《統計學》

參考值：

$z_{0.0062}=2.5$，$z_{0.025}=1.96$，$z_{0.05}=1.645$，$z_{0.1}=1.28$，$z_{0.1587}=1.0$，$z_{0.2743}=0.6$， $z_{0.3085}=0.5$

$t_{0.025,8}=2.306$，$t_{0.025,9}=2.262$，$t_{0.025,10}=2.228$，$t_{0.025,11}=2.201$，$t_{0.025,25}=2.060$，$t_{0.025,26}=2.056$

$t_{0.05,8}=1.860$，$t_{0.05,9}=1.833$，$t_{0.05,10}=1.812$，$t_{0.05,11}=1.796$，$t_{0.05,25}=1.708$，$t_{0.05,26}=1.706$

$\chi^2_{0.025,2}=7.378$，$\chi^2_{0.025,3}=9.348$，$\chi^2_{0.025,4}=11.143$，$\chi^2_{0.025,5}=12.833$，$\chi^2_{0.025,25}=40.647$，$\chi^2_{0.025,26}=41.923$

$\chi^2_{0.05,2}=5.991$，$\chi^2_{0.05,3}=7.815$，$\chi^2_{0.05,4}=9.488$，$\chi^2_{0.05,5}=11.070$，$\chi^2_{0.05,25}=37.652$，$\chi^2_{0.05,26}=38.885$

$\chi^2_{0.975,2}=0.0506$，$\chi^2_{0.975,3}=0.216$，$\chi^2_{0.975,4}=0.484$，$\chi^2_{0.975,5}=0.831$，$\chi^2_{0.975,25}=13.120$，$\chi^2_{0.975,26}=13.844$

$F_{0.025,1,10}=6.94$，$F_{0.025,2,9}=5.71$，$F_{0.05,1,10}=4.96$，$F_{0.05,2,9}=4.26$，$F_{0.1,1,10}=3.285$，$F_{0.1,2,9}=3.006$

一、假設某抗肥胖藥物可由政府補助之條件如下：

條件一：有心疾者、且身高體重指數（body-mass index，簡稱BMI）大於28；

條件二：無心疾者、且BMI大於32。

已知有心疾者占總人口的10%。若有心疾者的BMI平均值為25、標準差為5；無心疾者的BMI平均值為22、標準差為4。請回答下列問題：（每小題10分，共20分）

(一)試問無心疾者中，有多少比例可獲補助？

(二)若有心疾者與無心疾者的BMI分別都服從常態分配，試問可獲補助者中，屬於有心疾者的比例為何？

試題評析	本題屬於貝氏定理之計算題型，還考到單邊柴比雪夫不等式，中間再考到常態分配之事件機率，雖然計算難度不高，但容易讓考生不明題意而失去分數。
考點命中	《高點・高上統計學講義》第一回，趙治勳編撰，第五章第五節：單邊柴比雪夫不等式，頁110；第四章第四節：貝氏定理，頁65。

答：

令r.v.X表BMI值

(一) 在無心疾者下 $X\sim(22,4^2)$

由單邊柴比雪夫不等式可得

$$P(X>32)=P(X>22+10)\le\frac{4^2}{4^2+10^2}=0.1379$$

故無心疾者中，至多有 13.79%可獲得補助

(二) 有心疾者中，P(可獲得補助)= $P(X>28)=P(Z>\frac{28-25}{5})$

$$=P(Z>0.6)=0.2743$$

無心疾者中，P(可獲得補助)= $P(X>32)=P(Z>\frac{32-22}{4})$

$$=P(Z>2.5)=0.0062$$

P(有心疾者 | 可獲得補助)= $\dfrac{0.1\times0.2743}{0.1\times0.2743+0.9\times0.0062}=0.831$

二、甲、乙兩個長期照顧居家服務中心統計其每個月的服務人次，資料如下：

	樣本數（月）	樣本平均數	樣本標準差
甲	12	58	4
乙	15	65	5

假設兩母體皆服從常態分配，且變異數相同（$\sigma_1^2=\sigma_2^2=\sigma^2$）。請回答下列問題：（每小題10分，共20分）

(一)試求σ^2的估計值，以及其95%信賴區間。

(二)在顯著水準0.05之下，試檢定兩母體平均數是否相等。

試題評析	本題屬於兩母體平均數之假設檢定之計算題型，還有考到共同母體變異數σ^2之點估計與信賴區間，這部份是首次命題，同學需要透過S_p^2作出樞紐量，對於能夠舉一反三的同學來說，應該也可以輕鬆獲得滿分。
考點命中	《高點・高上統計學講義》第二回，趙治勳編撰，第十一章第三節，頁59。

