《統計學概要(統計)》

試題評析	今年統計行政統計學偏簡單，前三題是基本題型，大部分的題型在歷屆考古題都有出現，第4題Spearman相關計算在去年105年普考經建行政的統計學有考過，這份考卷要上榜的同學至少要考到90分以上。
考點命中	一、《高點‧高上統計學講義》第一回，楊老師編撰，2-4貝式定理，頁22-23。 二、《高點‧高上統計學講義》第五回，楊老師編撰，7-2檢定統計量法決策準則，頁34；6-2區間估計，頁18-19；7-2檢定統計量法決策準則，頁38-39。 三、《高點‧高上統計學講義》第七回，楊老師編撰，8-4二因子變異數分析，頁13-15。 四、《高點‧高上迴歸分析講義》第一回，楊老師編撰，9-2相關分析，頁18-19；歷屆考題計算題，頁67。

一、A、B和C公司生產的5.5吋手機市佔率各為35%、20%和45%。A公司所銷售的5.5吋手機消費者不滿意率為10%，B公司所銷售的5.5吋手機消費者不滿意率為15%，C公司所銷售的5.5吋手機消費者不滿意率為5%。

(一)消費者對買到的5.5吋手機滿意之機率為何？（5分）

(二)C公司售出的500件5.5吋手機中，消費者滿意的5.5吋手機個數為何？（5分）

(三)消費者買到不滿意的5.5吋手機分別是由A、B或C公司生產之機率為何？（10分）

答：

(一) $A\ B\ C$ 為生產 5.5 吋手機市場佔有之事件

E 為滿意之事件 ， E^C 為滿意之事件

$$P(E)=P(E\cap A)+P(E\cap B)+P(E\cap C)$$
$$=P(A)P(E\mid A)+P(B)P(E\mid B)++P(\mathrm{C})P(E\mid C)$$
$$=0.35\times 0.9+0.2\times 0.85+0.45\times 0.95=0.9125$$

(二) $500\times 0.9125=456.25$ 個

(三)不滿意之機率為 $P(E^C)=1-P(E)=1-0.9125=0.0875$

不滿意來自 A 公司機率

$$P(A\mid E^C)=\frac{P(A\cap E^C)}{P(E^C)}=\frac{P(A)P(E^C\mid A)}{P(E^C)}=\frac{0.35\times 0.1}{0.0875}=0.4$$

不滿意來自 B 公司機率

$$P(B\mid E^C)=\frac{P(B\cap E^C)}{P(E^C)}=\frac{P(B)P(E^C\mid B)}{P(E^C)}=\frac{0.2\times 0.15}{0.0875}=\frac{12}{35}$$

不滿意來自 C 公司機率

$$P(C\mid E^C)=1-0.4-\frac{12}{35}=\frac{9}{35}$$

二、工廠有兩條節能燈泡的生產線，老闆想知道這兩條生產線生產的節能燈泡之平均壽命是否有差異，於是自這兩條生產線分別隨機抽取5和6個節能燈泡並量測其壽命。兩條生產線抽取的節能燈泡壽命（單位：千小時）如表所示。假設兩條生產線的節能燈泡壽命呈常態分配。

	壽命	
燈泡	生產線1	生產線2
1	38	63
2	44	53
3	47	56
4	44	54

5	52	64
6	-	62

(一)檢定生產線2所生產的燈泡壽命之母體標準差是否超過10(千小時)。顯著水準為0.05。(6分)

(二)檢定這兩條生產線生產的燈泡壽命之母體平均值是否相等。顯著水準皆為0.1。(14分)

答:

(一)生產線 2 綜合統計量訊息

$n_2=6$,$\bar{x}_2=58.6667$,$s_2=4.8854$

① $\begin{cases} H_0:\sigma\ge 10 \\ H_1:\sigma<10 \end{cases} \Leftrightarrow \begin{cases} H_0:\sigma^2\ge 100 \\ H_1:\sigma^2<100 \end{cases}$,$\alpha=0.05$,$n_2=6$,$s_2{}^2=23.8667$

