《統計學概要(A)-統計》

試題評析	108普通考試統計組別的統計學概要有四大題，難度偏簡單，其中就有2.5題是應用統計學的五大重點，另外的1.5題是聯合分配的計算，一定要把握，但應用統計學中有兩題是卡方檢定，計算量比較大，請同學在平常的演練就要練好算計的速度。
考點命中	第一題：《高點・高上統計學講義》第三回，蘇建郎編撰，頁77，ch7-2檢定統計量法決策準則。 第二題：《高點・高上統計學講義》第四回，蘇建郎編撰，頁35~36，ch10-2齊一性檢定。 《高點・高上統計學講義》第三回，蘇建郎編撰，頁83~84，ch7-2檢定統計量法決策準則。 第三題：《高點・高上統計學講義》第二回，蘇建郎編撰，頁36，ch3-4條件分配。 第四題：《高點・高上統計學講義》第二回，蘇建郎編撰，頁24，ch3-3多維隨機變數。

一、某藥廠想知道肝癌病人服用新藥半年後是否可以有效降低肝癌指數（GOT），自肝癌病人群中隨機抽取 8 人，記錄其服用新藥前和新藥服用半年後的 GOT。肝癌病人 GOT 如表所示。

肝癌病人	1	2	3	4	5	6	7	8
新藥服用前的 GOT	45	60	120	200	150	100	80	50
新藥服用半年後的 GOT	30	30	80	150	120	100	20	20

(一) 請問這兩群（新藥服用前的 GOT 和新藥服用半年後的 GOT）是相依還是獨立？請說明原因。（5 分）

(二) 試檢定新藥服用半年後是否可以有效降低 GOT，顯著水準 0.05。（10 分）

(三) 試檢定新藥服用半年後之 GOT 是否比服用新藥前之 GOT 平均至少可以減少 30，顯著水準為 0.05。（10 分）

(四) 上述檢定結果若可信賴，則數據之假設條件為何？（5 分）

答：

(一)相依；8個相同的病人用藥前，用藥半年後是相同的實驗個體。

(二) μ_1 病人用藥前平均GOT；μ_2 病人用藥半年後平均GOT

令 $\mu_d = \mu_1 - \mu_2$，若 $\mu_1 > \mu_2$ 則表示藥有效，即 $\mu_d = \mu_1 - \mu_2 > 0$

① $\begin{cases} H_0 : \mu_1 - \mu_2 \le 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{cases} \Rightarrow \begin{cases} H_0 : \mu_d \le 0 \\ H_1 : \mu_d > 0 \end{cases}$，$\alpha = 0.05$

病人	1	2	3	4	5	6	7	8
前-後	15	30	40	50	30	0	60	30

$n = 8$，$\bar{d} = 31.875$，$s_d = 18.8864$

② $t^* = \dfrac{\bar{d}}{s_d/\sqrt{n}} = \dfrac{31.875}{18.8864/\sqrt{8}} = 4.77$

③ $\alpha = 0.05$，$C = \{t^* \mid t^* > t_\alpha(n-1) = t_{0.05}(7) = 1.895\}$

$t^* = 4.77 \in C$, $reject\ H_0$，

④故在顯著水準 $\alpha = 0.05$ 的情況下，根據樣本資料顯示，我們有證據說新藥服用半年後是有顯著降低GOT。

(三)

① $\begin{cases} H_0: \mu_d \le 30 \\ H_1: \mu_d > 30 \end{cases}$ ， $\alpha = 0.05$

病人	1	2	3	4	5	6	7	8
前-後	15	30	40	50	30	0	60	30

$n = 8$，$\bar{d} = 31.875$，$s_d = 18.8864$

② $t^* = \dfrac{\bar{d} - 30}{s_d / \sqrt{n}} = \dfrac{1.875}{18.8864/\sqrt{8}} = 0.281$

