95% Bonferroni simult. intervals

$$m = \frac{pg(g-1)}{2} = 12 \qquad t_{87, \frac{0.05}{24}} = 2.94$$

$\Lambda^* = 0.8301$

$g = 3$

$$\left(\frac{90-4-2}{4}\right)\left(\frac{1-\sqrt{0.8301}}{\sqrt{0.8301}}\right) = 2.049$$

$F_{8,168}$

so

$P(F > 2.049) = 0.044$

Reject null hypoth., so $\exists$ time effect

① Max breath $\Rightarrow$

$(131.36 - 132.36) \pm \ldots$

$\tau_{11} - \tau_{21}$ $\quad -1 \pm 2.94 \sqrt{\frac{1784.4}{87}\left(\frac{1}{30}+\frac{1}{30}\right)}$ $\quad -1 \pm 3.44$

$\tau_{11} - \tau_{31}$ $\quad -3.1 \pm 3.44$

$\tau_{21} - \tau_{31}$ $\quad -2.1 \pm 3.44$

Bas Height ②

$\tau_{12} - \tau_{22}$ $\quad 0.9 \pm 2.94 \sqrt{\frac{1924.3}{87}\left(\frac{1}{30}-\frac{1}{30}\right)} = 0.9 \pm 3.57$

$\tau_{12} - \tau_{32}$ $\quad -0.2 \pm 3.57$

$\tau_{22} - \tau_{32}$ $\quad -1.1 \pm 3.57$

4) find eigenvalues of the correlation matrix

$$\rho e_i = \lambda_i e_i \rightarrow \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix} \begin{bmatrix} e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix} = \lambda_i \begin{bmatrix} e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\lambda_1 = 1 + 2\rho$$

$$\lambda_2 = \lambda_3 = 1 - \rho$$

consistent w/ (8-16)/(8-17)

$$|\rho - \lambda I| = 0$$

$$\begin{vmatrix} 1-\lambda & \rho & \rho \\ \rho & 1-\lambda & \rho \\ \rho & \rho & 1-\lambda \end{vmatrix} = 0 \rightarrow \begin{vmatrix} 1-\lambda+2\rho & \rho & \rho \\ 1-\lambda+2\rho & 1-\lambda & \rho \\ 1-\lambda+2\rho & \rho & 1-\lambda \end{vmatrix} = 0$$

$$(1-\lambda+2\rho)(1-\lambda-\rho)^2 = 0$$

$$\lambda_1 = 1 + 2\rho \qquad \lambda_2 = \lambda_3 = 1 - \rho$$

$$\rho = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \cdots & & 1 \end{bmatrix}$$

eigenvalue $\lambda_1 \ldots \lambda_p$

eigenvector $e_i = \begin{bmatrix} e_{1i} \\ \vdots \\ e_{pi} \end{bmatrix}$

$$\rho e_i = \lambda_i e_i \rightarrow \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & & \ddots & \rho \\ \vdots & & & \\ \rho & \cdots & & 1 \end{bmatrix} \begin{bmatrix} e_{1i} \\ \\ e_{pi} \end{bmatrix} = \lambda_i \begin{bmatrix} e_{1i} \\ \vdots \\ e_{pi} \end{bmatrix}$$

$$e_{1i} + \rho\left(e_{2i} + \cdots + e_{pi}\right) = \lambda_i e_{1i}$$
$$\vdots \qquad\qquad \vdots$$
$$e_{pi} + \rho\left(e_{1i} + \cdots + e_{(p-1)i}\right) = \lambda_i e_{pi}$$

$$\lambda_i = 1 + (p-1)\rho$$
OR
$$\underbrace{e_{1i} + \cdots + e_{pi} = 0}$$
$$\downarrow$$
$$\lambda_i = 1-\rho \quad (i = 2, 3, \dots p)$$

$$\rightarrow \rho(p-1)\left(e_{1i} + \cdots + e_{pi}\right) = (\lambda_i - 1)\left(e_{1i} + \cdots + e_{pi}\right)$$

$$\lambda_i = 1 + (p-1)\rho$$
$$\rightarrow \lambda_2 = \dots = \lambda_p = 1 - \rho$$
$$\lambda_1 = 1 + (p-1)\rho \qquad (8.16)$$

$$\left(e_{1i} + \rho\left(e_{2i} + \cdots e_{pi}\right)\right) = \lambda_1 e_{1i}$$
$$\downarrow$$
$$e_{1i} = (\lambda_1 + \rho)\, e_{1i})$$

$$e_{1i} + \rho\left(e_{2i} + \cdots + e_{pi}\right) = (1 + (p-1)\rho)\, e_{1i}$$
$$\vdots \qquad\qquad \vdots$$
$$e_{pi} + \rho\left(e_{1i} \quad + e_{(p-1)i}\right) = (1 + (p-1)\rho)\, e_{pi}$$

$$\rightarrow e_{1i} + \cdots + e_{pi} = p e_{1i} = p e_{2i} = \cdots = e_{pi}$$
$$e_{1i} = \cdots = e_{pi} \rightarrow e_1' = \left[\frac{1}{\sqrt{p}}, \dots \frac{1}{\sqrt{p}}\right] \qquad (8.17)$$

b) <u>verify</u> $|\rho - \lambda I| = 0 \rightarrow \begin{vmatrix} 1-\lambda & & \rho \\ & \ddots & \\ \rho & & 1-\lambda \end{vmatrix} = 0 \rightarrow \begin{vmatrix} 1-\lambda+(p-1)\rho & & \rho \\ \vdots & 1-\lambda & \\ 1-\lambda+(p-1)\rho & & 1-\lambda \end{vmatrix} = 0$

$$\rightarrow \left(1-\lambda + (p-1)\rho\right)\left(1-\lambda-\rho\right)^{p-1} = 0$$
$$\lambda_1 = 1 + (p-1)\rho$$
$$\lambda_2 = \dots = \lambda_p = 1 - \rho \qquad (8.16)$$

$$\rho e = \lambda_1 e_1 \rightarrow \begin{pmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_p \end{pmatrix} = (1 + (p-1)\rho) \begin{pmatrix} e_1 \\ \vdots \\ e_p \end{pmatrix} \rightarrow \begin{array}{l} e_1 = e_2 = \cdots e_p \\ e_1' e_1 = 1 \\ \rightarrow e_1' = \left[\frac{1}{\sqrt{p}} \cdots \frac{1}{\sqrt{p}}\right] \end{array}$$

8-5.

