《統計學概要》

一、令Φ(z)為標準常態累積分配函數，Φ(−2) = 0.0228。計算回答下列各子題：（每小題10分，共30分）

（一）假設X_1和X_2是互相獨立的常態隨機變數，其分配分別為$N(12,4^2)$和$N(2,3^2)$。計算$P(X_1 > X_2)$和$P(X_1 + X_2 > 26)$。

（二）陳述中央極限定理（Central Limit Theorem）。（詳實敘明所需要的假設）

（三）令 $Y = \sum_{i=1}^{48} X_i$ ，X_i為服從齊一分配（uniform distribution）$U(0,4)$的隨機樣本，i =1,…,48。利用(二)所述定理，計算$P= (80 < Y < 112)$之近似機率。（需計算列出X_i的平均數與變異數）

試題評析	常態分配一直以來都是熱門考題，相關的性質和定理務必要掌握。

答：

(一)$X_1 \sim N(12, 4^2), X_2 \sim N(2, 3^2), X_1 \perp\!\!\!\perp X_2$

$X_1 - X_2 \sim N(10, 25), X_1 + X_2 \sim N(14, 25)$

則

$$P(X_1 > X_2) = P(X_1 - X_2 > 0) = P\left(\frac{(X_1 - X_2) - 10}{\sqrt{25}} > \frac{0 - 10}{\sqrt{25}}\right) = P(Z > -2) = 1 - \Phi(-2) = 0.9772$$

$$P(X_1 + X_2 > 26) = P\left(\frac{(X_1 + X_2) - 14}{\sqrt{25}} > \frac{26 - 14}{\sqrt{25}}\right) = P(Z > 2.4) = 0.0082$$

(二)給定隨機樣本 $X_1, X_2, \cdots, X_n$ 來自於平均數 μ 變異數 σ^2 的分配，其中 $\mu, \sigma^2 < \infty$。當樣本數 n 足夠大時，樣本平均的的抽樣分配會近似於常態分配。

(三)$\{X_i\}_{i=1}^{48} \overset{i.i.d.}{\sim} U(0, 4), Y = \sum_{i=1}^{48} X_i$

$$E(X_i) = \frac{0 + 4}{2} = 2, Var(X_i) = \frac{(4 - 0)^2}{12} = \frac{4}{3}$$

$$P(80 < Y < 112) = P\left(80 < \sum_{i=1}^{48} X_i < 112\right) = P\left(\frac{5}{3} < \bar{X} < \frac{7}{3}\right) = P\left(\frac{\frac{5}{3} - 2}{\sqrt{\frac{4}{3 \times 48}}} < \frac{\bar{X} - 2}{\sqrt{\frac{4}{3 \times 48}}} < \frac{\frac{7}{3} - 2}{\sqrt{\frac{4}{3 \times 48}}}\right)$$

$$\overset{by\ CLT}{\approx} P(-2 < Z < 2) = 0.9544$$

二、一家液體洗滌劑製造商生產的洗滌劑標示宣稱每瓶容量為450毫升（mL）。隨機抽取14瓶，測量其容量，資料列於下表：

447	459	439	443	462	449	437
458	453	461	445	467	456	448

數據的常態機率圖顯示可以假設內容量呈常態分配。μ表示該製造商生產的所有洗滌劑瓶子的平均容量。要確定平均容量是否少於標示所宣稱的容量，回答計算下列各子題：

（一）敘明虛無假設與對立假設。（5分）

（二）在顯著水準α=0.05下，依據(一)所敘明的假設執行統計檢定，含棄卻域和結論。（15分）

（三）如果常態分配假設不成立，但是資料的分配仍具有對稱分配時，可採用何種無母數統計檢定？（5分）

（四）在顯著水準α=0.05下，依(三)之統計檢定對此資料進行分析檢定。（13分）

（$t_{0.05,14}$=1.761，$t_{0.05,13}$=1.771，$t_{0.05,12}$=1.782，$z_{0.05}$=1.645，$z_{0.025}$=1.960，$w_{0.05,14}$=26，$w_{0.95,14}$=79。）

試題評析	本題為均數檢定與Wilcoxon符號等級檢定，注意資料中有等級相同的狀況。

答：

(一)$H_0: \mu = 450$ against $H_1: \mu < 450$

(二)$\varphi = \frac{\bar{X}-450}{\frac{S}{\sqrt{14}}} \sim t(14-1)$

$\alpha = 0.05, RR = \{\varphi | \varphi < -t_{0.05}(13) = -1.771\}$

$\bar{x} = \frac{1}{14}\sum_{i=1}^{14} x_i = \frac{6324}{14} = 451.71, s = \sqrt{\frac{\sum_{i=1}^{14} x_i^2 - 14 \cdot \bar{x}^2}{14-1}} = 9.12$

