《統計學》

試題評析	本年度考題：全部著重在觀念計算及數理推導，無母數考了三題比例上稍有偏高。全部考題都在講義中詳細論述說明，相信熟讀講義的同學應該可以考到90分以上。
考點命中	第一題：《高點・高上統計學講義》第三回，秦大成編撰，頁49《例題1》。 第二題：《高點・高上統計學講義》第四回，秦大成編撰，頁102《例題2》。 第三題：《高點・高上統計學講義》第七回，秦大成編撰，頁36《例題1》。 第四題：《高點・高上統計學講義》第四回，秦大成編撰，頁51《例題1》(3)。 第五題：《高點・高上統計學講義》第五回，秦大成編撰，頁51《例題1》。 第六題：《高點・高上統計學講義》第七回，秦大成編撰，頁8《例題3》。 第七題：《高點・高上統計學講義》第七回，秦大成編撰，頁16《例題2》。

一、一電阻能消耗的電力（W）與電壓（V）的平方成比例，即 $W=3V^2$。若通過電阻之電壓V呈平均值為6，變異數為1之常態分布。
（一）試求W之期望值，E(W)。（5分）
（二）試求W在120以上之機率。（10分）

答：

(一) $V \sim N(\mu=6, \sigma^2=1)$

$$EW = E(3V^2) = 3[Var(V)+(EV)^2] = 3(1+6^2) = 111$$

(二) $P(W>120) = P(3V^2>120) = P\left(V>\sqrt{40} \text{ 或 } V<-\sqrt{40}\right)$

$$= P\left(V>\sqrt{40}\right) + P\left(V<-\sqrt{40}\right) = P\left(Z>\frac{\sqrt{40}-6}{1}\right) + P\left(Z<\frac{-\sqrt{40}-6}{1}\right)$$

$$= P\left(Z>\frac{\sqrt{40}-6}{1}\right) = P(Z>0.3246)$$ (題目所給查表值查不到)

二、在一很大的母體中，已知某族群所占的比例介於（0.2,0.35）間，若要求抽樣結果該比例之估計誤差在1%內，而信賴度（confidence level）為95%，試求所需的最小樣本數。（10分）

答：

$0.2 < P < 0.35 \Rightarrow p_0 = 0.35$時 $P(1-P)$達max

$$b = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} \quad \Rightarrow \quad 0.01 = 1.96\sqrt{\frac{0.35(1-0.35)}{n}}$$

$\Rightarrow n = 8739.64$ $\quad \therefore$ 至少要取8740個樣本。

三、資料23　39　19　43　33　29　28　42　18　33　23　34　33　20　31　40為隨機抽取的一組樣本，試檢定其母體四分位數（quartile）是否為21，求其p-value。（答案不須乘開，寫出公式即可。）（10分）

答：

$n^- = 3$, $n^+ = 13$, $n' = 16$

① $\begin{cases} H_0：Q_1 = 21 \\ H_1：Q_1 \neq 21 \end{cases}$

② $\text{p}-\text{value}=2\cdot \text{P}\left(S\le 3 \middle| S\sim Bin(n'=16,\ p=\frac{1}{4}\right)$

$$=2\sum_{S=0}^{3} C_S^{16}\left(\frac{1}{4}\right)^S\left(\frac{3}{4}\right)^{16-S}\ \begin{cases}<\alpha \Rightarrow \text{reject}\ \ H_0\\ >\alpha \Rightarrow \text{Do not reject}\ \ H_0\end{cases}$$

四、若 $Y_1,Y_2,\ldots,Y_n$ 為獨立同分布之常態隨機變數，其期望值為 μ，變異數為 σ^2；μ 和 σ^2 皆未知。試求信賴水準為 $(1-\alpha)100\%$ 時，期望值 μ 之信賴區間長度的期望值。（10分）

答：

$$\mu \text{之}(1-\alpha)100\%\text{C.I.}=(\bar{Y}-t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}}\ ,\bar{Y}+t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}})$$