答：

令 X_1,X_2 分別甲、乙服務中心之每個月服務人次

母體：$X_1\sim N(\mu_1,\sigma^2)\perp X_2\sim N(\mu_2,\sigma^2)$　假設 (1)$X_1\perp X_2$(2)隨機樣本

樣本：$X_{11},\ldots,X_{112}\overset{iid}{\sim}N(\mu_1,\sigma^2)$　$X_{21},\ldots,X_{215}\overset{iid}{\sim}N(\mu_2,\sigma^2)$

(一) $\hat{\sigma}^2=S_p^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}=\dfrac{(12-1)(4^2)+(15-1)(5^2)}{12+15-2}=21.04$

樞紐量：$\dfrac{(25)S_p^2}{\sigma^2}=\dfrac{(n_1+n_2-2)S_p^2}{\sigma^2}=\dfrac{(n_1-1)S_1^2}{\sigma^2}+\dfrac{(n_2-1)S_2^2}{\sigma^2}\sim\chi^2_{(12+15-2=25)}$

機率區間：$P(\chi^2_{(25)0.975}\le\dfrac{(25)S_p^2}{\sigma^2}\le\chi^2_{(25)0.025})=0.95$

結論：σ^2之95%C.I.為 $(\dfrac{(25)S_p^2}{\chi^2_{(25)0.025}},\dfrac{(25)S_p^2}{\chi^2_{(25)0.975}})=(\dfrac{(25)(21.04)}{40.647},\dfrac{(25)(21.04)}{13.120})$

$=(12.941,40.091)$

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

T.S.：$T=\dfrac{(\bar{X}_1-\bar{X}_2)-(0)}{\sqrt{S_p^2(\frac{1}{12}+\frac{1}{15})}}\sim t_{(25)}$

R.R.：Reject H_0 at $\alpha=0.05$ if $|T^*|>t_{(25)0.025}=2.060$

$\because|T^*|=|\dfrac{(58\text{-}65)-(0)}{\sqrt{(21.04)(\frac{1}{12}+\frac{1}{15})}}|=3.94$　$\therefore$ reject H_0

我們有足夠證據去推論兩個長照服務中心之每月平均服務人次不相等。

三、COVID-19疫情期間，學校關閉或改為線上課程。某教育機構評估疫情後學生在閱讀方面的

能力。隨機抽取600名八年級學生進行滿分100分之閱讀測驗，記錄其成績，得樣本平均數56分、樣本標準差18分，分數分布如下：

分數	[0, 20]	(20, 40]	(40, 60]	(60, 80]	(80, 100]
人數	54	144	252	120	30

請回答下列問題：（每小題10分，共20分）

(一)在0.05顯著水準之下，試檢定此資料是否服從常態分配。

(二)若[0, 40]分為「待加強」，(40, 60]分為「基礎」，(60, 100]分為「精熟」。已知疫情前，此三種等級之比例分別為30%，50%，20%。在顯著水準0.05之下，試檢定疫情前後八年級學生閱讀能力之等級分布是否相同。

試題評析	本題屬於卡方適合度檢定之計算題型，是常考範圍之一，同學獲得滿分不難，但是所需查表值考卷內幾乎沒有提供。
考點命中	《高點・高上統計學講義》第三回，趙治勳編撰，第十三章第二節，頁9。

答：

(一)

	(0,20)	(20,40)	(40,60)	(60,80)	(80,100)	
O_i	54	144	252	120	30	600
E_i	13.8	96.6	241.8	192.6	55.2	600

其中 在H_0為真下, $X \sim N(56,18^2)$ 平均數與變異數需以樣本估計($c=2$)

$$E_1 = nP(X \le 20) = nP(Z \le -2) = 600(0.023) = 13.8$$

$$E_2 = nP(20 < X \le 40) = nP(-2 < Z \le -0.89) = 600(0.184 - 0.023) = 96.6$$

$$E_3 = nP(40 < X \le 60) = nP(-0.89 < Z \le 0.22) = 600(0.587 - 0.184) = 241.8$$

$$E_4 = nP(60 < X \le 80) = nP(0.22 < Z \le 1.33) = 600(0.908 - 0.587) = 192.6$$

H_0:資料是服從常態分配 vs H_1: not H_0

T.S.：$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{d}{\sim} \chi^2_{(5-1-2=2)}$

