《統計學》

一、設隨機變數X服從平均數 $\mu=0$ ，標準差為 σ 的常態分配；即 $X\sim N(0,\sigma^2)$ ，又 $X_1,X_2,\ldots,X_{n_1}$, $X_{n_1+1},X_{n_1+2},\ldots,X_{n_1+n_2}$ 為抽自X之一組大小為 n_1+n_2 之隨機樣本，令統計量：

$$S=\sum_{i=1}^{n_1}X_i/\sqrt{n_1}\sigma \text{ ， } T=\sum_{j=n_1+1}^{n_1+n_2}X_j^2/\sigma^2 \text{ ， } U=\frac{\sqrt{n_2}\sum_{i=1}^{n_1}X_i}{\sqrt{n_1}\sqrt{\sum_{j=n_1+1}^{n_1+n_2}X_j^2}} \text{ 及 } V=\frac{n_2\sum_{i=1}^{n_1}X_i^2}{n_1\sum_{j=1}^{n_1+n_2}X_j^2}$$

試問：（每小題5分，共20分）

(一)S之機率分配為何？

(二)T之機率分配為何？

(三)U之機率分配為何？

(四)V之機率分配為何？

試題評析	本題是抽樣分配的範圍內容，小心計算，獲得高分應該不難。
考點命中	《高點‧高上統計學講義》第二回，趙治勳編撰，第九章。

答：

(一) $\because \sum_{i=1}^{n_1}X_i \sim N(0,n_1\sigma^2)$

$$\therefore S=\frac{\sum_{i=1}^{n_1}X_i}{\sqrt{n_1}\sigma}=\frac{\sum_{i=1}^{n_1}X_i-0}{\sqrt{n_1\sigma^2}}\sim N(0,1)$$

(二) $\because \frac{(X_i-0)^2}{\sigma^2}=\frac{X_i^2}{\sigma^2}\overset{iid}{\sim}\chi^2_{(1)}$ ， $i=1,2,\cdots,n_1+n_2$

由 χ^2 加成性可得， $T=\frac{\sum_{j=n_1+1}^{n_1+n_2}X_j^2}{\sigma^2}=\frac{\sum_{i=1}^{n_1}X_i-0}{\sqrt{n_1\sigma^2}}\sim\chi^2_{(n_2)}$

(三) $$U=\frac{\sqrt{n_2}\sum_{i=1}^{n_1}X_i}{\sqrt{n_1}\sqrt{\sum_{j=n_1+1}^{n_1+n_2}X_j^2}}=\frac{\frac{\sum_{i=1}^{n_1}X_i}{\sqrt{n_1}\sigma}}{\sqrt{\frac{\sum_{j=n_1+1}^{n_1+n_2}X_j^2/\sigma^2}{n_2}}}\sim t_{(n_2)}$$

(四)題目可能有誤，原題目分母 $\sum_{j=1}^{n_1+n_2}X_j^2$ 應該是 $\sum_{j=n_1+1}^{n_1+n_2}X_j^2$

同(二)可知 $\dfrac{\sum_{i=1}^{n_1} X_i^2}{\sigma^2} \sim \chi^2_{(n_1)}$

$$\therefore V = \frac{n_2 \sum_{i=1}^{n_1} X_i^2}{n_1 \sum_{j=n_1+1}^{n_1+n_2} X_j^2} = \frac{\dfrac{\sum_{i=1}^{n_1} X_i^2}{\sigma^2} \Big/ n_1}{\dfrac{\sum_{j=n_1+1}^{n_1+n_2} X_j^2}{\sigma^2} \Big/ n_1} \sim F_{(n_1, n_2)}$$

二、設隨機變數X具有如下之機率分配f(x)：

x	1	2	3
f(x)	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中θ為未知參數，且0<θ<1。又已知抽自X之一組大小為n＝3的隨機樣本資料為2,1,3。請根據此樣本資料：（每小題12分，共24分）

(一)試以動差法（method of moments estimation）求θ之點估計值$\hat{\theta}$＝？

(二)試以最大概似法（method of maximum likelihood estimation）求θ之點估計值$\hat{\theta}$＝？

試題評析	本題是 MME 與 MLE 的範圍內容，獲得高分應該不難。
考點命中	《高點・高上統計學講義》第二回，趙治勳編撰，第十章。

答：

(一)令 $E(X) = 3 - 2\theta = \bar{X} \Rightarrow \hat{\theta}_{MME} = \dfrac{\bar{X} - 3}{-2}$

