《統計學概要》

一、令隨機變數X為具有期望值等於μ，變異數等於σ^2的常態分配。

(一)若X的第20百分位數（the 20-th percentile）為2.5，第60百分位數（the 60-th percentile）為5.5，試求μ及σ。（5分）

(二)從上述常態分配之母體隨機抽取4個樣本，並計算其樣本平均數。試求樣本平均數大於μ＋σ的機率為何？（7分）

(三)從上述常態分配之母體隨機抽取1個樣本，若此樣本與平均數的距離小於σ，則可以獲得1元的獎金；若此樣本與平均數的距離介於σ至2σ之間，則可以獲得100元的獎金；若此樣本與平均數的距離大於2σ以上，則可以獲得1000元的獎金。試計算期望獲得的獎金為何？（8分）

試題評析	本題涉及常態分配之機率計算題，難度不高，惟有理解題意就能夠獲得滿分。
考點命中	《高點・高上統計學講義》第二回第七章第三節，趙治勳編撰。

答：

$X \sim N(\mu,\sigma^2)$

(一) $P(X \le 2.5) = 0.2 \Rightarrow P(Z \le \frac{2.5-\mu}{\sigma})$

$\Rightarrow \frac{2.5-\mu}{\sigma} = z_{0.8} = -0.84$ ------(1)

$P(X \le 5.5) = 0.6 \Rightarrow P(Z \le \frac{5.5-\mu}{\sigma})$

$\Rightarrow \frac{5.5-\mu}{\sigma} = z_{0.4} = 0.25$ ----------(2)

(1)(2)解聯立方程可得 μ=4.812,σ=2.752

(二) 樣本：$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu,\sigma^2)$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{4})$

$P(\bar{X} > \mu+\sigma) = P(Z > \frac{\mu+\sigma-\mu}{\sigma/\sqrt{4}}) = P(Z > 2) = 0.0228$

(三) 樣本：$X_1 \sim N(\mu,\sigma^2)$

$P(|X_1-\mu| \le \sigma) = P(|Z| \le 1) = 1-2\times 0.1587 = 0.6826$

$P(\sigma < |X_1-\mu| \le 2\sigma) = P(1 < |Z| \le 2) = 2\times[0.1587-0.0228] = 0.2718$

$P(|X_1-\mu| > 2\sigma) = 1-P(|X_1-\mu| \le 2\sigma) = 1-0.6826-0.2718 = 0.0456$

令 P 表獲得的獎金

$P=p$	1	100	1000
$f_P(p)$	0.6826	0.2718	0.0456

$E(P) = 1\times 0.6826 + 100\times 0.2718 + 1000\times 0.0456 = 73.4626$元

二、假設隨機變數 X_1 與 X_2 的聯合機率質量函數（joint probability mass function）為

$$f(x_1,x_2)=\begin{cases} k(x_1+2x_2) & x_1=1,2,3,\ \ x_2=2,3 \\ 0 & \text{其他} \end{cases}$$

(一)試求k值。（5分）

(二)試求機率 $P(X_1 \ge X_2)$。（5分）

(三)試求 X_1 與 X_2 的共變異數（covariance）$Cov(X_1,X_2)$。（10分）

試題評析	本題涉及機率論中間斷型隨機變數之基礎計算題，難度不高，應該容易獲得滿分。
考點命中	《高點．高上統計學講義》第一回第五章第三節，趙治勳編撰。

答：

(一) 根據機率公理假設 $\sum_{x_1=1}^{3}\sum_{x_2=2}^{3}k(x_1+2x_2)=1$

$$\sum_{x_1=1}^{3}\sum_{x_2=2}^{3}k(x_1+2x_2)=42k=1 \qquad \Rightarrow k=\frac{1}{42}$$

(二) $P(X_1 \ge X_2)=P(X_1=2\cap X_2=2)+P(X_1=3\cap X_2=2)+P(X_1=3\cap X_2=3)$

$$=\frac{6}{42}+\frac{7}{42}+\frac{9}{42}=\frac{11}{21}$$

(三) $Cov(X_1,X_2)=E(X_1X_2)-E(X_1)E(X_2)=\frac{113}{21}-\frac{44}{21}\times\frac{18}{7}=-\frac{1}{147}$