(a). Find the eigenvalues of the correlation matrix

$$\rho = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

$$|\rho - \lambda I| = \begin{vmatrix} 1-\lambda & \rho & \rho \\ \rho & 1-\lambda & \rho \\ \rho & \rho & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda+2\rho & \rho & \rho \\ 1-\lambda+2\rho & 1-\lambda & \rho \\ 1-\lambda+2\rho & \rho & 1-\lambda \end{vmatrix}$$

$$= (1-\lambda+2\rho)\begin{vmatrix} 1 & \rho & \rho \\ 1 & 1-\lambda & \rho \\ 1 & \rho & 1-\lambda \end{vmatrix}$$

$$= (1-\lambda+2\rho)\begin{vmatrix} 1 & \rho & \rho \\ 0 & 1-\lambda-\rho & 0 \\ 0 & 0 & 1-\lambda-\rho \end{vmatrix}$$

$$= (1-\lambda+2\rho)(1-\lambda-\rho)^2 = 0$$

eigenvalues. $\lambda_1 = 1+2\rho \qquad \lambda_2 = \lambda_3 = 1-\rho$. , Consistent with (8-16), (8-17)

(8-16) $\lambda_1 = 1+(p-1)\rho = 1+2\rho$ where $p=3$.

(8-17) $\lambda_2 = \lambda_3 = 1-\rho$.

(b) Verify the eigenvalue-eigenvector pairs for the $p \times p$ matrix ρ given in (8-15)

$$\rho_{p\times p}\, e_j = \lambda_j \cdot e_j$$

for $j=1$, $LS = \begin{bmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{bmatrix}_{p\times p} \cdot \begin{bmatrix} \frac{1}{\sqrt{p}} \\ \vdots \\ \frac{1}{\sqrt{p}} \end{bmatrix}_{p\times 1} = \begin{bmatrix} \frac{1}{\sqrt{p}} + (p-1)\cdot\sqrt{p} \\ \vdots \\ \vdots \end{bmatrix} = \frac{1}{\sqrt{p}} + (p-1)\sqrt{p}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

$RS = 1+(p-1)\rho \begin{bmatrix} \frac{1}{\sqrt{p}} \\ \vdots \\ \frac{1}{\sqrt{p}} \end{bmatrix} = \frac{1}{\sqrt{p}}(1+(p-1)\rho)\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{\sqrt{p}} + (p-1)\sqrt{p}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

$\Big\}$ $LS = RS$.

for $j=2$,

$LS = \begin{bmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} - \frac{\rho}{\sqrt{2}} \\ \frac{\rho}{\sqrt{2}} - \frac{1}{\sqrt{2}} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$RS = (1-\rho)\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$\Big\}$ $LS = RS$.

....

for $j=i$. $LS = \begin{bmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{(i-1)i}} \\ \vdots \\ \frac{1}{\sqrt{(i-1)i}} \\ \frac{-(i-1)}{\sqrt{(i-1)i}} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{(i-1)i}} - \frac{\rho}{\sqrt{(i-1)i}} \\ \vdots \\ \frac{\rho(i-1)}{\sqrt{(i-1)i}} - \frac{i-1}{\sqrt{(i-1)i}} \\ 0 \end{bmatrix}$

$\Big\}$ $LS = RS$

$RS = (1-\rho)\begin{bmatrix} \frac{1}{\sqrt{(i-1)i}} \\ \frac{1}{\sqrt{(i-1)i}} \\ \vdots \end{bmatrix}$

For $j=p$

$$LS = \begin{bmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{(p-1)p}} \\ \vdots \\ \frac{-(p-1)}{\sqrt{(p-1)p}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{(p-1)p}} - \frac{\rho}{\sqrt{(p-1)p}} \\ \vdots \\ \frac{(p-1)\rho}{\sqrt{(p-1)p}} - \frac{(p-1)}{\sqrt{(p-1)p}} \end{bmatrix}$$

$$RS = (1-\rho) \begin{bmatrix} \frac{1}{\sqrt{(p-1)p}} \\ \vdots \\ \frac{-(p-1)}{\sqrt{(p-1)p}} \end{bmatrix}$$

$\Bigg\}\ LS = RS$

Moreover, $e_i \perp e_j \quad i \neq j$.

$|\rho-\lambda I|=0$

$\Rightarrow \begin{vmatrix} 1-\lambda & \rho & \rho \\ \rho & 1-\lambda & \rho \\ \rho & \rho & 1-\lambda \end{vmatrix}=0$

$\Rightarrow \begin{vmatrix} 1-\lambda+2\rho & \rho & \rho \\ 1-\lambda+2\rho & 1-\lambda & \rho \\ 1-\lambda+2\rho & \rho & 1-\lambda \end{vmatrix}=0$

$\Rightarrow (1-\lambda+2\rho)(1-\lambda-\rho)^2=0$

$\therefore \lambda_1=1+2\rho,\ \lambda_2=\lambda_3=1-\rho$

$\rho=\begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$ eigen value $\lambda_1, \lambda_2, \lambda_3$, eigen vector $e_i=\begin{pmatrix} e_{1i} \\ e_{2i} \\ e_{3i} \end{pmatrix}$

$\rho e_i=\lambda_i e_i \Rightarrow \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}\begin{bmatrix} e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}=\lambda_i\begin{bmatrix} e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix} \Rightarrow \lambda_1=1+2\rho,\ \lambda_2=\lambda_3=1-\rho$

Yes, this results are consistent with 8-16 and 8.17.

(b). $\rho=\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & & \vdots \\ \vdots & & \ddots & \rho \\ \rho & \rho & \cdots & 1 \end{bmatrix}$ eigen value $\lambda_1, \cdots \lambda_p$, eigen vector $e_i=\begin{bmatrix} e_{1i} \\ \vdots \\ e_{pi} \end{bmatrix}$

$\rho e_i=\lambda_i e_i \Rightarrow \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \rho & \cdots & \cdots & 1 \end{bmatrix}\begin{bmatrix} e_{1i} \\ \vdots \\ e_{pi} \end{bmatrix}=\lambda_i\begin{bmatrix} e_{1i} \\ \vdots \\ e_{pi} \end{bmatrix}$

$\Rightarrow \begin{matrix} e_{1i}+\rho(e_{2i}+\cdots+e_{pi})=\lambda_i e_{1i} \\ e_{pi}+\rho(e_{1i}+\cdots+e_{(p-1)i})=\lambda_i e_{pi} \end{matrix}$

$\Rightarrow \rho(p-1)(e_{1i}+\cdots+e_{pi})=(\lambda_i-1)(e_{1i}+\cdots+e_{pi}) \Rightarrow \begin{cases} \lambda_i=1+(p-1)\rho \\ \text{or} \\ e_{1i}+\cdots+e_{pi}=0 \end{cases}$

$\Downarrow$

$\lambda_i=1-\rho \quad (i=2,\cdots p)$

$(e_{1i}+\rho(e_{2i}+\cdots+e_{pi})=\lambda_i e_{1i}$

$\Downarrow$

$e_{1i}=(\lambda_i+\rho)e_{1i})$

$\Rightarrow \lambda_i=1+(p-1)\rho$ or $\lambda_2=\cdots=\lambda_p=1-\rho$

$\therefore \lambda_1=1+(p-1)\rho \quad$ (8.16)