故點估計值為 $\hat{\theta}_{MME} = \dfrac{\bar{x} - 3}{-2} = \dfrac{2-3}{-2} = \dfrac{1}{2}$

(二)由隨機樣本結果為 $x_1 = 2, x_2 = 1, x_3 = 3$

得概似函數為

$$L(\theta) = f(x_1) f(x_2) f(x_3) = 2\theta(1-\theta) \times \theta^2 \times (1-\theta)^2 = 2\theta^3 (1-\theta)^3$$

$$\ln L(\theta) = \ln 2 + 3\ln\theta + 3\ln(1-\theta)$$

令 $\dfrac{d \ln L(\theta)}{d\theta} = \dfrac{3}{\theta} - \dfrac{3}{1-\theta} = 0$ 且 $\dfrac{d^2 \ln L(\theta)}{d\theta^2} < 0$

$$\Rightarrow \hat{\theta}_{MLE} = \frac{1}{2}$$

三、甲公司A廠為日產500個小型變壓器的工廠，該廠品管經理隨機抽出10天的生產資料，並記錄其不良品數如下表：

天	1	2	3	4	5	6	7	8	9	10
不良品數	15	18	13	22	12	8	16	15	16	25

根據此樣本資料，是否顯示該廠每日產品的不良品數中位數（M）為20？試取顯著水準 $\alpha = 0.05$，以符號檢定法檢定之。（12分）

試題評析	本題是符號檢定的範圍內容，獲得高分應該不難。
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。

答：

$H_0: M = 20$ vs $H_1: M \neq 20$

T.S.: $T \sim Bin(n = 10, p = \frac{1}{2})$

R.R.: Reject H_0 at $\alpha = 0.05$ if $\alpha > p - value$

$T_0 = 2 \quad n = 10$

$$\because p - value = 2\min\{\sum_{t=0}^{T_0}\binom{n}{t}(\frac{1}{2})^n, \sum_{t=T_0}^{n}\binom{n}{t}(\frac{1}{2})^n\} = 2\sum_{t=0}^{2}\binom{10}{t}(\frac{1}{2})^{10} = 0.1094$$

$\therefore don'trejectH_0$

結論：我們沒有足夠的統計證據去推論不良品數中位數不為 20。

四、某公司先後引進3部機器生產甲產品，公司管理部門想了解其生產效率之差異，找了4位操作員分別在這3部機器上操作5次，這4位操作員在3部機器上的操作順序完全隨機安排，今記綠操作員完成生產的工作時間，共得到60筆資料。若資料適合做變異數分析，且經電腦分析得到下列ANOVA表：

變異來源	平方和（SS）	自由度（df）	均方（MS）	F值	$F_{0.05}(v_1, v_2)$
操作員（A）	16.80	3	5.60	5.83	2.81
機器（B）	6.22	2	3.11	3.24	3.20
A×B交互作用	3.66	6	0.61	0.64	2.30
誤差	46.08	48	0.96		
總變異	72.76	59			

在給定顯著水率 $\alpha = 0.05$ 下，則：

(一)操作員和機器間有無交互作用？請說明依據。（6分）

(二)若不考慮交互作用，而將機器當做集區（Block），再做變異數分析。在此模式下，請寫出其ANOVA表，並試問操作員之平均操作時間是否有顯著差異？請說明依據。（10分）

(三)若不考慮機器間的差異，只討論操作員的單一因子的變異數分析。則在此模式下，請寫出其ANOVA表，並試問操作員之平均操作時間是否有顯著差異？請說明依據。（10分）

試題評析	本題是變異數分析的範圍內容，獲得高分應該不難。
考點命中	《高點‧高上統計學講義》第三回，趙治勳編撰，第十二章。

答：

(一)假設模型: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$

$i = 1,2,3,\ j = 1,2,3,4 \quad k = 1,2,3,4,5$

H_0: 交互作用不存在 vs H_1: 交互作用存在

T.S.: $F=\dfrac{MSR}{MSE}\sim F_{(6,48)}$

R.R.: Reject H_0 at $\alpha=0.05$ if $F^*>F_{0.05(6,48)}=2.3$

$\because F^*$=0.64 $\quad\therefore don'trejectH_0$

結論：我們沒有足夠的統計證據去推論操作員與機器交互作用存在。

(二)假設模型: $y_{ijk}=\mu+\alpha_i+\beta_j+\varepsilon_{ijk},\varepsilon_{ijk}\overset{iid}{\sim}N(0,\sigma^2)$