② $\chi_0^2=\dfrac{(n-1)s^2}{\sigma_0^2}=\dfrac{5\times 23.8667}{100}=1.19333$

③ $\alpha=0.05$,$C=\{\chi_0^2 \mid \chi_0^2<\chi_{1-\alpha}^2(n-1)=\chi_{0.95}^2(5)=1.145476\}$

$\chi_0^2=1.19333\notin C$,$don't\ reject\ H_0$

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$\sigma<10$。

(二)生產線 1 與生產線 2 綜合統計量訊息

$n_1=5$,$\bar{x}_1=45$,$s_1{}^2=26$; $n_2=6$,$\bar{x}_2=58.6667$,$s_2{}^2=23.8667$

首先,檢定兩獨母體平均數是否相等,利用區間估計法判斷兩未知變異數是否相等。

變異數比值估計值為$\dfrac{s_1^2}{s_2^2}=\dfrac{26}{23.8667}=1.0894$

查表值 $F_{\alpha/2}(n_1-1,n_2-1)=F_{0.05}(4,5)=5.19$

$F_{1-\alpha/2}(n_A-1,n_B-1)=F_{0.95}(4,5)=\dfrac{1}{F_{0.05}(5,4)}=\dfrac{1}{6.26}=0.1597$

兩變異數比值σ_1^2/σ_2^2 90%信賴區間$\left(\dfrac{s_1^2}{s_2^2}\dfrac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\dfrac{s_1^2}{s_2^2}\dfrac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right)$

$(1.0894\times\dfrac{1}{5.19},1.0894\times\dfrac{1}{0.1597})\Rightarrow(0.2099,5.7479)$ 比值信賴區間包含 1,故隱含兩未知變異數是相等。

再者,檢定兩生產線平均值是否相等。

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

$n_1=5$,$\bar{x}_1=45$,$s_1{}^2=26$; $n_2=6$,$\bar{x}_2=58.6667$,$s_2{}^2=23.8667$

$s_p^2=\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}=\dfrac{4\times 26+5\times 23.8667}{5+6-2}=24.8148$

② $t^*=\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_1)}{\sqrt{s_p^2(\frac{1}{n_1}+\frac{1}{n_2})}}=\dfrac{(45-58.6667)-0}{\sqrt{24.8148(\frac{1}{5}+\frac{1}{6})}}=\dfrac{-13.667}{3.0164}=-4.53$

③ $\alpha=0.1$,$C=\{t^*:|t^*|>t_{\alpha/2}(n_1+n_2-2)=t_{0.05}(9)=1.833\}$

$t^* = -4.53 \in C$ ， $reject\ H_0$

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 $\mu_A \neq \mu_B$ ，即兩條生產線所生產燈泡壽命平均值是相等的。

三、三因子變異數分析表如下：

變異來源	平方和	自由度	均方	F值
X_1	200	2	A	F1
X_2	10	1	B	F2
X_3	48	2	C	F3
X_1X_2（交互作用）	876	D	E	F4
誤差	F	G	H	
總和	1358	35		

（假設資料符合變異數分析的假設，且各因子的水準數和水準值是由實驗者指定。）

(一)請填入表格中A、B、C、D、E、F、G和H的數值。（8分）

(二)寫出判定係數值（coefficient of determination R^2），因子X_1和X_2的水準數及實驗之反覆數（replicates）（或樣本大小(n)=？）。（12分）

(三)檢定因子X_2、X_3及X_1X_2效應是否顯著。顯著水準皆為0.05。（寫出虛無假設和對立假設，並說明檢定統計量之分配及檢定之結果。）（6分）

(四)母體標準差之估計值為何？（4分）

答：

(一) $A = \frac{200}{2} = 100$ ， $B = \frac{10}{1} = 10$ ， $C = \frac{48}{2} = 24$ ， $D = 2 \times 1 = 2$ ， $E = \frac{876}{2} = 438$ ，