$\begin{cases} e_{1i}+\rho(e_{2i}+\cdots+e_{pi})=(1+(p-1)\rho)e_{1i} \\ \vdots \\ e_{pi}+\rho(e_{1i}+\cdots+e_{(p-1)i})=(1+(p-1)\rho)e_{pi} \end{cases}$

$\Rightarrow e_{1i}+\cdots+e_{pi}=pe_{1i}=pe_{2i}=\cdots=e_{pi}$

$\Rightarrow e_{1i}=\cdots=e_{pi} \quad \Rightarrow e_1'=[\frac{1}{\sqrt{p}},\cdots,\frac{1}{\sqrt{p}}]$ (8.17)

(b) verify: $|\rho-\lambda I|=0 \Rightarrow \begin{vmatrix} 1-\lambda & & \rho \\ & \ddots & \\ \rho & & 1-\lambda \end{vmatrix}=0 \Rightarrow \begin{vmatrix} 1-\lambda+(p-1)\rho & & \rho \\ \vdots & 1-\lambda & \\ 1-\lambda+(p-1)\rho & & 1-\lambda \end{vmatrix}=0$

$\Rightarrow (1-\lambda+(p-1)\rho)(1-\lambda-\rho)^{p-1}=0$

$\Rightarrow \lambda_1=1+(p-1)\rho,\ \lambda_2=\cdots=\lambda_p=1-\rho \qquad$ (8.16)

$\rho e=\lambda_1 e_1 \Rightarrow \begin{pmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{pmatrix}\begin{pmatrix} e_1 \\ \vdots \\ e_p \end{pmatrix}=(1+(p-1)\rho)\begin{pmatrix} e_1 \\ \vdots \\ e_p \end{pmatrix}$

$\Rightarrow e_1=e_2=\cdots=e_p$

$\cancel{e_1'e_1=1}$ $e_1'e_1=1 \Rightarrow e_1'=[\frac{1}{\sqrt{p}},\cdots,\frac{1}{\sqrt{p}}]$ (8.17)

$P(A) - P(A \cap A \quad) \leq \sum P(\cap A) =$

$$\bar{x} = \begin{bmatrix} 2 \\ 14 \end{bmatrix} = \begin{bmatrix} \frac{-3+4+8}{3} \\ \frac{15+16+11}{3} \end{bmatrix}$$

$$S = \begin{bmatrix} 19 & -5 \\ -5 & 7 \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} \frac{7}{108} & \frac{5}{108} \\ \frac{5}{108} & \frac{19}{108} \end{bmatrix}$$

$$\mu_0' = [4, 14]$$

$$T^2 = n(x - \mu_0)^T S^{-1} (\bar{x} - \mu_0)$$

$$= 3 \cdot [-2 \ 0] \begin{bmatrix} S^{-1} \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$= \frac{7}{9}$$

STA 5707 HW#9. Li Fan

Exercise 6.18.

(1) $\bar{X}_1 = \begin{bmatrix} 136.04 \\ 102.58 \\ 52.04 \end{bmatrix}$ $\quad S_1 = \begin{bmatrix} 451.52 & 270.97 & 165.95 \\ 270.97 & 171.73 & 101.84 \\ 165.95 & 101.84 & 64.74 \end{bmatrix}$

$\bar{X}_2 = \begin{bmatrix} 113.38 \\ 88.29 \\ 40.71 \end{bmatrix}$ $\quad S_2 = \begin{bmatrix} 138.77 & 79.15 & 37.38 \\ 79.15 & 50.04 & 21.65 \\ 37.38 & 21.65 & 11.26 \end{bmatrix}$

$S_{pooled} = \frac{n_1-1}{n_1+n_2-2} S_1 + \frac{n_2-1}{n_1+n_2-2} S_2 = \begin{bmatrix} 295.14 & 175.06 & 101.66 \\ 175.06 & 110.89 & 61.75 \\ 101.66 & 61.75 & 38.00 \end{bmatrix}$ Where $n_1 = n_2 = 24$

and $c^2 = \frac{(n_1+n_2-2)\cdot p}{n_1+n_2-p-1} \cdot F_{p, n_1+n_2-p-1}(\alpha) = \frac{46\cdot 3}{44} \cdot F_{3,44}(0.05)$ $\quad p=3$

$= 31.36 \cdot 2.816$

under $H_0: \mu_1 = \mu_2$. $\quad = 8.832$

$T^2 = [\bar{X}_1 - \bar{X}_2 - 0]' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{pooled}\right]^{-1} [\bar{X}_1 - \bar{X}_2 - 0]$

$= [22.66, 14.29, 11.33] \begin{bmatrix} 0.86 & -0.80 & -0.996 \\ -0.80 & 1.89 & -0.92 \\ -0.996 & -0.92 & 4.475 \end{bmatrix} \begin{bmatrix} 22.66 \\ 14.29 \\ 11.33 \end{bmatrix}$

$= 72.33 \quad \sim \frac{(n_1+n_2-2)\cdot p}{n_1+n_2-p-1} \cdot F_{p, n_1+n_2-p-1}(\alpha)$ where $n_1 = n_2 = 24$, $p=3$, $\alpha = 0.05$

Since $T^2 = 72.33 > 8.832$
reject H_0, conclude that the two population mean vectors are not equal at level α

(2) The most critical linear combination leading to the reject of H_0 has coefficient vector. $\hat{a} \propto \left(\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2\right)^{-1} (\bar{X}_1 - \bar{X}_2) = \begin{bmatrix} -3.27 \\ -1.63 \\ 14.99 \end{bmatrix}$

(3) 95% Simultaneous CI for the population mean differences are.

$\mu_1 - \mu_2 = \begin{bmatrix} \mu_{11} - \mu_{21} \\ \mu_{12} - \mu_{22} \\ \mu_{13} - \mu_{23} \end{bmatrix}$

$\mu_{11} - \mu_{21}: \quad (\bar{X}_{11} - \bar{X}_{21}) \pm \sqrt{c^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{11,pooled}} = 22.66 \pm \sqrt{8.832} \cdot \sqrt{\left(\frac{1}{24} + \frac{1}{24}\right) \cdot 295.14}$
$= 22.66 \pm 14.74$

$\mu_{12} - \mu_{22}: \quad 14.29 \pm \sqrt{8.832} \cdot \sqrt{\left(\frac{1}{24} + \frac{1}{24}\right) \cdot 110.89} = 14.29 \pm 9.03$

$\mu_{13} - \mu_{23}: \quad 11.33 \pm \sqrt{8.832} \cdot \sqrt{\left(\frac{1}{24} + \frac{1}{24}\right) \cdot 38} = 11.33 \pm 5.29$

95% Bonferroni Simultaneous CI for the population mean differences are.