$i=1,2,3,\ j=1,2,3,4\quad k=1,2,3,4,5$

ANOVA TABLE				
source	SS	d.f.	MS	F
操作員	16.8	3	5.60	F_1^*=6.087
機器	6.22	2	3.11	F_2^*=3.38
Error	49.74	54	0.92	
Total	72.76	59		

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

T.S.：$F=\dfrac{MSR}{MSE}\sim F_{(3,54)}$

R.R.：Reject H_0 at $\alpha=0.05$ if $F^*>F_{0.05(3,54)}\approx 2.76$

$\because F^*$=6.087 $\quad\therefore rejectH_0$

結論：我們有足夠的統計證據去推論操作員之平均操作時間有顯著差異。

(三)假設模型: $y_{ij}=\mu+\alpha_i+\varepsilon_{ij},\varepsilon_{ijk}\overset{iid}{\sim}N(0,\sigma^2)$

$i=1,2,3,\ k=1,2,\cdots,15$

ANOVA TABLE				
source	SS	d.f.	MS	F
操作員	16.8	3	5.60	F^*=5.606
Error	55.96	56	0.999	
Total	72.76	59		

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

T.S.：$F=\dfrac{MSR}{MSE}\sim F_{(3,56)}$

R.R.：Reject H_0 at $\alpha=0.05$ if $F^*>F_{0.05(3,54)}\approx 2.76$

$\because F^*$=5.606 $\quad\therefore rejectH_0$

結論：我們有足夠的統計證據去推論操作員之平均操作時間有顯著差異。

五、桃園國際機場在106～108年各季之入境來臺外國旅客人數（單位：萬人）統計如下表

年＼季	1	2	3	4
106	238	242	243	289
107	262	242	251	297
108	270	282	267	289

交通部統計單位根據上述資料，以移動平均法求算得到桃園國際機場在106～108年各季之入境來臺外國旅客人數季節指數（單位：%）如下表：

季	1	2	3	4
季節指數	98.67	96.47	94.57	110.29

(一)試求經季節調整後之108年第4季在桃園國際機場入境來臺外國旅客人數 $y_{108.4}=?$（5分）

(二)試以最小平方法求經季節調整後，在桃園國際機場入境來臺外國旅客人數之長期趨勢方程式。（8分）

(三)請根據(二)之結果，試預測經季節調整後之109年第2季在桃園國際機場入境來臺外國旅客人數 $y_{109.2}=?$（5分）

試題評析	本題是時間序列的範圍內容，獲得高分應該不難。
考點命中	《高點・高上統計學講義》第三回，趙治勳編撰，第十四章。

答：

(一) $y_{108,4}=\dfrac{289}{1.1029}=262.036$ (萬人)

(二)

年	季	調整前人數	調整後人數	t
106	1	238	241.21	1
	2	242	250.86	2
	3	243	256.95	3
	4	289	262.04	4
107	1	262	265.53	5
	2	242	250.86	6
	3	251	265.41	7
	4	297	269.29	8
108	1	270	273.64	9
	2	282	292.32	10
	3	267	282.33	11
	4	289	262.04	12

$$\sum t_i=78,\sum t_i^{\ 2}=650,\sum Y_i=3172.48,\sum Y_i^{\ 2}=840902.783,\sum X_iY_i=21043.01$$

$$SS_t=\sum t_i^2-\frac{(\sum t_i)^2}{n}=143 \qquad SS_{tY}=\sum t_iY_i-\frac{(\sum t_i)(\sum Y_i)}{n}=421.89$$

$$\hat{\beta}_1=\frac{SS_{tY}}{SS_t}=\frac{421.89}{143}=2.95$$

$$\hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{t}=\frac{3172.48}{12}-2.95\times\frac{78}{12}=245.2$$

已除季節之長期趨勢方程式為 $\hat{Y}_t^*=245.2+2.95\times t$

(三) $y_{109,2}=245.2+2.95\times 14=286.5$