《統計學概要》

參考值：$z_{0.2}=0.84$，$z_{0.05}=1.645$，$z_{0.1}=1.28$，$\chi^2_{1,0.05}=3.8415$，$\chi^2_{2,0.05}=5.9915$，$\chi^2_{3,0.05}=7.8147$

一、若機率密度函數$f(x)=3x^2$，$0\le x\le 1$。試求25及75百分位數各為多少？（8分）

試題評析	本題涉及單一連續型隨機變數機率函數及累積分配函數的應用。
考點命中	《高點‧高上統計學講義第一回》，趙治勳編撰，頁49。

答：

$$F_X(x)=P(X\le x)=\int_0^x 3t^2dt=x^3,0\le x\le 1$$

令$\eta_{0.25}$爲第25百分位數，$F_X(\eta_{0..25})=0.25\Rightarrow \eta_{0.25}^3=0.25\Rightarrow \eta_{0..25}=\sqrt[3]{0.25}=0.63$

令$\eta_{0.75}$爲第75百分位數，$F_X(\eta_{0..75})=0.75\Rightarrow \eta_{0.75}^3=0.75\Rightarrow \eta_{0..75}=\sqrt[3]{0.75}=0.909$

二、生產一批貨品其重量（單位為0.1公斤）如下莖葉圖（stem-and-leaf display）所示:

9	2 7
8	1 4 8
7	0 2 3 4 6 6 7
6	1 2 3 5 5 8
5	8
4	
3	8

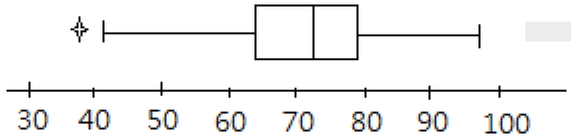
試畫出箱型圖（box plot），並判定是否有離群值？（12分）

試題評析	本題涉及利用盒形圖判斷離群值的方法，講義中完全命中。
考點命中	《高點‧高上統計學講義第一回》，趙治勳編撰，頁26（註）2。

答：

最小值=38，最大值=97，$Q_1=64,Q_2=72.5,Q_3=79$

$IQR=15$，$1.5\times IQR=22.5$，$3\times IQR=45$



觀察值38落在$(Q_1-3\times IQR,Q_1-1.5\times IQR)=(19,41.5)$間，故可以判斷其爲離群值

三、假設有兩條互相獨立的生產線是製造泡麵的，且製造出的泡麵重量各自服從常態分配，其變異數為 σ_1^2 及 σ_2^2。已知 $\sigma_1^2/\sigma_2^2=1.5$，求當 $P(s_1^2/s_2^2<k)=0.05$，k 值等於多少？s_1^2、s_2^2 分別是 σ_1^2 及 σ_2^2 的不偏估計量，定義回答中所用的符號。（12分）

試題評析	本題涉及兩獨立常態母體且均隨機樣本下，有關樣本變異數比之抽樣分配。
考點命中	《高點・高上統計學講義第二回》，趙治勳編撰，頁14。

答：

令 X_1, X_2 分別表該兩條製造泡麵的生產線 [假設隨機樣本]

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

樣本：$X_{11}, X_{12},\cdots,X_{1n_1} \overset{iid}{\sim} N(\mu_1,\sigma_1^2)$，$X_{21}, X_{22},\cdots,X_{2n_2} \overset{iid}{\sim} N(\mu_2,\sigma_2^2)$

$S_1^2 = \dfrac{\sum(X_{1i}-\overline{X}_1)^2}{n_1-1}, S_2^2 = \dfrac{\sum(X_{2i}-\overline{X}_2)^2}{n_2-1}$ 分別為 σ_1^2, σ_2^2 之不偏估計量