$\mu_{1i}-\mu_{2i}$: $(\bar{X}_{1i}-\bar{X}_{2i}) \pm t_{n_1+n_2-2}\left(\frac{\alpha}{2p}\right)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{ii,pooled}}$

$\mu_{11}-\mu_{21}$: $22.66 \pm t_{46}(0.0083)\cdot\sqrt{\left(\frac{1}{24}+\frac{1}{24}\right)\cdot 29514} = 22.66 \pm 2.486\cdot\sqrt{\left(\frac{1}{24}+\frac{1}{24}\right)\cdot 29514}$

$= 22.66 \pm 12.33$

$\mu_{12}-\mu_{22}$: $14.29 \pm 2.486\cdot\sqrt{\left(\frac{1}{24}+\frac{1}{24}\right)110.89} = 14.29 \pm 7.56$

$\mu_{13}-\mu_{23}$: $11.33 \pm 2.486\cdot\sqrt{\left(\frac{1}{24}+\frac{1}{24}\right)38} = 11.33 \pm 4.42$

STA 5707 HW#9 Li Fan

Exercise 6_18

```
title1 'EXR6_18_T6_9';
Data turtles;
 turtle+1;
 infile 'E:\sta5707\data\T6-9.dat';
 Input x1 x2 x3 gender $;
 Lx1=Log(x1);
 Lx2=Log(x2);
 Lx3=Log(x3);
run;

data turtles_female;
 set turtles;
 where gender='female';
run;
 proc corr data=turtles_female cov;
  var x1 x2 x3;
run;

 data turtles_male;
  set turtles;
  where gender='male';
run;

proc corr data=turtles_male cov;
  var x1 x2 x3;
run;

proc iml;
 use turtles_female;
 read all var{x1 x2 x3} into x_f;
 close turtles_female;
 print x_f;
 x_fbar={136.04,102.58,52.04};
 e=J(24,1);
 s1=(x_f-e*x_fbar`)`*(x_f-e*x_fbar`)/23;
 use turtles_male;
 read all var{x1 x2 x3} into x_m;
 close turtles_male;
 print x_m;
 x_mbar={113.38,88.29,40.71};
 e=J(24,1);
 s2=(x_m-e*x_mbar`)`*(x_m-e*x_mbar`)/23;
 print s1 s2;
 spool=(24-1)/(24+24-2)*(s1+s2);
 print spool;
 xbardiff=x_fbar-x_mbar;
 print xbardiff;
 s=(1/24+1/24)*spool;
 s_1=inv(s);
 print s_1;
 Tsquare=xbardiff`*s_1*xbardiff;
 print Tsquare;
 ahat=inv(1/24*s1+1/24*s2)*xbardiff;
 print ahat;
```

```
run;
quit;
```

output:

```
        S1                                  S2

451.51993 270.97464 165.95471 138.76633  79.14673 37.375009
270.97464  171.7319 101.84421  79.14673  50.04167 21.653983
165.95471 101.84421 64.737322 37.375009 21.653983 11.259061


                         SPOOL

               295.14313 175.06069 101.66486
               175.06069 110.88678 61.749096
               101.66486 61.749096 37.998191


                         XBARDIFF

                            22.66
                            14.29
                            11.33


                            S_1

               0.8612592 -0.804773 -0.996516
               -0.804773 1.8903822 -0.918787
               -0.996516 -0.918787 4.4750796


                          TSQUARE

                         72.330743


                            AHAT

                         -3.274595
                         -1.632455
                         14.992132
```

rcise 6.18 (after log-transformation)

(1) $\bar{X}_1 = \begin{bmatrix} 4.90 \\ 4.62 \\ 3.94 \end{bmatrix}$ $S_1 = \begin{bmatrix} 0.026 & 0.020 & 0.025 \\ 0.020 & 0.016 & 0.019 \\ 0.025 & 0.019 & 0.025 \end{bmatrix}$

$\bar{X}_2 = \begin{bmatrix} 4.73 \\ 4.48 \\ 3.70 \end{bmatrix}$ $S_2 = \begin{bmatrix} 0.011 & 0.008 & 0.008 \\ 0.008 & 0.006 & 0.006 \\ 0.008 & 0.006 & 0.0068 \end{bmatrix}$

$$S_{pooled} = \frac{n_1-1}{n_1+n_2-2} S_1 + \frac{n_2-1}{n_1+n_2-2} S_2 = \begin{bmatrix} 0.0187 & 0.014 & 0.0165 \\ 0.014 & 0.0113 & 0.0127 \\ 0.0165 & 0.0127 & 0.0159 \end{bmatrix}$$

and $$c^2 = \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} \cdot F_{p,n_1+n_2-p-1}(\alpha) = \frac{46 \cdot 3}{44} \cdot F_{3,44}(0.05) = 8.832$$

under $H_0: \mu_1 = \mu_2$

$$T^2 = [\bar{X}_1 - \bar{X}_2 - 0]' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{pooled} \right]^{-1} [\bar{X}_1 - \bar{X}_2 - 0]$$

$$= [0.17, 0.14, 0.24] \begin{bmatrix} 13095.746 & -9602.33 & -5953.40 \\ -9602.33 & 17711.484 & -4185.98 \\ -5953.40 & -4185.98 & 10315.81 \end{bmatrix} \begin{bmatrix} 0.17 \\ 0.14 \\ 0.24 \end{bmatrix}$$

$$= 95.64 \sim \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} \cdot F_{p,n_1+n_2-p-1}(\alpha)$$ where $n_1 = n_2 = 24$ $p=3$, $\alpha = 0.0$

Since $T^2 = 95.64 > 8.832$

reject H_0. Conclud. that the two population mean vectors are not equal. at level 0.05.

(2) The most critical linear Combination leading to the reject of H_0 has Coefficient vector. $\hat{a} \propto \left(\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2\right)^{-1} (\bar{X}_1 - \bar{X}_2) = \begin{bmatrix} -546.87 \\ -157.42 \\ 877.68 \end{bmatrix}$

(3) 95% Simultaneous CI for the population mean difference after log transformation are:

$\mu_{11} - \mu_{21}$: $0.17 \pm \sqrt{8.832} \cdot \sqrt{\left(\frac{1}{24} + \frac{1}{24}\right) \cdot 0.0187} = 0.17 \pm 0.117$

$\mu_{12} - \mu_{22}$: $0.14 \pm \sqrt{8.832} \cdot \sqrt{\left(\frac{1}{24} + \frac{1}{24}\right) \cdot 0.0113} = 0.14 \pm 0.091$

$\mu_{13} - \mu_{23}$: $0.24 \pm \sqrt{8.832} \cdot \sqrt{\left(\frac{1}{24} + \frac{1}{24}\right) \cdot 0.0159} = 0.24 \pm 0.108$

