《統計學(統計)》

參考值：

$z_{0.025}=1.96$，$z_{0.05}=1.645$，$z_{0.1}=1.28$，
$t_{0.025,8}=2.306$，$t_{0.025,9}=2.262$，$t_{0.025,10}=2.228$，$t_{0.05,8}=1.860$，$t_{0.05,9}=1.833$，$t_{0.05,10}=1.812$
$F_{0.025,2,8}=6.059$，$F_{0.025,4,8}=6.053$，$F_{0.05,2,8}=4.459$，$F_{0.05,4,8}=3.838$

一、令與為具獨立同分布、期望值1/的指數（exponential）隨機變數。令Y1 = X1 – X2以及 Y2= X2。（每小題 10 分，共 20 分）
(一)試求Y1與Y2之聯合機率密度函數。
(二)試求Y1之邊際機率密度函數。

試題評析	本題為隨機變數之變數變換，講義都有相關例題，獲得滿分不難
考點命中	《高點・高上統計學講義》第一回，趙治勳編撰，第五章P.158例19。

答：

(一) $$f_{X_1X_2}(x_1,x_2)\overset{X_1\perp X_2}{=}f_{X_1}(x_1)f_{X_2}(x_2)=\lambda e^{-\lambda x_1}\times\lambda e^{-\lambda x_2}=\lambda^2e^{-\lambda(x_1+x_2)}\qquad,0<x_1,x_2$$

$$\begin{matrix}Y_1=X_1-X_2\\Y_2=X_2\end{matrix}\Rightarrow\begin{matrix}X_1=Y_1+Y_2\\X_2=Y_2\end{matrix}\quad,\quad J=\begin{vmatrix}1&1\\0&1\end{vmatrix}=1$$

$$\begin{matrix}0<x_1<\infty\\0<x_2<\infty\end{matrix}\Rightarrow\begin{matrix}0<y_1+y_2<\infty\\0<y_2<\infty\end{matrix}\Rightarrow\begin{matrix}-y_2<y_1<\infty\\0<y_2<\infty\end{matrix}$$

$$f_{Y_1Y_2}(y_1,y_2)=f_{X_1X_2}(x_1=y_1+y_2,x_2=y_2)\times|1|=\lambda^2e^{-\lambda(y_1+2y_2)}\quad,\quad\begin{matrix}-y_2<y_1<\infty\\0<y_2<\infty\end{matrix}$$

(二) $$f_{Y_1}(y_1)=\int_{y_2}f_{Y_1Y_2}(y_1,y_2)dy_2=\begin{cases}\displaystyle\int_{-y_1}^{\infty}\lambda^2e^{-\lambda(y_1+2y_2)}dy_2=\frac{\lambda}{2}e^{\lambda y_1},-\infty<y_1<0\\ \displaystyle\int_{0}^{\infty}\lambda^2e^{-\lambda(y_1+2y_2)}dy_2=\frac{\lambda}{2}e^{-\lambda y_1},0<y_1<\infty\end{cases}$$

二、令$X_1,X_2,\dots,X_n$為一組隨機抽自常態分配 $N(0,\sigma^2)$之樣本。假設$\sigma_1^2>\sigma_0^2$，在顯著水準 0.05 下：
(一)試求檢定 $H_0:\sigma^2=\sigma_0^2$ vs. $H_1:\sigma^2=\sigma_1^2$ 的最強力檢定（most powerful test）。（10 分）
(二)試求檢定 $H_0:\sigma^2=\sigma_0^2$ vs. $H_1:\sigma^2>\sigma_0^2$ 的齊一最強力檢定（uniformly most powerful test）。（5分）
(三)說明檢定 $H_0:\sigma^2=\sigma_0^2$ vs. $H_1:\sigma^2\neq\sigma_0^2$ 的齊一最強力檢定是否存在。（5 分）

試題評析	本題是假設檢定原理，屬數統範圍之考題，但講義已經有完全相同之例題，只要同學有準備必能獲得高分。第(三)小題於講義中也有相關練習題教同學如何說明UMPT不存在。
考點命中	《迴歸分析申論題完全制霸》，高點文化出版，趙治勳編著，(一)P.5-16例12，(二)P.5.5例2，

	(三)P.5-9例4及P.5-13例9。

答：

母體: $X \sim N(0,\sigma^2)$

樣本: $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(0,\sigma^2)$

(一)