可得 $F=\dfrac{S_1^2}{S_2^2}\dfrac{\sigma_2^2}{\sigma_1^2} \sim F_{(n_1-1,n_2-1)}$

故 $P(F<F_{(n_1-1,n_2-1)0.95})=0.05 \Rightarrow P(\dfrac{S_1^2}{S_2^2}\dfrac{\sigma_2^2}{\sigma_1^2}<F_{(n_1-1,n_2-1)0.95})=0.05$

$\Rightarrow P(\dfrac{S_1^2}{S_2^2}<F_{(n_1-1,n_2-1)0.95}\dfrac{\sigma_1^2}{\sigma_2^2})=0.05$

$\Rightarrow k=F_{(n_1-1,n_2-1)0.95}\dfrac{\sigma_1^2}{\sigma_2^2}$

當 $\dfrac{\sigma_1^2}{\sigma_2^2}=1.5$ 時，$k=1.5\times F_{(n_1-1,n_2-1)0.95}$

四、相互獨立的樣本比例 $\hat{p}_1=x_1/n$ 及 $\hat{p}_2=x_2/n$ 是母體比例 p_1 與 p_2 的估計量，x_i 服從二項式分配(n, p_i)，i=1, 2。區間 $\hat{p}_1-\hat{p}_2\pm0.05$ 包含 p_1-p_2 的機率至少有80%，求 n 值為多少？（12分）

試題評析	本題涉及信賴區間之觀念，計算一個合理樣本數。
考點命中	《高點・高上統計學講義第二回》，趙治勳編撰，頁52。

答：

母體：$Y_1 \sim Ber(p_1) \perp Y_2 \sim Ber(p_2)$ [假設隨機樣本]

樣本：$Y_{11}, Y_{12},\cdots,Y_{1n} \overset{iid}{\sim} Ber(p_1)$，$Y_{21}, Y_{22},\cdots,Y_{2n} \overset{iid}{\sim} Ber(p_2)$

令 $X_1=\sum_{i=1}^{n}Y_{1i} \sim Bin(n,p_1)$ 及 $X_2=\sum_{i=1}^{n}Y_{2i} \sim Bin(n,p_2)$

$\hat{p}_1-\hat{p}_2=\dfrac{X_1-X_2}{n} \underset{by C.L.T.}{\sim} N(p_1-p_2, \dfrac{p_1(1-p_1)+p_2(1-p_2)}{n})$

$P(|(\hat{p}_1-\hat{p}_2)-(p_1-p_2)|\le 0.05)\ge 0.8$

$$\Rightarrow P(|Z| \le \frac{0.05}{\sqrt{\frac{p_1(1-p_1)+p_2(1-p_2)}{n}}}) \ge 0.8$$

$$\Rightarrow \frac{0.05}{\sqrt{\frac{p_1(1-p_1)+p_2(1-p_2)}{n}}} \ge z_{0..1} = 1.28$$

$$\Rightarrow n \ge \frac{(1.28)^2[p_1(1-p_1)+p_2(1-p_2)]}{(0.05)^2} \overset{p_1,p_2\text{代入}\frac{1}{2}}{=} 327.68\text{取}328$$

故 n 至少328個

五、陳老師上課時，突然讓70位同學小考，共出3題4選1的選擇題。學生答對題數如下：

答對的題數	0	1	2	3
學生人數	30	27	10	3

(一)試以 $a = 0.05$ 檢定學生是否每題都在猜答案？（12分）

(二)試以 $a = 0.05$ 檢定這3題答對的題數是否為二項式分配？（12分）

試題評析	本題涉及卡方適合度檢定，本題測試考生判題的能力，此檢定多數考生在考古題中必有練習過，但可能考生無法從本題目的敘述判斷出來。
考點命中	《高點・高上統計學講義第二回》，趙治勳編撰，頁116。

答：

(一)

