《統計學概要》

試題評析

第一題：本題比較特別的是利用Z分數判斷離群值，當$|Z|>3$時視爲離群值。
第二題：本題是考柴比雪夫不等式，只要有記定理的敘述就能夠解題。
第三題：本題是考抽樣分配，國考考古題中有相似的題目，有做考古題者應該可以獲得分數。
第四題：本題是迴歸分析基本計算題，跟講義中的例一相同。
第五題：本題是無母數統計學中的適合度檢定，作答時只要注意有兩個參數額外估計，檢定統計量的自由度要多減一個2。

$z_{0.05}=1.645$， $z_{0.025}=1.96$， $z_{0.25}=0.67$， $t_{0.025}(3)=3.182$， $t_{0.05}(3)=2.353$， $t_{0.025}(4)=2.776$，
$t_{0.05}(4)=2.1322$， $\chi^2_{0.05}(1)=3.84$， $\chi^2_{0.05}(2)=5.99$， $\chi^2_{0.05}(3)=7.81$， $F_{0.05}(1,3)=10.13$， $F_{0.05}(3,1)=215.71$

一、假設有5個觀察值：1, 2, 3, 4, 20
(一)請計算出上述觀察值之Z分數（Z-score）。（10分）
(二)請說明20是否可以視為是一個離群值（outlier）？（5分）

答：(一)

觀察值	1	2	3	4	20
Z分數	$\frac{1-6}{7.071}=-0.707$	-0.566	-0.424	-0.283	1.980

(二)由於觀察值20之Z分數絕對值沒有大於3，故不會視爲離群值。

二、(一)假設有一隨機變數U，而且已知其$E(U)=4$，$E(U^2)=25$，請計算出$\Pr(0<U<8)$機率之下界值？（10分）
(二)假設有一隨機變數V，而且已知其$\Pr(V\geq 8)=0.4$，$\Pr(V\leq 4)=0.2$，$E(V)=6$，請計算出V的變異數（σ_V^2）之下界值？（10分）

答：(一)由題目資訊得知, $\mu_U=E(U)=4$, $\sigma_U^2=V(U)=E(U^2)-[E(U)]^2=9$

$$\Pr(0<U<8)=\Pr(4-(\frac{4}{3})3<U<4+(\frac{4}{3})3)$$

$$=\Pr(|U-4|<(\frac{4}{3})3)>1-\frac{1}{(\frac{4}{3})^2}=\frac{7}{16}$$

故$\Pr(0<U<8)$的下界爲$\frac{7}{16}$

(二)本題有兩個方法可以解：(1)柴比雪夫不等式(2)單邊柴比雪夫不等式

(1)柴比雪夫不等式

由題意得知 $\Pr(4<V<8)=0.4 \Rightarrow \Pr(|V-6|<2)=0.4>1-\frac{1}{k^2}$

得 $0.4>1-\frac{1}{k^2} \Rightarrow k>1.29$

由柴比雪夫不等式得知，$2<1.29\sigma_V \Rightarrow \sigma_V^2>2.404$

(2)單邊柴比雪夫不等式

由題意得知 $\Pr(V\ge 8)=\Pr(V\ge 6+2)=0.4\le \frac{\sigma_V^2}{\sigma_V^2+2^2} \Rightarrow \sigma_V^2 \ge 2.667$

又 $\Pr(V\le 4)=\Pr(V\le 6-2)=0.2\le \frac{\sigma_V^2}{\sigma_V^2+2^2} \Rightarrow \sigma_V^2\ge 1$

合併以上兩個下界，得知 $\sigma_V^2\ge 2.667$

三、假設一個袋子裡有五粒球，球的外表都一樣，球的編號為0, 1, 1, 1, 2；假設某甲隨機分別抽出兩顆球，抽球方式是允許放回方式（with replacement）。假設 X_1 與 X_2 分別表示某甲第一次抽的球號與第二次抽到的球號。

(一)請計算出 $\bar{X}=\frac{X_1+X_2}{2}$ 之抽樣分配？（10分）

(二)請計算出 S^2 之抽樣分配？（提示 $S^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$）（10分）

答：(一)樣本組合有(0,0),(1,1),(2,2),(0,1),(0,2),(1,2)

故 $\bar{X}$ 之隨機變量有0,0.5,1,1.5,2

$P(\bar{X}=0)=\frac{1}{5}\times\frac{1}{5}=\frac{1}{25}$，

$P(\bar{X}=0.5)=2\times\frac{1}{5}\times\frac{3}{5}=\frac{6}{25}$ (乘2是考慮(0,1),(1,0)兩個組合)，

$P(\bar{X}=1)=2\times\frac{1}{5}\times\frac{1}{5}+\frac{3}{5}\times\frac{3}{5}=\frac{11}{25}$，$P(\bar{X}=1.5)=2\times\frac{3}{5}\times\frac{1}{5}=\frac{6}{25}$，

