《統計學概要》

一、假設隨機變數$X\sim N(\mu,\sigma^2)$

令$U=\mu+k\sigma$，且$L=\mu-k\sigma$

定義 $I=\dfrac{U-L}{6\sigma}$

（每小題5分，共15分）

(一)當I=1，k值及P(L<X<U)機率值為何？

(二)當I=0.5，k值及P(L<X<U+σ)機率值為何？

(三)當I=1.5，k值及P(L<X<U)機率值為何？

試題評析	本題是常態分配機率計算題，拿滿分不難。
考點命中	《高上統計學講義第二回》，趙治勳編撰，第六章第三節。

答：

$$I=\frac{U\text{-}L}{6\sigma}=\frac{\mu+k\sigma\text{-}\mu+k\sigma}{6\sigma}=\frac{k}{3}\Rightarrow k=3I$$

(一) $k=3(1)=3$, $P(L<X<U)=P(\mu-3\sigma<X<\mu+3\sigma)=P(-3<Z<3)=0.9974$

(二) $k=3(0.5)=1.5$, $P(L<X<U+\sigma)=P(\mu-1.5\sigma<X<\mu+1.5\sigma+\sigma)=P(-1.5<Z<2.5)=0.927$

(三) $k=3(1.5)=4.5$, $P(L<X<U)=P(\mu-4.5\sigma<X<\mu+4.5\sigma)=P(-4.5<Z<4.5)\approx 1$

二、抽取6位成年人作為樣本，且詢問他們每星期花在休閒活動之時間，他們之回應如下：

14,36,18,16,20,28（小時）

假設他們每星期休閒活動之時間是服從常態分配。試求：

(一)成年人每星期花在休閒活動之平均時間之點估計值為何？（5分）

(二)成年人每星期休閒活動之樣本標準差為何？（5分）

(三)檢定成年人每星期休閒活動時間的標準差是否超過5小時（α＝0.05）？。（7分）

(四)建立成年人每星期休閒活動平均時間之90%信賴區間。（8分）

試題評析	本題是單一母體平均數與標準差之假設檢定及信賴區間一般計算題，拿滿分不難。
考點命中	《高上統計學講義第三回》，趙治勳編撰，第八章第五節。

答：

(一) $\bar{x}=22$ (小時)

(二) $s=8.39$ (小時)

(三)令 X 表成年人每星期花在休閒活動之時間(小時)　[假設隨機樣本]

母體：$X\sim N(\mu,\sigma^2)$

樣本：$X_1,X_2,\cdots,X_6\overset{iid}{\sim}N(\mu,\sigma^2)$

點估計：$S^2=\dfrac{\sum(X_i-\bar{X})^2}{n\text{-}1}$

$H_0:\sigma\le 5$ vs $H_1:\sigma>5$　等價於 $H_0:\sigma^2\le 5^2$ vs $H_1:\sigma^2>5^2$

T.S.: $\chi^2=\dfrac{(6\text{-}1)S^2}{5^2}\sim\chi^2_{(6\text{-}1=5)}$

R.R.: Reject H_0 at $\alpha = 0.05$ if $\chi^{2*} > \chi^2_{0.05(5)} = 11.071$

$\because \chi^{2*} = \dfrac{(6\text{-}1)(8.39)^2}{5^2} = 14.078 \qquad \therefore reject H_0$

結論: 我們有足夠證據去推論成年人每星期休閒活動時間之標準差超過5小時。

(四)點估計: $\overline{X} \sim N(\mu, \dfrac{\sigma^2}{6})$

樞紐量: $T = \dfrac{\overline{X} - \mu}{S/\sqrt{6}} \sim t_{(6\text{-}1=5)}$

機率區間: $P(-t_{(5)0.05} \le \dfrac{\overline{X} - \mu}{S/\sqrt{6}} \le t_{(5)0.05}) = 0.9$

信賴區間: $P(\overline{X} - t_{(5)0.05} S/\sqrt{6} \le \mu \le \overline{X} + t_{(5)0.05} S/\sqrt{6}) = 0.9$

結論: μ 之90%信賴區間

$(\overline{X} - t_{(5)0.05} S/\sqrt{6}, \overline{X} + t_{(5)0.05} S/\sqrt{6}) = (15.098, 28.902)$

三、調查400位隨機抽取的智慧型手機使用者之手機製造廠牌資料，整理如下：

廠牌	A	B	C	D	總和
人數	280	40	20	60	400

欲知使用者是否有廠牌偏好，（每小題5分，共15分）

(一)寫出虛無假設（H_0）和對立假設（H_1）。

(二)檢定統計量在（H_0）為真下之分配為何？

(三)在α=0.01下，說明手機使用者是否有廠牌偏好？

試題評析	本題是卡方適合度檢定一般計算題，拿滿分不難。
考點命中	《高上統計學講義第四回》，趙治勳編撰，第十一章第二節。

答：

(一)令 p_i 分別表偏好於品牌 A,B,C,D 手機之比例，$i = 1,2,3,4$

$$H_0: p_1 = p_2 = p_3 = p_4 = \frac{1}{4} \quad \text{vs} \quad H_1: \text{至少一個 } p_i \ne \frac{1}{4}$$