$H_0:\sigma^2=\sigma_0^2$ vs $H_1:\sigma^2=\sigma_1^2$

$$令\ \lambda=\frac{L(\sigma^2=\sigma_0^2)}{L(\sigma^2=\sigma_1^2)}=\frac{(\frac{1}{\sqrt{2\pi}\sigma_0})^n e^{-\frac{\sum x_i^2}{2\sigma_0^2}}}{(\frac{1}{\sqrt{2\pi}\sigma_1})^n e^{-\frac{\sum x_i^2}{2\sigma_1^2}}}=(\frac{\sigma_1}{\sigma_0})^n e^{-\frac{\sum x_i^2}{2}(\frac{1}{\sigma_o^2}-\frac{1}{\sigma_1^2})}\le k$$

$$\Rightarrow \sum x_i^2 \ge k'$$

$$\alpha=0.05=P(\sum_{i=1}^n X_i^2\ge c\,|\,\sigma^2=\sigma_0^2)\quad 其中\ \frac{\sum_{i=1}^n X_i^2}{\sigma_0^2}\overset{H_0 is true}{\sim}\chi^2_{(n)}$$

$$=P(\chi^2_{(n)}\ge\frac{c}{\sigma_0^2})$$

$$\Rightarrow \frac{c}{\sigma_0^2}=\chi^2_{(n)0.05}\Rightarrow c=\sigma_0^2\chi^2_{(n)0.05}$$

由Neyman-Pearson Lemma，

$\therefore C=\{\sum_{i=1}^n x_i^2\,|\,\sum_{i=1}^n x_i^2\ge\sigma_0^2\chi^2_{(n)0.05}\}$ 為在檢定水準 $\alpha=0.05$ 下 $H_0:\sigma^2=\sigma_0^2$ vs $H_1:\sigma^2=\sigma_1^2$ 之MPT

(二)

$H_0:\sigma^2=\sigma_0^2$ vs $H_1:\sigma^2>\sigma_0^2(=\sigma_1^2)$

$$令\ \lambda=\frac{L(\sigma^2=\sigma_0^2)}{L(\sigma^2=\sigma_1^2)}=\frac{(\frac{1}{\sqrt{2\pi}\sigma_0})^n e^{-\frac{\sum x_i^2}{2\sigma_0^2}}}{(\frac{1}{\sqrt{2\pi}\sigma_1})^n e^{-\frac{\sum x_i^2}{2\sigma_1^2}}}=(\frac{\sigma_1}{\sigma_0})^n e^{-\frac{\sum x_i^2}{2}(\frac{1}{\sigma_o^2}-\frac{1}{\sigma_1^2})}\le k$$

$$\Rightarrow \sum x_i^2 \ge k'$$

$$\alpha=0.05=P(\sum_{i=1}^n X_i^2\ge c\,|\,\sigma^2=\sigma_0^2)\quad 其中\ \frac{\sum_{i=1}^n X_i^2}{\sigma_0^2}\overset{H_0 is true}{\sim}\chi^2_{(n)}$$

$$=P(\chi^2_{(n)}\ge\frac{c}{\sigma_0^2})$$

$$\Rightarrow \frac{c}{\sigma_0^2}=\chi^2_{(n)0.05}\Rightarrow c=\sigma_0^2\chi^2_{(n)0.05}$$

由Neyman-Pearson Lemma，

$\therefore C=\{\sum_{i=1}^{n}x_i^2 \mid \sum_{i=1}^{n}x_i^2\ge\sigma_0^2\chi^2_{(n)0.05}\}$為在檢定水準$\alpha=0.05$下$H_0:\sigma^2=\sigma_0^2$ vs $H_1:\sigma^2>\sigma_0^2$之UMPT

(三)

$H_0:\sigma^2=\sigma_0^2$ vs $H_1:\sigma^2\ne\sigma_0^2(=\sigma_1^2)$

令$\lambda=\dfrac{L(\sigma^2=\sigma_0^2)}{L(\sigma^2=\sigma_1^2)}=\dfrac{(\frac{1}{\sqrt{2\pi}\sigma_0})^n e^{-\frac{\sum x_i^2}{2\sigma_0^2}}}{(\frac{1}{\sqrt{2\pi}\sigma_1})^n e^{-\frac{\sum x_i^2}{2\sigma_1^2}}}=(\dfrac{\sigma_1}{\sigma_0})^n e^{-\frac{\sum x_i^2}{2}(\frac{1}{\sigma_o^2}-\frac{1}{\sigma_1^2})}\le k$