③ $\alpha = 0.05$，$C = \{t^* \mid t^* > t_\alpha(n-1) = t_{0.05}(7) = 1.895\}$

$t^* = 0.281 \notin C$， don't reject H_0，

④故在顯著水準$\alpha = 0.05$的情況下，根據樣本資料顯示，我們沒有證據說新藥服用半年後可以降低30的GOT。

(四)病人用藥前GOT與病人用藥後半年GOT都是常態分配。

二、令 A、B、C、D 及 E 表示五種品牌的手機，自各品牌的使用者分別隨機抽取 500 人調查其下次會再購買相同品牌的意願。資料如表所示。

品　牌	A	B	C	D	E
會再購買相同品牌人數	450	300	250	100	50
抽取人數	500	500	500	500	500

(一) 請估計各品牌的使用者下次會再購買相同品牌手機的機率。（5 分）

(二) 請檢定各品牌的使用者下次會再購買相同品牌手機的機率是否都相同。顯著水準$\alpha = 0.05$。（10 分）

(三) 假設只存在 A 和 E 兩品牌下，請以和子題(二)不同的檢定方法檢定 A 和 E 品牌手機的使用者下次會再購買相同品牌手機的機率是否相同，並說明何以可用此方法檢定。顯著水準$\alpha = 0.05$。（10 分）

答：

(一) $\hat{p}_A = 0.9$，$\hat{p}_B = 0.6$，$\hat{p}_C = 0.5$，$\hat{p}_D = 0.2$，$\hat{p}_E = 0.1$

(二)① 設p_A, p_B, p_C, p_D, p_E分別表示五種品牌再購買比率。

$H_0: p_A = p_B = p_C = p_D = p_E$ 。

H_1: 五種品牌再購買比率不全相同，$i = 1, 2, 3$， $a = 0.05$

②

品牌(o_{ij} / e_{ij})	A	B	C	D	E	合計
會再買人數	450/230	300/230	250/230	100/230	50/230	1150
不會再買人數	50/270	200/270	250/270	400/270	450/270	1350
抽取人數	500	500	500	500	500	2500

$$c^2 = \sum_{i=1}^{2} \sum_{j=1}^{5} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = \frac{(450-230)^2}{230} + \cdots + \frac{(450-270)^2}{270} = 829.31$$

③自由度$df = (r-1((c-1) = (2-1)\times(5-1) = 1\times 4 = 4$

$a = 0.05$，拒絕域$C = \left\{c^2 \mid c^2 > c^2_{0.05}(4) = 9.488\right\}$，$c^2 \in C$， $reject\ H_0$

④故在顯著水準$a = 0.05$的情況下，根據樣本資料顯示，我們有證據說H_1： 五種品牌再購買比率不全相同。

(三) p_A, p_E分別表示A,E品牌再購買比率。

① $\begin{cases} H_0 : p_A = p_E \\ H_1 : p_A \neq p_E \end{cases}$

$\alpha = 0.05$，$\hat{p}_A = 0.9$，$\hat{p}_E = 0.1$，$\hat{p} = \dfrac{450+50}{500+500} = 0.5$

② $z^* = \dfrac{0.9-0.1}{\sqrt{0.5\times 0.5(\frac{1}{500}+\frac{1}{500})}} = 25.3$

③ $\alpha = 0.05$，$C = \{z^* : |z^*| > z_{\alpha/2} = z_{0.025} = 1.96\}$

$z^* = 25.3 \in C$, $reject\ H_0$

④故在顯著水準$\alpha = 0.05$的情況下，根據樣本資料顯示，我們有證據說A,E品牌再購買比率有差異。

三、自 W 城市的土地所有者中隨機抽取 1,200 人，調查其土地位置（X）及土地售價（Y）。收集之資料整理如表所示：

X \ Y	1,000 萬元	1,500 萬元	2,000 萬元	人數
位置 1	60	500	40	600
位置 2	40	300	60	400
位置 3	50	100	50	200
人數	150	900	150	1,200