其中 $E(X_1X_2)=1\times2\times\frac{5}{42}+1\times3\times\frac{7}{42}+........+3\times3\times\frac{9}{42}=\frac{113}{21}$

$$E(X_1)=1\times\frac{12}{42}+2\times\frac{14}{42}+3\times\frac{16}{42}=\frac{44}{21}$$

$$E(X_2)=2\times\frac{18}{42}+3\times\frac{24}{42}=\frac{18}{7}$$

三、欲了解民眾對於不同品牌手機的偏好情況，針對不同性別進行最喜歡的手機品牌調查，得到人數資料如下表所示。（每小題10分，共20分）

		最喜歡的手機品牌			
		A	B	C	D
性別	男	49	44	49	39
	女	41	46	36	44

(一)試檢定男生對於四種品牌的手機偏好程度是否相同。（α＝0.05）

(二)試檢定性別與最喜歡的手機品牌是否有關。（α＝0.05）

試題評析	本題涉及卡方檢定的計算題，只要考生能夠由題意中判斷出來，就能夠獲得滿分。
考點命中	《高點．高上統計學講義》第三回第十三章，趙治勳編撰。

答：

(一)

品牌	A	B	C	D	
O_i	49	44	49	39	181
E_i	45.25	45.25	45.25	45.25	181

令 p_i 表喜歡品牌 i 之機率，$i = A, B, C, D$

$H_0: p_A = p_B = p_C = p_D$ vs H_1：至少一個$p_i \neq p_j, i \neq j$

T.S.：$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \overset{d}{\sim} \chi^2_{(4-1=3)}$

R.R.：Reject H_0 at $\alpha = 0.05$ if $\chi^{2*} > \chi^2_{0.05(3)} = 7.81$

$\because \chi^{2*} = 1.519$ $\therefore$ don't reject H_0

結論：我們沒有足夠證據去推論男生對於四種品牌手機偏好程度是不同的

(二)

O_i / E_i	A	B	C	D	
男	49/46.81	44/46.81	49/44.21	39/43.17	181
女	41/43.19	46/43.19	36/40.79	44/39.83	167
	90	90	85	83	348

H_0：性別與喜歡手機品牌無關 vs H_1：性別與喜歡手機品牌有關

T.S.：$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{4} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \overset{d}{\sim} \chi^2_{((2-1)(4-1)=3)}$

R.R.：Reject H_0 at $\alpha = 0.05$ if $\chi^{2*} > \chi^2_{0.05(3)} = 7.81$

$\because \chi^{2*} = 2.2205$ $\therefore$ don't reject H_0

結論：我們沒有足夠證據去推論性別與喜歡手機品牌有關

四、蒐集自變數X與應變數Y的資料，並將結果整理如下：（每小題10分，共20分）

$\sum_{i=1}^{10} x_i = 460$，$\sum_{i=1}^{10} x_i^2 = 23634$，$\sum_{i=1}^{10} y_i = 760$，$\sum_{i=1}^{10} y_i^2 = 59816$，$\sum_{i=1}^{10} x_i y_i = 36854$

(一)試求X與Y間的相關係數，並檢定母體相關係數是否為0。

(二)假設迴歸模型為 $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$，試求最小平方迴歸線，並檢定模型是否顯著。（$\alpha =$ 0.05）

試題評析	本題涉及簡迴歸的計算題，考古題中已經出現過很多次，大部份考生應該能夠獲得滿分。
考點命中	《迴歸分析申論題完全制霸》，高點文化出版，趙治勳編著，第二章。

答：

$SS_X = 23634 - \frac{(460)^2}{10} = 2474$，$SS_Y = 59816 - \frac{(760)^2}{10} = 2056$，

$SS_{XY} = 36854 - \frac{(460)(760)}{10} = 1894$

(一) $r_{XY} = \frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}} = 0.8398$