$$\Rightarrow \sum x_i^2\le k' \quad or \quad \sum x_i^2\ge k''$$

$\alpha=0.05=P(\sum_{i=1}^{n}X_i^2\le c_1 \quad or \quad \sum_{i=1}^{n}X_i^2\ge c_2 \mid \sigma^2=\sigma_0^2)$ 其中$\dfrac{\sum_{i=1}^{n}X_i^2}{\sigma_0^2}\overset{H_0 is true}{\sim}\chi^2_{(n)}$

$$=P(\chi^2_{(n)}\le\frac{c_1}{\sigma_0^2} \quad or \quad \chi^2_{(n)}\ge\frac{c_2}{\sigma_0^2})$$

$$\Rightarrow\begin{cases}\dfrac{c_1}{\sigma_0^2}=\chi^2_{(n)0.975}\\[2ex] \dfrac{c_2}{\sigma_0^2}=\chi^2_{(n)0.025}\end{cases}\Rightarrow\begin{cases}c_1=\sigma_0^2\chi^2_{(n)0.975}\\ c_2=\sigma_0^2\chi^2_{(n)0.025}\end{cases}$$

$\therefore C=\{\sum_{i=1}^{n}x_i^2 \mid \sum_{i=1}^{n}x_i^2\le\sigma_0^2\chi^2_{(n)0.975} \quad or \quad \sum_{i=1}^{n}x_i^2\ge\sigma_0^2\chi^2_{(n)0.025}\}$為在檢定水準$\alpha=0.05$下$H_0:\sigma^2=\sigma_0^2$ vs $H_1:\sigma^2\ne\sigma_0^2$之LRT

$$power_1(\sigma^2)=P(\sum_{i=1}^{n}X_i^2\le\sigma_0^2\chi^2_{(n)0.975} \quad or \quad \sum_{i=1}^{n}X_i^2\ge\sigma_0^2\chi^2_{(n)0.025} \mid \sigma^2=\sigma_0^2)$$

$$power_2(\sigma^2)=P(\sum_{i=1}^{n}X_i^2\le\sigma_0^2\chi^2_{(n)0.95} \mid \sigma^2=\sigma_0^2)$$

$$power_3(\sigma^2)=P(\sum_{i=1}^{n}X_i^2\ge\sigma_0^2\chi^2_{(n)0.05} \mid \sigma^2=\sigma_0^2)$$

當$\sigma^2<\sigma_0^2$時， $power_2(\sigma^2)>power_1(\sigma^2)$

當$\sigma^2>\sigma_0^2$時， $power_3(\sigma^2)>power_1(\sigma^2)$

故$H_0:\sigma^2=\sigma_0^2$ vs $H_1:\sigma^2\ne\sigma_0^2$並不存在UMPT

三、以下是(X,Y)兩變數之觀測資料：

X	11	9	14	10	12	15	7	5	13	8	6
Y	7.46	6.77	12.74	7.11	7.81	8.84	6.08	5.39	8.15	6.42	5.73

以下考慮皮爾森相關係數（Pearson's correlation coefficient r）與皮爾曼等級相關係數（Spearman's rank correlation coefficient r_S）。

(一)試畫出(X,Y)之散布圖，並試計算 r 與 r_S。（10分）

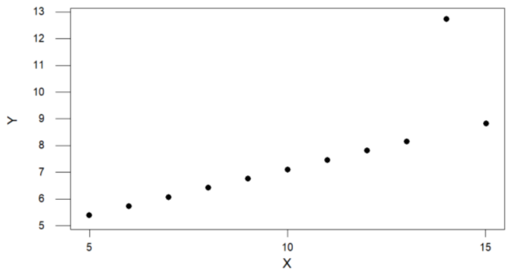
(二)試刪去本數據中之離群子後，重新計算 r 與 r_S。（5分）

(三)試問 r 與 r_S何者容易受離群子影響？（5分）

試題評析	本題為相關分析之計算題型，除了常考之Pearson相關係數外，還考到Spearman等級相關係數，只要同學有準備講義第13.6節，要獲得滿分不難
考點命中	《高點・高上統計學講義》第四回，趙治勳編撰，第十三章，P.30例18,19。

答：

(一)



$$\sum X_i = 110, \sum X_i^{\ 2} = 1210, \sum Y_i = 82.5, \sum Y_i^{\ 2} = 659.9762, \sum X_i Y_i = 879.97$$

$$SS_X = \sum X_i^2 - \frac{(\sum X_i)^2}{n} = 110 \qquad SS_{XY} = \sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n} = 54.97$$

$$SS_Y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = 41.2262$$

$$r = \frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}} = 0.8163$$

$rank_X$	7	5	10	6	8	11	3	1	9	4	2
$rank_Y$	7	5	11	6	8	10	3	1	9	4	2
d_i	0	0	-1	0	0	1	0	0	0	0	0