(一) 將表中的人數計算為機率值。（4 分）

(二) 計算條件期望值 $E(Y \mid X \le 2)$。（8 分）

(三) 試檢定土地位置（X）及土地售價（Y）是否相關。顯著水準 0.05。（8 分）

答：

(一)

X \ Y	1000萬	1500萬	2000萬	人數
位置1	$1/20$	$5/12$	$1/30$	$1/2$
位置2	$1/30$	$1/4$	$1/20$	$1/3$
位置3	$1/24$	$1/12$	$1/24$	$1/6$
人數	$1/8$	$3/4$	$1/8$	1

(二)

$$f_{Y|X}(y\,|\,1)=\begin{cases}1/10, & y=1000\\ 5/6, & y=1500\\ 1/15, & y=2000\\ 0, & o.w.\end{cases},\quad f_{Y|X}(y\,|\,2)=\begin{cases}1/10, & y=1000\\ 3/4, & y=1500\\ 3/20, & y=2000\\ 0, & o.w.\end{cases}$$

$$E(Y\,|\,X=1)=1000\times\frac{1}{10}+1500\times\frac{5}{6}+2000\times\frac{1}{15}=\frac{4450}{3}=1483.3333$$

$$E(Y\,|\,X=2)=1000\times\frac{1}{10}+1500\times\frac{3}{4}+2000\times\frac{3}{20}=1525$$

(三)

①H_0：土地位置與土地售價是無關

H_1：土地位置與土地售價是有關　，$a=0.05$

②

o_{ij} \ e_{ij}	1000萬	1500萬	2000萬	人數
位置1	60 75	500 450	40 75	600
位置2	40 50	300 300	60 50	400
位置3	50 25	100 150	50 25	200
人數	150	900	150	1200

$$c^2=\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}=\frac{(60-75)^2}{75}+\cdots+\frac{(50-25)^2}{25}=95.55$$

③自由度$df=(r-1((c-1)=(3-1)\times(3-1)=2\times2=4$

$a=0.05$，拒絕域$C=\left\{c^2\,|\,c^2>c^2_{0.05}(4)=9.488\right\}$，$c^2\in C$，　$reject\ H_0$

④故在顯著水準$a=0.05$的情況下，根據樣本資料顯示，我們有證據說土地位置與土地售價是有關。

四、假設(X, Y)的聯合機率密度函數（joint probability density function, pdf）為

$$f(x,y)=c(3x+2y),\qquad 0\le x,y\le 1.$$

(一)計算 c 值。（5 分）

(二)計算機率 $P(0\le X\le 0.5, 0\le Y\le 0.5)$。（5 分）

(三)推導求得 X 的邊際機率密度函數（Marginal pdf of X）。（5 分）

(四)計算 X 的期望值和變異數。（10 分）

答：

(一)

$$1=\int_0^1\int_0^1 c(3x+2y)dxdy=c\frac{5}{2} \Rightarrow c=\frac{2}{5}$$

且 $f_{X,Y}(x,y)=\frac{2}{5}(3x+2y)\ge 0$，$\forall 0\le x,y\le 1$

(二) $P(0\le X\le 0.5, 0\le Y\le 0.5)=\int_0^{0.5}\int_0^{0.5}\frac{2}{5}(3x+2y)dxdy=\frac{1}{8}=0.125$

(三) $f_X(x)=\int_0^1\frac{2}{5}(3x+2y)dy=\frac{2}{5}(3x+1)$，$0\le x\le 1$

(四)

$E(X)=\int_0^1 x\frac{2}{5}(3x+1)dx=\frac{3}{5}=0.6$，$E(X^2)=\int_0^1 x^2\frac{2}{5}(3x+1)dx=\frac{13}{30}$

$Var(X)=EX^2-(EX)^2=\frac{13}{30}-(\frac{3}{5})^2=\frac{11}{150}$