$H_0: \rho_{XY} = 0$ vs $H_1: \rho_{XY} \neq 0$

T.S.：$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(10-2=8)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $T^*>t_{0.025(8)}=2.306$

$\because |T^*|=4.375$ $\therefore$ reject H_0

結論：我們有足夠證據去推論母體相關係數不為 0

(二) $\hat{\beta}_1=\dfrac{SS_{XY}}{SS_X}=0.7656$ $\hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{X}=40.7824$

$\therefore \hat{y}=40.7824+0.7656x$

由於在簡迴歸下，H_0:模型不適當 vs H_1:模型適當 會等價於 $H_0:\rho_{XY}=0$ vs $H_1:\rho_{XY}\neq 0$,故可得 reject H_0

結論：我們有足夠證據去推論模型是顯著的。

五、分別自兩個獨立的常態母體抽取樣本，得到樣本資料如下表所示。（每小題10分，共20分）

母體	樣本數	樣本平均數及樣本變異數
$N(\mu_1,\sigma_1^2)$	15	$\overline{x}_1=6.5, s_1^2=0.65$
$N(\mu_2,\sigma_2^2)$	10	$\overline{x}_2=5.9, s_2^2=0.75$

(一)試檢定兩個母體的變異數是否相同。（α＝0.05）

(二)試檢定兩個母體的平均數是否相同。（α＝0.05）

試題評析	本題涉及兩獨立母體平均數與變異數的假設檢定之計算題，考古題中已經出現過很多次，大部份考生應該能夠獲得滿分。
考點命中	《高點・高上統計學講義》第三回第十一章，趙治勳編撰。

答：

母體： $X_1\sim N(\mu_1,\sigma_1^2)\perp X_2\sim N(\mu_2,\sigma_2^2)$

樣本： $X_{11},\ldots\ldots,X_{115}\overset{iid}{\sim}N(\mu_1,\sigma_1^2)\perp X_{21},\ldots\ldots,X_{210}\overset{iid}{\sim}N(\mu_2,\sigma_2^2)$

點： $\overline{X}_1\sim N(\mu_1,\dfrac{\sigma_1^2}{15})$ $\overline{X}_2\sim N(\mu_2,\dfrac{\sigma_2^2}{10})$

$$S_1^2=\frac{\sum(X_{1i}-\overline{X}_1)^2}{15-1}\qquad S_2^2=\frac{\sum(X_{2i}-\overline{X}_2)^2}{10-1}$$

(一) $H_0:\sigma_1^2=\sigma_2^2$ vs $H_1:\sigma_1^2\neq\sigma_2^2$

T.S.： $F=\dfrac{S_1^2}{S_2^2}\sim F_{(14,9)}$

R.R.：Reject H_0 at $\alpha=0.05$ if

$F^*<F_{0.975(14,9)}=\dfrac{1}{F_{0.025(9,14)}}$ 或 $F^*>F_{0.025(14,9)}=3.03$

$\because F^*=0.8667$ $\therefore$ don't reject H_0

（考卷未附 0.025 之 F 表，但由 0.05 之 F 可得不拒絕 H_0 ，就可以得知在由 0.025 之 F 也會得到不拒絕 H_0 之結論）

結論：我們沒有足夠證據去推論兩母體變異數不相同。

(二) $H_0:\mu_1=\mu_2$ vs $H_1:\mu_1\neq\mu_2$

T.S.：$T=\dfrac{(\overline{X}_1-\overline{X}_2)-(0)}{\sqrt{S_p^2(\dfrac{1}{15}+\dfrac{1}{10})}}\sim t_{(15+10-2=23)}$　其中 $S_p^2=\dfrac{(15-1)S_1^2+(10-1)S_2^2}{15+10-2}=0.6891$

R.R.：Reject H_0 at $\alpha=0.05$ if $|T^*|>t_{0.025(23)}=2.069$

$\because |T^*|=1.7705$ $\therefore$ don't reject H_0

結論：我們沒有足夠證據去推論兩母體平均數不相同。