(二)自由度爲 3 之卡方分配 $\chi^2_{(3)}$

(三) $H_0: p_1 = p_2 = p_3 = p_4 = \dfrac{1}{4}$ vs H_1: 至少一個 $p_i \ne \dfrac{1}{4}$

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

R.R.: Reject H_0 at $\alpha = 0.01$ if $\chi^{2*} > \chi^2_{(3)0.01} = 11.345$

$\because \chi^{2*} = 440 \qquad \therefore reject H_0$

結論：我們有足夠證據去推論手機使用者對廠牌有偏好的。

四、X和Y的可能值及對應的聯合機率如表所示：

Y \ X	0	1
1	0.2	0.1
2	0.3	0.2
3	0.1	0.1

(一)計算X=0之機率值。（2分）
(二)計算Y的期望值。（3分）

試題評析	本題是二元間斷型隨機變數一般計算題，拿滿分不難。
考點命中	《高上統計學講義第一回》，趙治勳編撰，第四章第三節。

答：

(一) $P(X=0)=0.2+0.3+0.1=0.6$

(二) $E(Y)=(1)(0.3)+(2)(0.5)+(3)(0.2)=1.9$

五、變異數分析表如下所列：

變異來源	平方和	自由度	均方	F值
處理	700	B	C	E
誤差	A	9	D	
總和	970	11		

試作：
(一)請寫出表中A～E的數字。（5分）
(二)因子的個數及其水準數為何？（4分）
(三)實驗的總次數為何？每個處理下的實驗次數為何？如何決定所有實驗的順序？（6分）
(四)在顯著水準0.1下，因子是否顯著？（5分）

試題評析	本題是變異數分析一般計算題，拿滿分不難。
考點命中	《高上統計學講義第四回》，趙治勳編撰，第十章第三節。

答：

(一)(A)270　(B)2　(C)350　(D)30　(E)11.667
(二)因子個數爲1，水準數爲3
(三)實驗總次數爲12，每個處理下實驗次數爲4，研究者可以把4張1號、4張2號及4張3號並12張號碼牌放入袋子中，以隨機的方式一次抽取一個出來以決定實驗順序。
(四)假設模型: $y_{ij}=\mu+\alpha_i+\varepsilon_{ij}$ ， $\varepsilon_{ij}\overset{iid}{\sim}N(0,\sigma^2)$

$H_0:\mu_1=\mu_2=\mu_3$ vs H_1:至少一個$\mu_i\neq\mu_j$

T.S.: $F=\dfrac{MSR}{MSE}\sim F_{(2,9)}$

R.R.: Reject H_0 at $\alpha=0.1$ if $F^*>F_{0.1(2,9)}=3.007$

$\because F^*$=11.667　　$\therefore rejectH_0$

結論：我們有足夠的統計證據去推論該因子爲影響應變數的重要因素

六、變數x, y的配對資料如表所示：

No.	x	Y
1	1	1
2	2	3
3	3	5
4	4	6
5	4	10

試作：

(一)畫出y和x之散佈圖。（5分）

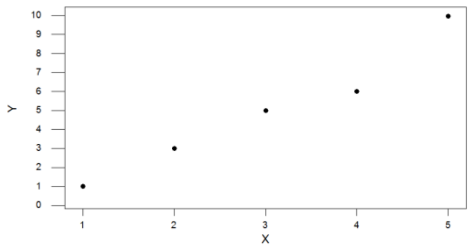
(二)依(一)的散佈圖，寫出y和x的迴歸模式及假設。（5分）

(三)就(二)的迴歸係數寫出其最小平方誤差法的估計量及估計值。（10分）

試題評析	本題是簡單線性迴歸分析一般計算題，拿滿分不難。
考點命中	《迴歸分析熱門題庫》，趙治勳編撰，第二篇第二章。

答：

(一)



(二)模型：$Y_i=\beta_0+\beta_1 X_i+\varepsilon_i,\ \ i=1,2,3,4,5$

假設：$\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$

(1)常態性假設

(2)變異數齊一性假設

(3)獨立性假設

(4) $E(\varepsilon_i)=0$

(5)模型之正確性

(三)估計量：

$$\hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{X}$$

$$\hat{\beta}_1=\frac{\sum X_iY_i-\frac{\sum X_i\sum Y_i}{n}}{\sum X_i^2-\frac{(\sum X_i)^2}{n}}=\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sum(X_i-\overline{X})^2}=\frac{SS_{XY}}{SS_X}$$

估計值：

$$\hat{\beta}_1=\frac{SS_{XY}}{SS_X}=2.1,\ \ \hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{X}=\text{-}1.3$$