$\Rightarrow \varphi^* = \frac{451.71 - 450}{\frac{9.12}{\sqrt{14}}} = 0.702 \notin RR$

∴ 無法拒絕虛無假設

我們沒有足夠證據推論平均容量少於標示所宣稱的容量

(三)Wilcoxon符號等級檢定

(四)

X_i	447	459	439	443	462	449	437	458	453	461	445	467	456	448
d_i	-3	9	-11	-7	12	-1	-13	8	3	11	-5	17	6	-2
$\|d_i\|$	3	9	11	7	12	1	13	8	3	11	5	17	6	2
R_i	3.5	9	10.5	7	12	1	13	8	3.5	10.5	5	14	6	2

令 m_e 表瓶子容量的中位數

$H_0: m_e = 450$ against $H_1: m_e < 450$

$\alpha = 0.05, RR = \{W \big| W < w_{0.05,14} = 26\}$

$W^+ = 9 + 12 + 8 + 3.5 + 10.5 + 14 + 6 = 63$

$W^- = 3.5 + 10.5 + 7 + 1 + 13 + 5 + 2 = 42$

$W = 63 \notin RR$

∴ 無法拒絕虛無假設

我們沒有足夠證據推論容量中位數少於標示所宣稱的容量

三、某一特徵被認為存在於三種族群，某研究欲檢定各族群具有此一特徵之比例均為20%。分別從此三個族群中抽取60、120和60的隨機樣本進行測試，檢驗結果如下表所示：

	有顯現	無顯現
族群一	28	32
族群二	30	90
族群三	25	35

(一)寫出虛無假設與對立假設。(5分)

(二)在顯著水準α=0.05下，寫出檢定統計量、計算過程、棄卻域和結論。(15分)

($\chi^2_{0.05,4} = 9.49$，$\chi^2_{0.05,3} = 7.81$，$\chi^2_{0.05,2} = 5.99$，$t_{0.05,3} = 2.353$，$t_{0.05,4} = 2.132$，$t_{0.05,2} = 2.920$。)

試題評析	本題為卡方檢定中的適合度檢定，務必按照題目要求把過程寫清楚。

答：

(一)H_0: 各族群具有此一特徵比例均為20% against H_1: 各族群具有此一特徵比例不均為20%

(二)

	族群一	族群二	族群三
O_i	28	30	25
E_i	12	24	12

$$\varphi=\sum_{i=1}^{3}\frac{(O_i-E_i)^2}{E_i}\sim\chi^2(2)$$

$\alpha=0.05, RR=\{\varphi|\varphi>\chi^2_{0.05}(2)=5.99\}$

$$\varphi^*=\frac{(28-12)^2}{12}+\frac{(30-24)^2}{24}+\frac{(25-12)^2}{12}=\frac{443}{12}\in RR$$

$\therefore$ 拒絕虛無假設

我們有足夠證據推論各族群具有此一特徵比例不均為20%

四、某種抹片檢查用於檢測女性某種癌症，假設對於患有這種癌症的女性，大約有15%的假陰性（false negative）檢測結果。對於沒有這種癌症的女性，大約有20%的假陽性（false positive）檢測結果。假設每100,000人中大約有8名女性患有這種癌症。在某一抹片檢查呈陽性的情況下，計算此女性得此癌症的條件機率。（請定義各事件的符號，並敘明所採用的計算公式。）（12分）

試題評析	經典貝式定理題目，依照題目要求詳細說明，即可得分。

答：

令 A 表患有該癌症，A^c 表沒有該癌症，$+$ 表檢測結果陽性，$-$ 表檢測結果陰性

由題目可知

$P(-|A)=0.15, P(+|A^c)=0.2, P(A)=0.00008$

則所求

$$\begin{aligned}P(A|+)=\frac{P(A\cap +)}{P(+)}=\frac{P(A\cap +)}{P(A\cap +)+P(A^c\cap +)}&=\frac{P(+|A)\cdot P(A)}{P(+|A)\cdot P(A)+P(+|A^c)\cdot P(A^c)}\\&=\frac{[1-P(-|A)]\cdot P(A)}{[1-P(-|A)]\cdot P(A)+P(+|A^c)\cdot[1-P(A)]}\\&=\frac{0.85\cdot 0.00008}{0.85\cdot 0.00008+0.2\cdot 0.99992}\\&=\frac{17}{50013}\\&\approx 0.0003\end{aligned}$$