$F = 1358 - 200 - 10 - 48 - 876 = 224$

$G = 35 - 2 - 1 - 2 - 2 = 28$ ， $H = \frac{224}{28} = 8$

(二)判定係數值 $R^2 = \frac{SSTr}{SSTO} = \frac{1358-224}{1358} = 0.8351 = 83.51\%$

因子 X_1 有 3 個水準與因子 X_2 有 2 個水準共有 6 種配方，反覆數為總樣本數除以配方數為 $\frac{36}{6} = 6$ 。

(三)

①H_0 :因子 X_2 效應不存在　v.s.　H_1 :因子 X_2 效應存在

H_0 :因子 X_3 效應不存在　v.s.　H_1 :因子 X_3 效應存在

H_0 :因子 X_1 與 X_2 無交互作用 v.s. H_1 :因子 X_1 與 X_2 有交互作用

②由 ANOVA 表可知檢定統計量

$F2 = \frac{10}{8} = 1.25$ ， $F3 = \frac{24}{8} = 3$ ， $F4 = \frac{438}{8} = 54.75$

③ $\alpha = 0.05$, $C2 = \{F^* \mid F^* > F_{0.05}(1,2) = 18.5\}$,

$C3 = C4 = \{F^* \mid F^* > F_{0.05}(2,2) = 19\}$

$F2 = 1.25 \notin C2$, $F3 = 3 \notin C3$, $F4 = 54.75 \in C4$

④因子 X_2 效應不存在與因子 X_3 效應不存在，且因子 X_1 與 X_2 有交互作用。

(四)母體標準差估計值為 $\sqrt{MSE} = \sqrt{8} = 2.8284$

四、隨機抽取六位大學男生修習統計和英語的成績如下：

男學生	1	2	3	4	5	6
統計	57	53	40	77	98	50
英語	60	36	18	84	99	10

(一)計算統計和英語成績的皮爾遜（Pearson）相關係數（r）。（6分）

(二)將統計和英語成績由小到大分別排序，並計算思匹爾門（Spearman）等級相關係數。（r_{sp}）。（7分）

(三)對(二)檢定這兩科目的成績等級是否存在顯著關係。顯著水準 $\alpha = 0.1$。（需寫出假設檢定的虛無假設和對立假設。）（7分）

(四)說明r和r_{sp}使用上的差異。（10分）

答：

(一) $n = 6$，$\sum x_i = 375$，$\sum x_i^2 = 25691$，$\sum y_i = 305$，$\sum y_i^2 = 22109$，$\sum x_i y_i = 22638$

$$SS_{xx} = \sum(x_i - \overline{x})^2 = \sum x_i^2 - n\overline{x}^2 = 2253.5$$

$$SS_{xy} = \sum(x_i - \overline{x})(y_i - \overline{y}) = \sum x_i y_i - n\overline{xy} = 3575.5$$

$$SS_{yy} = \sum(y_i - \overline{y})^2 = \sum y_i^2 - n\overline{y}^2 = 6604.833$$

相關係數 $r_{X,Y} = \dfrac{SS_{XY}}{\sqrt{SS_{XX}SS_{YY}}} = \dfrac{3575.5}{\sqrt{2253.5 \times 6604.833}} = 0.9268$

(二)

統計 X	57(4)	53(3)	40(1)	77(5)	98(6)	50(2)
英語 Y	60(4)	36(3)	18(2)	84(5)	99(6)	10(1)
d=X-Y	0	0	-1	0	0	1

$$r_{sp} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 = \frac{6 \times 2}{6 \times (36-1)} = 1 - \frac{2}{35} = 0.9429$$

故 Spearman 等級相關係數為 0.9429

(三)

① $\begin{cases} H_0 : \rho_{SP} = 0 \\ H_1 : \rho_{SP} \neq 0 \end{cases} \quad \alpha = 0.1$

② $r_{sp} = 0.9429$

③ $\alpha = 0.1$，$n = 6$ 雙尾，Spearman 等級相關係數臨界值為 0.829

$r_{sp} = 0.9429 > 0.829 \quad reject\ H_0$

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 $\rho_{SP} \neq 0$，即兩科目的成績存在顯著關係。

(四) Pearson 相關係數常用來呈現量變數之間線性相關程度；Spearman 等級相關係數則用來順序尺度變數之間的關聯性。