其中 $d_i = rank_X - rank_Y$

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6(2)}{(11)(11^2-1)} = 0.9909$$

(二) 刪除離群值$(X,Y) = (14, 12.74)$

$$\sum X_i = 96, \sum X_i^{\ 2} = 1014, \sum Y_i = 69.75, \sum Y_i^{\ 2} = 497.6686, \sum X_i Y_i = 701.61$$

$$SS_X = \sum X_i^2 - \frac{(\sum X_i)^2}{n} = 92.4 \qquad SS_{XY} = \sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n} = 32.01$$

$$SS_Y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = 11.16235$$

$$r=\frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}}=0.9967$$

$rank_X$	7	5	6	8	10	3	1	9	4	2
$rank_Y$	7	5	6	8	10	3	1	9	4	2
d_i	0	0	0	0	1	0	0	0	0	0

其中 $d_i=rank_X-rank_Y$

$$r_S=1-\frac{6\sum d_i^2}{n(n^2-1)}=1-\frac{6(1)}{(10)(10^2-1)}=0.9939$$

(三)

由於有無刪除離群值 r_S 變化較 r 為小，故 r 較容易受到離群值之影響

四、甲公司之零件製造部門有三台機器，輪流由五名員工（ABCDE）負責操作。李主任擬研究不同機器以及不同員工之生產量是否不同。以下是隨機抽取之生產量資料：

	機器一	機器二	機器三
員工 A	31	25	35
員工 B	33	26	33
員工 C	28	24	30
員工 D	30	29	28
員工 E	28	26	27

(一)試寫出 ANOVA 表（Analysis of Variance Table）。(5 分)

(二)在顯著水準 0.05 下，試檢定不同機器之生產量是否不同。(5 分)

(三)在顯著水準 0.05 下，試檢定不同員工之生產量是否不同。(5 分)

(四)試寫出模型假設。(5 分)

四、甲公司之零件製造部門有三台機器，輪流由五名員工（ABCDE）負責操作。李主任擬研究不同機器以及不同員工之生產量是否不同。以下是隨機抽取之生產量資料：

	機器一	機器二	機器三
員工 A	31	25	35
員工 B	33	26	33
員工 C	28	24	30
員工 D	30	29	28
員工 E	28	26	27

(一)試寫出 ANOVA 表（Analysis of Variance Table）。（5 分）

(二)在顯著水準 0.05 下，試檢定不同機器之生產量是否不同。（5 分）

(三)在顯著水準 0.05 下，試檢定不同員工之生產量是否不同。（5 分）

(四)試寫出模型假設。（5 分）

試題評析	本題為二因子變異數分析之計算題，屬基礎題型，只要考生計算上沒有失誤，可以輕鬆獲得滿分
考點命中	《高點・高上統計學講義》考題補充Ｑ１，趙治勳編撰，例298。

答：

(一)

ANOVA TABLE				
Source	SS	d.f.	MS	F
機器	62.533	2	31.2665	$F_1^* = 5.755$
員工	33.733	4	8.43325	$F_2^* = 1.552$
Error	43.467	8	5.433375	
Total	139.733	14		

(二)

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

T.S.: $F_1 = \dfrac{MSA}{MSE} \sim F_{(2,8)}$

R.R.: Reject H_0 at $\alpha = 0.05$ if $F_1^* > F_{0.05(2,8)} = 4.459$

$\because F_1^*$=5.755 $\quad \therefore rejectH_0$

結論:我們有足夠的統計證據去推論不同機器之生產量是不同的

(三)

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ vs H_1:至少一個$\mu_i \neq \mu_j, i \neq j$

T.S.: $F_2 = \dfrac{MSB}{MSE} \sim F_{(4,8)}$

R.R.: Reject H_0 at $\alpha = 0.05$ if $F_2^* > F_{0.05(4,8)} = 3.838$

$\because F_2^*$=1.552 $\quad \therefore don'trejectH_0$

結論:我們沒有足夠的統計證據去推論不同員工之生產量不同

(四)

模型: $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$ ，$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$， $i = 1,2,3$， $j = 1,2,3,4,5$

五、乙公司從2018年第一季至2020年第四季之銷售量如下表所示：

	第一季	第二季	第三季	第四季
2018	1,600	2,500	2,800	2,970
2019	2,100	3,100	3,650	3,350
2020	2,250	3,250	3,840	3,860