95% Bonferroni Simultaneous CI for the population mean difference, after log-transformation are.

$$\mu_{11}-\mu_{21} : 0.17 \pm 2.486 \cdot \sqrt{\left(\frac{1}{24}+\frac{1}{24}\right)\cdot 0.0187} = 0.17 \pm 0.098$$

$$\mu_{12}-\mu_{22} : 0.14 \pm 2.486 \cdot \sqrt{\left(\frac{1}{24}+\frac{1}{24}\right)\cdot 0.0113} = 0.14 \pm 0.076$$

$$\mu_{13}-\mu_{23} : 0.24 \pm 2.486 \cdot \sqrt{\left(\frac{1}{24}+\frac{1}{24}\right)\cdot 0.0159} = 0.24 \pm 0.09$$

Exercise6_18 **(after log transformation)**

```
title1 'EXR6_18_T6_9_after log transformation';
Data turtles;
 turtle+1;
 infile 'E:\sta5707\data\T6-9.dat';
 Input x1 x2 x3 gender $;
 Lx1=Log(x1);
 Lx2=Log(x2);
 Lx3=Log(x3);
run;

data turtles_female;
 set turtles;
 where gender='female';
run;
 proc corr data=turtles_female cov;
  var Lx1 Lx2 Lx3;
run;

 data turtles_male;
  set turtles;
  where gender='male';
run;

proc corr data=turtles_male cov;
  var Lx1 Lx2 Lx3;
run;

/*exercise 6_18(1)*/
proc iml;
 use turtles_female;
 read all var{Lx1 Lx2 Lx3} into x_f;
 close turtles_female;
 print x_f;
 x_fbar={4.90,4.62,3.94};
 e=J(24,1);
 s1=(x_f-e*x_fbar`)`*(x_f-e*x_fbar`)/23;
 use turtles_male;
 read all var{Lx1 Lx2 Lx3} into x_m;
 close turtles_male;
 print x_m;
 x_mbar={4.73,4.48,3.70};
 e=J(24,1);
 s2=(x_m-e*x_mbar`)`*(x_m-e*x_mbar`)/23;
 print s1 s2;
 spool=(24-1)/(24+24-2)*(s1+s2);
 print spool;
 xbardiff=x_fbar-x_mbar;
 print xbardiff;
 s=(1/24+1/24)*spool;
 s_1=inv(s);
 print s_1;
 Tsquare=xbardiff`*s_1*xbardiff;
 print Tsquare;
 ahat=inv(1/24*s1+1/24*s2)*xbardiff;
 print ahat;
```

```
run;
quit;
```

Output:

```
        S1                              S2

0.0264061  0.020114 0.0249178 0.0110937 0.0080307 0.0081445
 0.020114 0.0161993 0.0194252 0.0080307 0.0064229 0.0059972
0.0249178 0.0194252 0.0249399 0.0081445 0.0059972 0.0067833


                    SPOOL

      0.0187499 0.0140723 0.0165311
      0.0140723 0.0113111 0.0127112
      0.0165311 0.0127112 0.0158616


                   XBARDIFF

                      0.17
                      0.14
                      0.24


                     S_1

      13095.746 -9602.331 -5953.404
      -9602.331 17711.484 -4185.984
      -5953.404 -4185.984 10315.808


                   TSQUARE

                   95.635817


                     AHAT

                   -546.8664
                   -157.4249
                   877.67746
```

Exercise 6_33

(a) Perform a two-factor MANOVA, testing for a species effect, a time effect and species-time interaction. Use α =0.05.

Since the p-value<0.001 for three test, there are significant effects of species, time and species-time interaction.

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall species Effect
H = Type III SSCP Matrix for species
E = Error SSCP Matrix

S=2 M=-0.5 N=12

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.06877382	36.57	4	52	<.0001
Pillai's Trace	0.96119962	12.49	4	54	<.0001
Hotelling-Lawley Trace	13.10458966	84.21	4	30.19	<.0001
Roy's Greatest Root	13.07124729	176.46	2	27	<.0001

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall time Effect
H = Type III SSCP Matrix for time
E = Error SSCP Matrix

S=2 M=-0.5 N=12

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.04916603	45.63	4	52	<.0001
Pillai's Trace	0.99198604	13.29	4	54	<.0001
Hotelling-Lawley Trace	18.50224287	118.89	4	30.19	<.0001
Roy's Greatest Root	18.45689384	249.17	2	27	<.0001

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall species*time Effect
H = Type III SSCP Matrix for species*time
E = Error SSCP Matrix

S=2 M=0.5 N=12

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.08707032	15.53	8	52	<.0001
Pillai's Trace	0.92115522	5.76	8	54	<.0001
Hotelling-Lawley Trace	10.39049958	33.07	8	34.899	<.0001
Roy's Greatest Root	10.38139964	70.07	4	27	<.0001

(b) From the formal normality test and plot, the usual MANOVA assumptions are not satisfied. We need to consider the covariance structure for the repeated measures.

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.862909	Pr < W	0.0004
Kolmogorov-Smirnov	D	0.178779	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.307814	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.72642	Pr > A-Sq	<0.0050

```
                 The UNIVARIATE Procedure
                      Variable:  E1

                   Normal Probability Plot
 5.5+                                                 *
    |
    |                                                 +++++
 2.5+                                            +++++++
    |                                      +++++++ * *
    |                                 ************ *
-0.5+                      ***********
    |                    +++++++
    |              +++*+*+*
-3.5+++++*
     +----+----+----+----+----+----+----+----+----+----+
         -2        -1         0        +1        +2
```

```
                 The UNIVARIATE Procedure
                      Variable:  E2

                   Normal Probability Plot
 22.5+                                                *
     |                                            *    ++++
     |                                          *++++++
  7.5+                                   ++++++*+
     |                              ************
     |                     *********
 -7.5+             * ++++*+++
     |        +++++++
     |++++
-22.5+     *
      +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

(c) For x1(percent spectral reflectance at wavelength 560nm), there is significant interaction between species and time. But for x2(percent spectral reflectance at wavelength 720nm), there is no significant interaction between species and time.

The GLM Procedure

Dependent Variable: x1 percent spectral reflectance at wavelength 560mm(green)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	3036.236856	379.529607	133.67	<.0001
Error	27	76.658775	2.839214		
Corrected Total	35	3112.895631			

R-Square	Coeff Var	Root MSE	x1 Mean
0.975374	11.78205	1.684997	14.30139

Source	DF	Type I SS	Mean Square	F Value	Pr > F
species	2	965.181172	482.590586	169.97	<.0001
time	2	1275.247739	637.623869	224.58	<.0001
species*time	4	795.807944	198.951986	70.07	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
species	2	965.181172	482.590586	169.97	<.0001
time	2	1275.247739	637.623869	224.58	<.0001
species*time	4	795.807944	198.951986	70.07	<.0001

Dependent Variable: x2 percent spectral reflectance at wacelength 720mm(near infrared)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	7794.211339	974.276417	14.86	<.0001
Error	27	1769.642225	65.542305		
Corrected Total	35	9563.853564			

R-Square	Coeff Var	Root MSE	x2 Mean
0.814966	22.62157	8.095820	35.78806

Source	DF	Type I SS	Mean Square	F Value	Pr > F
species	2	2026.856372	1013.428186	15.46	<.0001
time	2	5573.805706	2786.902853	42.52	<.0001
species*time	4	193.549261	48.387315	0.74	0.5741

Source	DF	Type III SS	Mean Square	F Value	Pr > F
species	2	2026.856372	1013.428186	15.46	<.0001
time	2	5573.805706	2786.902853	42.52	<.0001
species*time	4	193.549261	48.387315	0.74	0.5741

(d) Repeated measures design is more appropriate way to analyze these data. Time trend and covariance structure can be taken into account.

SAS code:

```
Title1 'Exercise 6_33 Two-way MANOVA';

Data skull;
infile 'E:\sta5707\data\T6-18.dat';
 Input x1 x2 species $ time replicate @@;
 Label x1='percent spectral reflectance at wavelength 560mm(green)'
          x2='percent spectral reflectance at wacelength 720mm(near
infrared)';
run;

Proc Print Data=skull Label;
Run;

/*Produce summary statistics of the responses.
Compare the correlations here with the partial correlations reported by
GLM below.*/

Title2 "Summary Statistics";
Proc Corr Data=skull;
 Var x1 x2;
Run;

/*Analyze the data as a 2 factor factorial, but we wish to use all 2
responses with replication simultaneously.*/

Title2 "Multivariate Analysis of Variance";
Proc GLM Data=skull;
 Class species time;
 Model x1 x2 = species time species*time / NoUni;
 Manova H=species*time species time/ PrintE PrintH;
```

```
 repeated replicate;
 /* Compute Bonferroni CIs over all Responses.    */
 /* Note that the intervals are simultaneous only  over each factor
separately.   */
 /* Alpha=0.05/2=0.025 was used for p=2.       */
 LSMeans species time / PDiff Adjust=Bon CL Alpha=0.025;
 Output Out=Diag R=E1 E2; /* output residuals for diagnostics */
Run;
Quit;

proc univariate data=diag plot normal;
 var E1 E2;
 run;

Title2 'univariate two-factor ANOVA';
Proc GLM Data=skull;
 Class species time;
 Model x1 = species time species*time;
 *Manova H=species*time species time/ PrintE PrintH;
 *repeated replicate;
Run;
Quit;

Proc GLM Data=skull;
 Class species time;
 Model x2 = species time species*time;
 *Manova H=species*time species time/ PrintE PrintH;
 *repeated replicate;
Run;
Quit;
```

STA 5707 HW10 Li Fan

Exercise 6_24

Construct 95% Simultaneous CI, it is found that

(1)A difference in MaxBreath exists between period 4000 B.C. and period 1850 B.C.; but no difference is observed between period 4000 B.C. and period 3300 B.C.; between period 3300 B.C. and period 1850 B.C.

(2)A difference in BasLength exists between period 4000 B.C. and period 3300 B.C.; between period 4000 B.C. and period 1850 B.C., but no difference is observed between period 3300 B.C. and period 1850 B.C.

The Usual MANOVA assumptions looks satisfied through checking the normality of residuals for each component.

For MaxBreath:

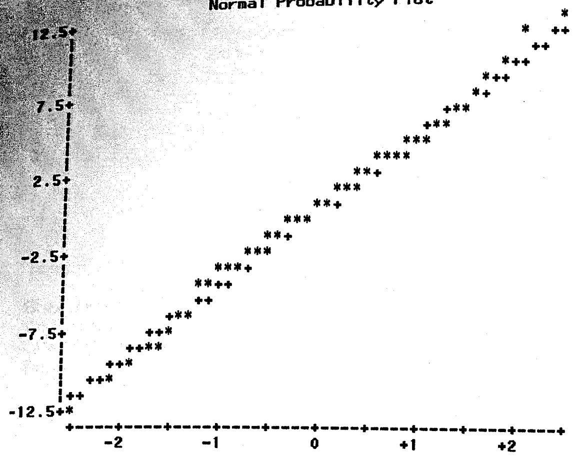
```
        Least Squares Means for Effect TimePeriod

                  Difference
                     Between     95% Confidence Limits for
   i     j             Means        LSMean(i)-LSMean(j)

   1     2         -1.000000        -3.324841      1.324841
   1     3         -3.100000        -5.424841     -0.775159
   2     3         -2.100000        -4.424841      0.224841

                        Tests for Normality

 Test                  --Statistic---     -----p Value------

 Shapiro-Wilk          W     0.982491     Pr < W      0.2678
 Kolmogorov-Smirnov    D      0.07529     Pr > D     >0.1500
 Cramer-von Mises      W-Sq  0.081441     Pr > W-Sq   0.2036
 Anderson-Darling      A-Sq  0.482023     Pr > A-Sq   0.2322

                The UNIVARIATE Procedure
                     Variable:  R1

                     Normal Probability Plot
   15+                                                    *
     |
     |                                                   +++
     |                                              ++++*+
     |                                           +***
     |                                       ++*** *
     |                                    ******
    1+                                *****
     |                            ******
     |                         ****+
     |                     +****
     |                  +**+*
     |              +*+*+*
     |+++
  -13+*
      +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

For BasHeight:

```
        Least Squares Means for Effect TimePeriod

                  Difference
                     Between     95% Confidence Limits for
   i     j             Means        LSMean(i)-LSMean(j)

   1     2          0.900000        -1.513581      3.313581
   1     3         -0.200000        -2.613581      2.213581
   2     3         -1.100000        -3.513581      1.313581

                        Tests for Normality

 Test                  --Statistic---     -----p Value------

 Shapiro-Wilk          W     0.988847     Pr < W      0.6462
 Kolmogorov-Smirnov    D     0.065033     Pr > D     >0.1500
 Cramer-von Mises      W-Sq  0.058122     Pr > W-Sq  >0.2500
 Anderson-Darling      A-Sq  0.389462     Pr > A-Sq  >0.2500
```

The UNIVARIATE Procedure
Variable: R2

Normal Probability Plot



For BasLength:

Least Squares Means for Effect TimePeriod

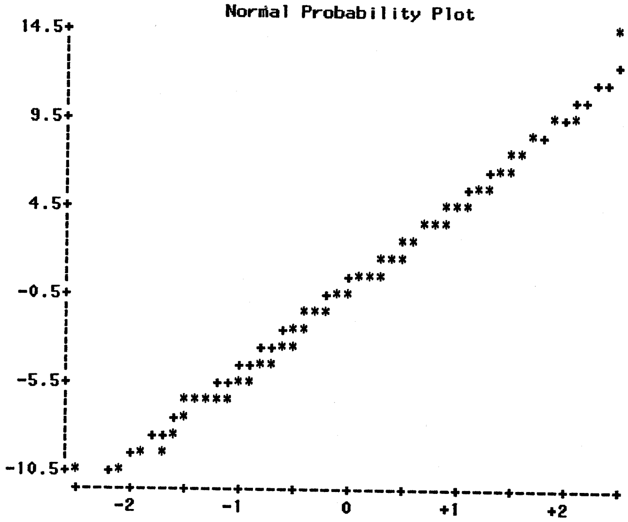
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.100000	-2.452981	2.652981
1	3	3.133333	0.580352	5.686314
2	3	3.033333	0.480352	5.586314

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.988946	Pr < W	0.6535
Kolmogorov-Smirnov	D	0.058599	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.037097	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.255298	Pr > A-Sq	>0.2500

The UNIVARIATE Procedure
Variable: R3

Normal Probability Plot



For NasHeight:

Least Squares Means for Effect TimePeriod

i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.300000	-1.294839	1.894839
1	3	-0.033333	-1.628172	1.561506
2	3	-0.333333	-1.928172	1.261506

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.986421	Pr < W	0.4765
Kolmogorov-Smirnov	D	0.068903	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.054824	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.336162	Pr > A-Sq	>0.2500

The UNIVARIATE Procedure
Variable: R4

Normal Probability Plot

```
 9.5+                                                     *
    |                                                  +++
    |                                               +++
    |                                            *+* *
    |                                         ***
    |                                     ***+
    |                                 ****
    |                              **+
 1.5+                          +****
    |                       ****
    |                    ***
    |                 ***
    |             +***
    |          +****
    |    * *+* *
-6.5+*  +++
     +----+----+----+----+----+----+----+----+----+----+
         -2        -1         0        +1        +2
```

```
Proc GLM Data=EgyptianSkulls Order=Internal;
 Class TimePeriod;
 Model MaxBreadth BasHeight BasLength NasHeight = TimePeriod;
 MANOVA H=TimePeriod;
 LSMeans TimePeriod / PDiff CL Alpha=0.05;
 Output Out=Diagnostics R=R1-R4; /* output the residuals */
Run;
Quit;

proc univariate data=diagnostics plot normal;
 var R1-R4;
run;
```

Exercise 6_25
Since the assumption is invalid when constructing the model based on the original data, it needs to consider the data transformation as following: x3→sqrt(x3) x4→log(x4).

For vanadium(x1),there is difference between zone of submuli and Wilhelm or upper. No difference is observed between zone of Wilhelm and upper.

Least Squares Means for Effect zone

i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-2.780861	-4.072781	-1.488941
1	3	1.216883	-0.607503	3.041269
2	3	3.997744	2.445746	5.549742

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.976806	Pr < W	0.3523
Kolmogorov-Smirnov	D	0.107089	Pr > D	0.1074
Cramer-von Mises	W-Sq	0.08181	Pr > W-Sq	0.1998
Anderson-Darling	A-Sq	0.478881	Pr > A-Sq	0.2328

The UNIVARIATE Procedure
Variable: R1

Normal Probability Plot

```
 4.25+                                               +++
     |                                            *++ *
 3.25+                                          ++
     |                                       +*+*
 2.25+                                    ****
     |                                  **
 1.25+                               ***
     |                           *****
 0.25+                         ***
     |                      ***
-0.75+                   ****
     |                **+*
-1.75+              **+
     |            +*+
-2.75+         +++*
     |       +* **
-3.75+    ++*
     | +++
-4.75++
     |
-5.75+  *
      +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

For iron(x2), there is difference among zone of Wilhelm, submuli, and upper with each other.

Least Squares Means for Effect zone

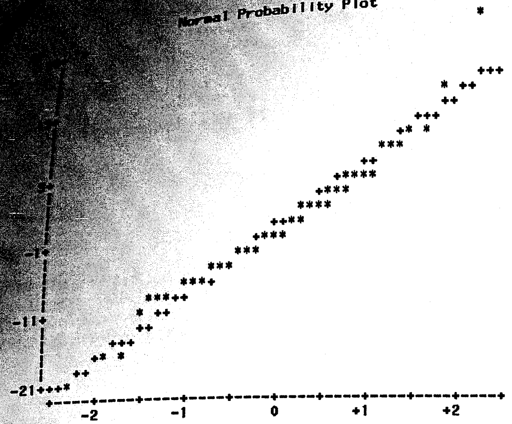
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	10.838278	4.709643	16.966912
1	3	-10.480519	-19.135074	-1.825965
2	3	-21.318797	-28.681194	-13.956400

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.967757	Pr < W	0.1387
Kolmogorov-Smirnov	D	0.084206	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.075886	Pr > W-Sq	0.2349
Anderson-Darling	A-Sq	0.505874	Pr > A-Sq	0.2027

The UNIVARIATE Procedure
Variable: R2

Normal Probability Plot



For beryllium(x3), there is difference between zone of submuli and Wilhelm or upper. No difference is observed between zone of Wilhelm and upper.

Least Squares Means for Effect zone

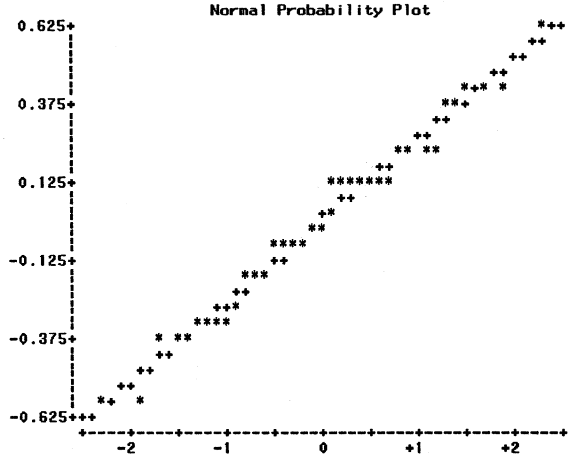
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-0.254154	-0.432632	-0.075676
1	3	0.040663	-0.211375	0.292700
2	3	0.294817	0.080409	0.509224

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.976082	Pr < W	0.3281
Kolmogorov-Smirnov	D	0.121427	Pr > D	0.0391
Cramer-von Mises	W-Sq	0.103257	Pr > W-Sq	0.0996
Anderson-Darling	A-Sq	0.580635	Pr > A-Sq	0.1301

The UNIVARIATE Procedure
Variable: R3

Normal Probability Plot



For saturated hydrocarbons(x4), there is difference between zone of submuli and Wilhelm or upper. No difference is observed between zone of Wilhelm and upper.

Least Squares Means for Effect zone

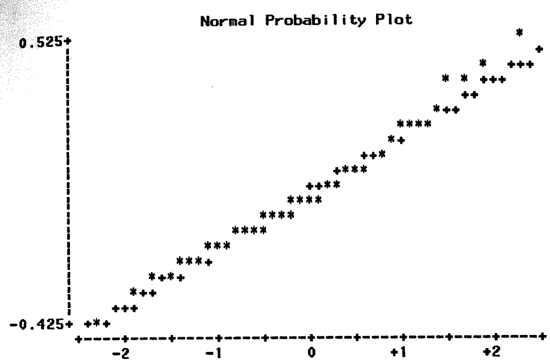
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.346943	0.216119	0.477767
1	3	-0.043061	-0.227804	0.141682
2	3	-0.390004	-0.547164	-0.232843

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.976267	Pr < W	0.3341
Kolmogorov-Smirnov	D	0.103106	Pr > D	0.1418
Cramer-von Mises	W-Sq	0.079567	Pr > W-Sq	0.2131
Anderson-Darling	A-Sq	0.471947	Pr > A-Sq	0.2405

The UNIVARIATE Procedure
Variable: R4

Normal Probability Plot



For aromatic hydrocarbons(x5), there is difference between zone of upper and Wilhelm or submuli. No difference is observed between zone of Wilhelm and submuli.

Least Squares Means for Effect zone

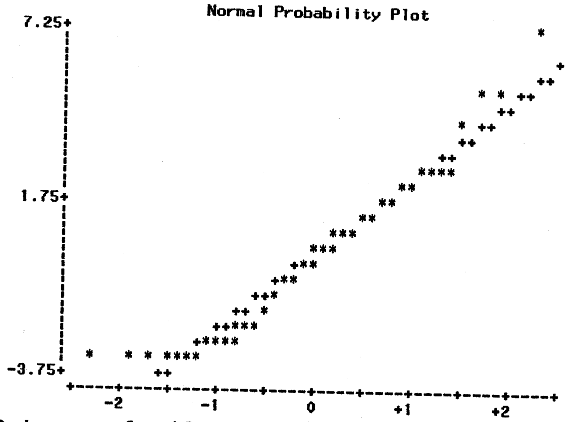
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-0.284258	-2.018545	1.450029
1	3	-6.056364	-8.505438	-3.607289
2	3	-5.772105	-7.855524	-3.688687

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.947023	Pr < W	0.0157
Kolmogorov-Smirnov	D	0.105569	Pr > D	0.1205
Cramer-von Mises	W-Sq	0.07268	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.658993	Pr > A-Sq	0.0846

The UNIVARIATE Procedure
Variable: R5

Normal Probability Plot