$\Rightarrow$ 信賴區間長度 $L=2\cdot t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}}$

令 $X=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

$$E\left(\frac{\sqrt{n-1}\,S}{\sigma}\right)=E(\sqrt{X})=\int_0^\infty\sqrt{x}\cdot\frac{x^{\frac{n-1}{2}-1}e^{-\frac{x}{2}}}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}}dx=\int_0^\infty\frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}}dx$$

$$=\frac{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}}\int_0^\infty\frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}dx=\frac{\Gamma(\frac{n}{2})\cdot\sqrt{2}}{\Gamma(\frac{n-1}{2})}\quad\Rightarrow\quad E[S]=\frac{\Gamma\left(\frac{n}{2}\right)\sqrt{2}}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}}\sigma$$

$$\therefore\ \ E[L]=E[2\cdot t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}}]=\frac{2\cdot t_{\frac{\alpha}{2}}(n-1)\cdot\Gamma\left(\frac{n}{2}\right)\sqrt{2}}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}\cdot\sqrt{n}}\sigma$$

五、下表分別為十位參加減重班的學員參加前與參加後的體重紀錄（Kg）：

參加減重班前	70	73	95	80	60	86	65	90	78	66
參加減重班後	65	71	89	82	60	83	66	86	75	65

若體重皆符合常態分配，試根據上述資料檢定該減重班是否具有顯著的減重功效？（$\alpha=5\%$）（10分）

答：

X_i	70	73	95	80	60	86	65	90	78	66
Y_i	65	71	89	82	60	83	66	86	75	65
$D_i=X_i-Y_i$	5	2	6	-2	0	3	-1	4	3	1

$\bar{D}=\frac{\sum_{i=1}^n D_i}{n}=2.1$ ， $S_D^2=\frac{\sum_{i=1}^n D_i^2-n\bar{D}^2}{n-1}=\frac{105-10\times 2.1^2}{9}=0.7519$

① $\begin{cases}H_0：\mu_D\le 0\\ H_1：\mu_D>0\end{cases}$ (減重班有顯著的減重功效)

② $C=\left\{T\middle|T>\underset{=}{t_{0.05}}(10-1)=1.833\right\}$

③ $T=\frac{\bar{D}}{S_D/\sqrt{n}}=\frac{2.1}{\sqrt{0.7519}/\sqrt{10}}=7.6584\in C$

∴ reject H_0 ，有充分的證據顯示：減重班有顯著的減重功效。

六、下表是投擲一骰子300次所得各點數出現的次數。

點數	1	2	3	4	5	6
次數	35	60	52	65	55	33

在顯著水準α＝5%下，請檢定該骰子是否為公平的骰子。（10分）

答：

令 P_i：出現ith面機率 ，$i=1,2,3,4,5,6$

① $\begin{cases}H_0：P_i=P_2=\cdots=P_6=\frac{1}{6} & \text{(骰子公正)}\\ H_1：H_0\text{不成立}\end{cases}$

② $C=\{Q|Q>\chi^2_{0.05}(6-1)=11.07\}$

③ $Q=\sum_{i=1}^{6}\frac{(O-E_i)^2}{E_i}=17.36\in C$

∴ reject H_0 ，有充分的證據顯示：骰子不公正。

七、在一項實驗中，欲研究使用一種藥性貼布是否有助於紓緩疼痛。實驗過程中隨機安排患者使用外型相同的藥性貼布或無藥貼布，實驗結束時得數據如下：

	紓緩疼痛	未紓緩	合計
使用藥性貼布	25	31	56
使用無藥貼布	10	40	50

試問實驗數據是否證實該藥性貼布可以顯著達到紓緩疼痛的效果？試以p-value做結論。（15分）

答：

P_1:使用藥布舒緩疼痛 ，P_2:使用無藥布舒緩疼痛

$\widehat{P}_1=\frac{25}{56}=0.4464$ ， $\widehat{P}_2=\frac{10}{50}=0.2$ ， $\widehat{P}=\frac{25+10}{56+50}=0.33$