$P(\bar{X}=2)=\frac{1}{5}\times\frac{1}{5}=\frac{1}{25}$

$\bar{X}=\bar{x}$	0	0.5	1	1.5	2
$f_{\bar{X}}(\bar{x})$	1/25	6/25	11/25	6/25	1/25

(二)樣本組合有(0,0),(1,1),(2,2),(0,1),(0,2),(1,2)

故 S^2 之隨機變量有0,0.5,2

$P(S^2=0)=2\times\frac{1}{5}\times\frac{1}{5}+\frac{3}{5}\times\frac{3}{5}=\frac{11}{25}$，

$$P(S^2=0.5)=2\times\frac{1}{5}\times\frac{3}{5}+2\times\frac{3}{5}\times\frac{1}{5}=\frac{12}{25},$$

$$P(S^2=2)=2\times\frac{1}{5}\times\frac{1}{5}=\frac{2}{25}$$

$S^2=s^2$	0	0.5	2
$f_{S^2}(s^2)$	11/25	12/25	2/25

四、假設給定如下資料：

X	6	10	14	18	22
Y	8.6	6.1	8.4	14.2	16.3

(一)請求X與Y之樣本相關係數 $r_{x,y}$ ？（5分）

(二)請求出迴歸線 $\hat{Y}=b_0+b_1X$ ？（5分）

(三)請求出斜率之95%信賴區間？（5分）

(四)請用F-檢定迴歸線是否顯著？（顯著水準= 0.05）（10分）

答：$SS_{XY}=\sum X_iY_i-\frac{\sum X_i\sum Y_i}{n}=94$，$SS_X=\sum X_i^2-\frac{(\sum X_i)^2}{n}=160$，

$$SS_Y=\sum Y_i^2-\frac{(\sum Y_i)^2}{n}=74.468$$

(一) $r_{XY}=\frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}}=0.8612$

(二) $b_1=\frac{SS_{XY}}{SS_X}=0.5875$，$b_0=\overline{Y}-b_1\overline{X}=2.495$　　$\therefore\hat{y}=2.495+0.5875x$

(三)斜率之95%信賴區間為 $(b_1\mp t_{(3)0.025}\sqrt{\frac{MSE}{SS_X}})=(0.5875\mp 3.182\sqrt{\frac{6.414}{160}})$

$$=(-0.0495,1.2245)$$

其中 $MSE=\frac{SSE}{5-1-1}=\frac{SST-SSR}{3}=\frac{SS_Y-b_1^2SS_X}{3}=6.414$

(四) H_0：模型是不適當的 vs H_1：模型是適當的

T.S.: $F=\frac{MSR}{MSE}\sim F_{(1,3)}$

R.R.: Reject H_0 at α =0.05 if $F^*>F_{(1,3)0.05}=10.128$

$\because F^*=\frac{SSR/1}{MSE}=\frac{b_1^2SS_X/1}{MSE}=8.61$　　$\therefore don't\, reject H_0$

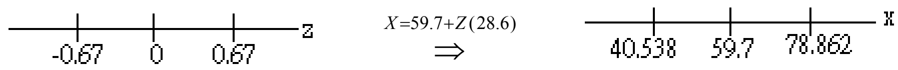
結論：我們沒有足夠證據去推論模型是適當的（顯著的）。

五、假設以下20筆資料是某班級同學統計學期末考分數：

17, 18, 22, 27, 30, 30, 43, 46, 54, 63, 66, 71, 75, 82, 82, 88, 91, 93, 97, 99

請以顯著水準= 0.05檢定這20位同學統計學期末考分數是否為常態分配？（此20筆之平均數 $\overline{X}=59.7$，標準差 $S=28.6$）。（20分）

答：本題是考卡方適合度檢定，首先會面臨到分組的問題，由考卷上給的查表值 $z_{0.25}=0.67$ 預測老師希望考生用三個四分位數分成四組，以下先算出組界：



令X～N（$59.7, 28.6^2$）

	$X<40.538$	$40.538\le X<59.7$	$59.7\le X<78.862$	$78.862\le X$
觀察次數 O_i	6	3	4	7
期望次數 E_i	5	5	5	5

H_0 :資料服從常態分配 vs H_1 : not H_0

T.S.: $\chi^2=\sum_{i=1}^{4}\frac{(O_i-E_i)^2}{E_i}\sim\chi^2_{(4-2-1=1)}$ (這裡也可以考慮Yate's修正)

R.R.: Reject H_0 at $\alpha=0.05$ if $\chi^{2*}>\chi^2_{0.05(1)}=3.84$

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

我們沒有足夠證據去推論期末考分數不是來自於常態分配。