， $don't\ reject\ H_0$

(4)故在 $\alpha = 0.05$ 情況下，我們沒有顯著證據說不同溫度對軟化程度有差異。

四、某便利商店為了瞭解廣告的促銷效果，在臺北市隨機抽取15家分店比較廣告前的銷售量 x_i 與廣告後的銷售量 y_i ， $i = 1,2,\ldots,15$ 。資料整理如下所示：

	廣告前的銷售量	廣告後的銷售量
樣本平均數	$\bar{x} = 800$	$\bar{y} = 1200$
樣本變異數	$S_x^2 = 1400$	$S_y^2 = 2000$

另知廣告前後銷售量的樣本相關係數為 $r = 0.85$。擬以簡單線性迴歸模型 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ， $\varepsilon_i \overset{i.i.d.}{\sim} N(0,\sigma^2)$ ， $i = 1,2,\ldots,15$,分析資料，請回答下列問題：

(一)求斜率（ β_1 ）的最小平方估計值。（10分）

(二)檢定廣告前後的平均銷售量是否相等？（假設廣告前後的銷售量的母體均滿足變異數均等的常態分配，顯著水準 $\alpha = 0.10$。）（5分）

(三)已知廣告前的銷售量是900，求其廣告後的平均銷售量是多少？（10分）

（參考的 t 值：$t_{30,0.05}=1.697,\ t_{29,0.05}=1.699,\ t_{15,0.05}=1.753,\ t_{14,0.05}=1.761$）

答：

(一) $r_{X,Y}=\hat{\beta}_1\dfrac{S_X}{S_Y} \Leftrightarrow \hat{\beta}_1=r_{X,Y}\dfrac{S_Y}{S_X}=0.85\sqrt{\dfrac{2000}{1400}}=1.0159$

(二) μ_1 廣告前的銷售量；μ_2 廣告後的銷售量

(1) $\begin{cases}H_0:\mu_1=\mu_2\\ H_1:\mu_1\neq\mu_2\end{cases}$，$\alpha=0.1$，

$\overline{d}=\overline{x}-\overline{y}=800-1200=-400$

$s_d^2=s_x^2+s_y^2-2rs_xs_y=1400+2000-2\times0.85\times\sqrt{1400}\times\sqrt{2000}=555.356$

$s_d=23.566$

(2) $t^*=\dfrac{\overline{d}}{s_d/\sqrt{n}}=\dfrac{-400}{23.566/\sqrt{15}}=\dfrac{-400}{18.9841}=-65.73$

(3) $\alpha=0.1$，$C=\{t^*:|t^*|>t_{\alpha/2}(n-1)=t_{0.05}(14)=1.761\}$

$t^*=-65.73\in C$， $reject\ H_0$

(4)故在顯著水準 $\alpha=0.1$ 的情況下，根據樣本資料顯示，我們有證據說 $\mu_1\neq\mu_2$ 則表示廣告前後的銷售量有差異。

(三)由(一)得 $\hat{\beta}_1=1.0159$，且

樣本迴歸線 $\hat{y}=387.28+1.0159x$

$\hat{E}(Y\mid X=900)=387.28+1.0159\times900=1301.59$

故廣告後平均銷售量為1301.59