假設該公司近年第一至四季的季節指數分別為 74.720、103.978、123.761、97.540。

(一)試計算去除季節因子之銷售量。(5分)

(二)考慮簡單線性模型 $y_t = \alpha + \beta t + e_t$，試求出去除季節因子之銷售量的趨勢估計式，並在0.05顯著水準下檢定斜率是否為零。(10分)

(三)試預測2021年第一至四季之銷售量。(5分)

試題評析	本題為時間序列分析中有關季節因素之計算題，跟課本例題完全相同，考生獲得滿分應該不難
考點命中	《高點‧高上統計學講義》第四回，趙治勳編撰，第十四章，P.37例。

答：

(一) 注意:題目給的季節指數的和為400，表示單位是%

年	季	銷售量	季節指數(%)	去除季節因子之銷

				售量
2018	一	1600	74.72	2141.328
	二	2500	103.978	2404.355
	三	2800	123.761	2262.425
	四	2970	97.54	3044.905
2019	一	2200	74.72	2944.325
	二	3100	103.978	2981.400
	三	3650	123.761	2949.233
	四	3350	97.54	3434.488
2020	一	2250	74.72	3011.242
	二	3250	103.978	3125.661
	三	3840	123.761	3102.755
	四	3860	97.54	3957.351

(二)

年	季	去除季節因子之銷售量(Y)	t
2018	一	2141.328	1
	二	2404.355	2
	三	2262.425	3
	四	3044.905	4
2019	一	2944.325	5
	二	2981.400	6
	三	2949.233	7
	四	3434.488	8
2020	一	3011.242	9
	二	3125.661	10
	三	3102.755	11
	四	3957.351	12

$\sum t_i = 78, \sum t_i^{\,2} = 650$, $\sum Y_i = 35359.468, \sum Y_i^{\,2} = 106932751.3$, $\sum t_i Y_i = 246623.798$

$$SS_t = \sum t_i^2 - \frac{(\sum t_i)^2}{n} = 143 \qquad SS_{tY} = \sum t_i Y_i - \frac{(\sum t_i)(\sum Y_i)}{n} = 16787.256$$

$$SS_Y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = 2741753.217$$

$$\hat{\beta} = \frac{SS_{tY}}{SS_t} = 117.393 \qquad \hat{\alpha} = \overline{Y} - \hat{\beta}\overline{t} = 2183.568$$

$\hat{y} = 2183.568 + 117.393t$

$H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$

T. S. : $T = \dfrac{\hat{\beta} - 0}{\sqrt{\dfrac{MSE}{SS_t}}} \sim t_{(12-2=10)}$

R. R. : Reject H_0 at $\alpha = 0.05$ if $|T^*| > t_{0.025(10)} = 2.228$

$\because T^* = \dfrac{117.393 - 0}{\sqrt{\dfrac{77105.356}{143}}} = 5.056 \qquad \therefore reject H_0$

其中 $MSE = \dfrac{SSE}{n-2} = \dfrac{SS_Y - \hat{\beta}^2 SS_t}{12-2} = 77105.356$

結論:我們有足夠的統計證據去推論斜率不為零

(三)

2021年第一季($t = 13$)已消除季節影響銷售量之趨勢值為

$\hat{y}^* = 2183.568 + 117.393(13) = 3709.677$

因此，2021年第一季($t = 13$)銷售量之估計值為 $\hat{y} = \hat{y}^* \times S_1 = 3709.677 \times 74.72\% = 2771.871$

2021年第二季($t = 14$)已消除季節影響銷售量之趨勢值為

$\hat{y}^* = 2183.568 + 117.393(14) = 3827.07$

因此，2021年第二季($t = 14$)銷售量之估計值為 $\hat{y} = \hat{y}^* \times S_2 = 3827.07 \times 103.978\% = 3979.311$

2021年第三季($t = 15$)已消除季節影響銷售量之趨勢值為

$\hat{y}^* = 2183.568 + 117.393(15) = 3944.463$

因此，2021年第三季($t = 15$)銷售量之估計值為 $\hat{y} = \hat{y}^* \times S_3 = 3944.463 \times 123.761\% = 4881.707$

2021年第四季($t = 16$)已消除季節影響銷售量之趨勢值為

$\hat{y}^* = 2183.568 + 117.393(16) = 4061.856$

因此，2021年第四季($t = 16$)銷售量之估計值為 $\hat{y} = \hat{y}^* \times S_4 = 4061.856 \times 97.54\% = 3961.934$