答對題數	0	1	2	3
O_i	30	27	10	3
E_i	17.5	17.5	17.5	17.5

其中 E_i 爲當 H_0 爲真下之期望成功次數，$E_i = 70 \times \frac{1}{4} = 17.5,\ i = 1,2,3,4$

H_0：資料來自 $Bin(n=70, p=\frac{1}{4})$ vs H_1： not H_0

T.S.: $\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \overset{d}{\sim} \chi^2_{(4-1-0=3)}$

R.R.: Reject H_0 at $\alpha = 0.05$ if $\chi^{2*} > \chi^2_{0.05(3)} = 7.8147$

$\because \chi^{2*} = 29.314 \quad \therefore$ reject H_0

我們有足夠證據去推論學生不是用猜的。

(二)

答對題數	0	1	2	3
O_i	30	27	10	3
E_i	0.56	6.72	26.88	35.84

其中 E_i 爲當 H_0 爲真下之期望成功次數，

$X \overset{H_o isture}{\sim} Bin(n=70, \hat{p} = \bar{x} = 0.8)$ (註:額外估計一個參數)

$$E_1 = 70P(X=0) = 70\binom{3}{0}(0.8)^0(0.2)^3 = 70(0.008) = 0.56$$

$$E_2 = 70P(X=1) = 70\binom{3}{1}(0.8)^1(0.2)^2 = 70(0.096) = 6.72$$

$$E_3 = 70P(X=2) = 70\binom{3}{2}(0.8)^2(0.2)^1 = 70(0.384) = 26.88$$

$$E_4 = 70P(X=3) = 70\binom{3}{3}(0.8)^3(0.2)^0 = 70(0.512) = 35.84$$

H_0:資料來自二項分配vs H_1: not H_0

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

R.R.: Reject H_0 at $\alpha = 0.05$ if $\chi^{2*} > \chi^2_{0.05(3)} = 5.9915$

$\because \chi^{2*} = 1001.893$ $\therefore$ reject H_0

我們有足夠證據去推論學生答對的題數不為二項分配。

六、若X與Y之聯合機率密度函數為 $f(x,y)=8xy$ ，$0<x<y<1$：

(一)分別求X與Y的邊際機率密度函數。(8分)

(二)求X與Y的相關係數。(9分)

試題評析	本題想測試考生在多元隨機變數上之基礎運算能力。
考點命中	《高點・高上統計學講義第一回》，趙治勳編撰，頁63。

答：

(一)

$$f_X(x) = \int_x^1 8xy\,dy = 4x(1-x^2), 0<x<1$$

$$f_Y(y) = \int_0^y 8xy\,dx = 4y^3, 0<y<1$$

(二)

$$E(X) = \int_0^1 x4x(1-x^2)dx = \frac{8}{15},\ E(X^2) = \int_0^1 x^2 4x(1-x^2)dx = \frac{1}{3}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{11}{225}$$

$$E(Y) = \int_0^1 y4y^3dy = \frac{4}{5},\ E(Y^2) = \int_0^1 y^2 4y^3 dy = \frac{2}{3}$$

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{2}{75}$$

$$E(XY)=\int_0^1\int_x^1 xy8xydydx=\frac{4}{9}$$

$$\rho_{XY}=\frac{E(XY)-E(X)E(Y)}{\sqrt{V(X)}\sqrt{V(Y)}}=0.4924$$

七、X與Y為兩個獨立變數，且$\mathrm{E}(X)=\mathrm{E}(Y)=0$，$\mathrm{Var}(X)=5$，$\mathrm{Var}(Y)=6$。設$W=3X+2Y$，而$W$與$X$有直線關係，則$W$對$X$的斜率及截距係數各為多少？（15分）

試題評析	本題涉及迴歸分析，考生要先瞭解母體迴歸線之定義，然後計算過程中有用到條件期望值之觀念。
考點命中	《高點・高上統計學講義第一回》，趙治勳編撰，頁62。

答：

$$E(W\mid X)=E(3X+2Y\mid X)=3E(X\mid X)+2E(Y\mid X)=3X+2E(Y)=3X$$

截距項爲0，斜率爲3