① $\begin{cases}H_0：P_1-P_2\le 0\\ H_1：P_1-P_2>0 \ (\because \widehat{P}_1-\widehat{P}_2>0)\end{cases}$

② $P-\text{value}=P\left(Z>\frac{\widehat{P}_1-\widehat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_1}\right)}}\right)=P\left(Z>\frac{0.4464-0.2}{\sqrt{0.33(1-0.33)\left(\frac{1}{56}+\frac{1}{50}\right)}}\right)$

$=P(Z>2.73)\cong 0.0036<\alpha$

∴ reject H_0 ，有充分的證據顯示：藥性貼布可以顯著舒緩疼痛的功效。

八、若已知獨立資料(x_i,y_i)，$i=1,\dots,5$，分別為$(-2,0)$，$(-1,0)$，$(0,1)$，$(1,1)$，$(2,3)$，且滿足線性模型$y_i=1+\beta x_i+\varepsilon_i$，其中$\varepsilon_i$為獨立同分布之$N(0,\sigma^2)$隨機變數，$i=1,\dots,5$。（每小題10分，共20分）

(一)試求斜率β之最小平方估計（Least Squares Estimate）。

(二)試檢定上述迴歸模型是否顯著，並寫出檢定統計量之分布。

答：

(一) 令 $Q=\sum_{i=1}^{n}(Y_i-\widehat{Y}_i)^2=\sum_{i=1}^{n}(Y_i-1-\widehat{\beta}x_i)^2$

$$\frac{dQ}{d\widehat{\beta}_1}=-2\sum_{i=1}^{n}\left[(Y_i-1-\widehat{\beta}x_i)\cdot x_i\right]=0$$

$$\sum_{I=1}^{n}x_iY_i-\sum_{i=1}^{n}x_i-\widehat{\beta}\sum_{i=1}^{n}x_i^2=0$$

$$\widehat{\beta}=\frac{\sum_{i=1}^{n}x_iY_i-\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_i^2}=\frac{7-0}{10}=0.7$$

(二)

$$\mathrm{Var}(\widehat{\beta})=\mathrm{Var}\left(\frac{\sum_{i=1}^{n}x_iY_i-\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_i^2}\right)=\frac{\sum_{i=1}^{n}x_i^2\cdot\mathrm{Var}(Y_i)}{(\sum_{i=1}^{n}x_i^2)^2}=\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{(\sum_{i=1}^{n}x_i^2)^2}=\frac{\sigma^2}{\sum_{i=1}^{n}x_i^2}$$

$$\mathrm{SSTO}=\sum_{i=1}^{n}Y_i^2-n\bar{Y}^2=11-5\times1^2=6$$

$$\mathrm{SSR}=\frac{S_{XY}^2}{S_{XX}}=\frac{(\sum_{i=1}^{n}x_iY_i-\sum_{i=1}^{n}x_i)^2}{\sum_{i=1}^{n}x_i^2}=\frac{(7-0)^2}{10}=4.9$$

$$\mathrm{MSE}=\frac{SSTO-SSR}{n-1}=\frac{1.1}{4}=0.275$$

假設$\alpha=0.05$

①$\begin{cases}H_0: \beta=0\\ H_1: \beta\neq0\ \text{(迴歸模型不顯著)}\end{cases}$

②$C=\left\{T\,\middle|\,|T|>t_{\underset{=}{0.05}}(5-1)=2.776\right\}$

③ $|T|=\left|\frac{\widehat{\beta}}{\sqrt{\frac{\mathrm{MSE}}{\sum_{i=1}^{n}x_i^2}}}\right|=\left|\frac{0.7}{\sqrt{\frac{0.275}{10}}}\right|=4.22\in C$

$\therefore$ reject H_0，有充分的證據顯示：迴歸模型不顯著。