R.R.：Reject H_0 at $\alpha = 0.05$ if $\chi^{2*} > \chi^2_{(2)0.05} = 5.991$

$\because \chi^{2*} = 179.66$ $\therefore$ reject H_0

我們有足夠證據去推論資料不是服從常態分配

(二)

	待加強	基礎	精熟	
O_i	198	252	150	600
E_i	180	300	120	600

其中 在H_0為真下, $X \sim$ 三項(n=600,$p_1 = 0.3, p_2 = 0.5, p_3 = 0.2$)

沒有參數估計($c = 0$)

$E_1 = np_1 = 600 \times 0.3 = 180$ ， $E_2 = np_2 = 600 \times 0.5 = 300$

$H_0: p_1 = 0.3, p_2 = 0.5, p_3 = 0.2$ vs H_1: not H_0

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

R.R.：Reject H_0 at $\alpha=0.05$ if $\chi^{2*}>\chi^2_{(2)0.05}=5.991$

$\because \chi^{2*}=16.98$ $\therefore$ reject H_0

我們有足夠證據去推論疫情前後閱讀能力之等級分佈不相同。

四、王先生蒐集過去12個月甲市新成屋的交易價格，得每坪平均交易價格$y_1, y_2,\ldots, y_{12}$（單位：萬元）。已知此樣本之平均數與標準差分別為$\bar{y}=63, s_y=12$。又將每坪平均交易價格對時間做線性迴歸，得到截距項係數估計值為50。請回答下列問題：（每小題10分，共40分）

(一)試求斜率係數估計值。

(二)試分別預測接下來兩個月的每坪平均交易價格。

(三)在0.05顯著水準之下，試檢定斜率係數是否為正值。

(四)接下來兩個月，若每坪平均交易價格的真實值分別為77萬元與72萬元，試計算(二)中預測結果的平均絕對誤差（mean absolute deviation）。

試題評析	本題屬於簡迴歸之計算題型，同學應該可以獲得滿分，其中(四)之平均絕對誤差在講義中已經有講過
考點命中	1.《高點・高上迴歸分析講義》，趙治勳編撰，第貳章。 2.《高點・高上統計學講義》第三回，趙治勳編撰，第十四章，頁41。

答：

假設模型：$Y_i=\beta_0+\beta_1X_i+\varepsilon_i,\quad \varepsilon_i\overset{iid}{\sim}N(0,\sigma^2)$

其中 $X_i=i, i=1,2,\cdots,12$表示第 i 個月，$\bar{X}=\frac{1+2+\cdots+12}{12}=6.5$，$S_X^2=13$

(一) $\hat{\beta}_0=50=\bar{Y}-\hat{\beta}_1\bar{X}=63-\hat{\beta}_1(6.5)\quad\Rightarrow\hat{\beta}_1=2$

(二) $\hat{y}=50+2x$

第 13 個月($x=13$)之每坪平均交易價格之預測值為 $\hat{y}=50+2(13)=76$

第 14 個月($x=14$)之每坪平均交易價格之預測值為 $\hat{y}=50+2(14)=78$

(三) $H_0:\beta_1\le 0$ vs $H_1:\beta_1>0$

T.S.：$T=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\sim t_{(12-2=10)}$

R.R.：Reject H_0 at $\alpha=0.05$ if $T^*>t_{(10)0.05}=1.812$

$\because T^*=\frac{0.6\sqrt{12-2}}{\sqrt{1-(0.6)^2}}=2.372$ $\therefore$ reject H_0

其中 $\hat{\beta}_1=r\frac{S_Y}{S_X}\Rightarrow r=\hat{\beta}_1\frac{S_X}{S_Y}=2\times\frac{\sqrt{13}}{12}=0.6$

我們有足夠證據去推論斜率係數為正值。

(四) $MAD=\frac{\sum_{i=1}^{n}|Y_i-\hat{Y}_i|}{n}=\frac{|76-77|+|72-78|}{2}=3.5$