```
Data crudeoil;
 infile 'E:\sta5707\data\T11-7.dat';
 Input x1-x5 zone $;
 Label x1="vanadium (in percnet ash)"
        x2 ="iron(in percent ash)"
          x3 ="beryllium(in percent ash)"
          x4="saturated hydrocarbons(in percent area)"
            x5="aromatic hydrocarbons(in percent area)";
Sx3=sqrt(x3);
Lx4=log(x4);
run;
Proc GLM Data=crudeoil;
 Class zone;
 Model x1 x2 Sx3 Lx4 x5 = zone;
 MANOVA H=zone;
 LSMeans zone / PDiff CL Adjust=Tukey Alpha=0.01; /* 0.05/5 */
 Output Out=Diagnostics_trans R=R1-R5; /* output the residuals */
Run;
Quit;

proc univariate data=diagnostics_trans plot normal;
 var R1-R5;
run;
```

HW10 Xin Li

6.24.

One Way MANOVA test: p value is 0.044. The null hypothesis is rejected at 5% level implying there is a time period effect.

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.83010268	2.05	8	168	0.0436
Pillai's Trace	0.17221185	2.00	8	170	0.0489
Hotelling-Lawley Trace	0.20188200	2.11	8	117.7	0.0405
Roy's Greatest Root	0.18696913	3.97	4	85	0.0053

From the SAS output, p value for MaxBreadth and BasLength is smaller than 0.05, for the other two variable is bigger than 0.05. It implies than any differences over time periods are probably due to changes in MaxBreadth and BasLength.

95% simultaneous confidence interval: (values from the SAS output)

$m = pg(g-1)/2 = 12$

$t_{87}(0.05/24) = 2.94$

MaxBreadth:

$$\tau_{11} - \tau_{21} : (131.36 - 132.36) \pm 2.94\sqrt{\frac{1785.4}{87}\left(\frac{1}{30} + \frac{1}{30}\right)} \Rightarrow -1 \pm 3.44$$

$\tau_{11} - \tau_{31} : -3.1 \pm 3.44$

$\tau_{21} - \tau_{31} : -2.1 \pm 3.44$

BasHeight:

$$\tau_{12} - \tau_{22} : -0.9 \pm 2.94\sqrt{\frac{1924.3}{87}\left(\frac{1}{30} + \frac{1}{30}\right)} \Rightarrow -0.9 \pm 3.57$$

$\tau_{12} - \tau_{32} : -0.2 \pm 3.57$

$\tau_{22} - \tau_{32} : -1.1 \pm 3.57$

BasLength:

$$\tau_{13} - \tau_{23} : -0.1 \pm 2.94\sqrt{\frac{2153}{87}\left(\frac{1}{30} + \frac{1}{30}\right)} \Rightarrow -0.1 \pm 3.78$$

$\tau_{13} - \tau_{33} : 3.14 \pm 3.78$

$\tau_{23} - \tau_{33} : 3.03 \pm 3.78$

NasHeight:

$$\tau_{14} - \tau_{24} : -0.3 \pm 2.94\sqrt{\frac{840.2}{87}\left(\frac{1}{30} + \frac{1}{30}\right)} \Rightarrow -0.3 \pm 2.36$$

$\tau_{14} - \tau_{34} : -0.03 \pm 2.36$

$\tau_{24} - \tau_{34} : -0.33 \pm 2.36$

All the intervals includes 0. Evidence for changes over time is marginal.

From the residue plot below, the data seems normal. The usual MANOVA assumptions are realistic these data.

SAS output:

Dependent Variable: **MaxBreadth** Maximum Breadth of Skull (mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	150.200000	75.100000	3.66	0.0298
Error	87	1785.400000	20.521839		

Dependent Variable: **BasHeight** Basibregmatic Height of Skull (mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	20.600000	10.300000	0.47	0.6293
Error	87	1924.300000	22.118391		

Dependent Variable: **BasLength** Basialveolar Length of Skull (mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	190.288889	95.144444	3.84	0.0251
Error	87	2153.000000	24.747126		

Dependent Variable: **NasHeight** Nasal Height of Skull (mm)

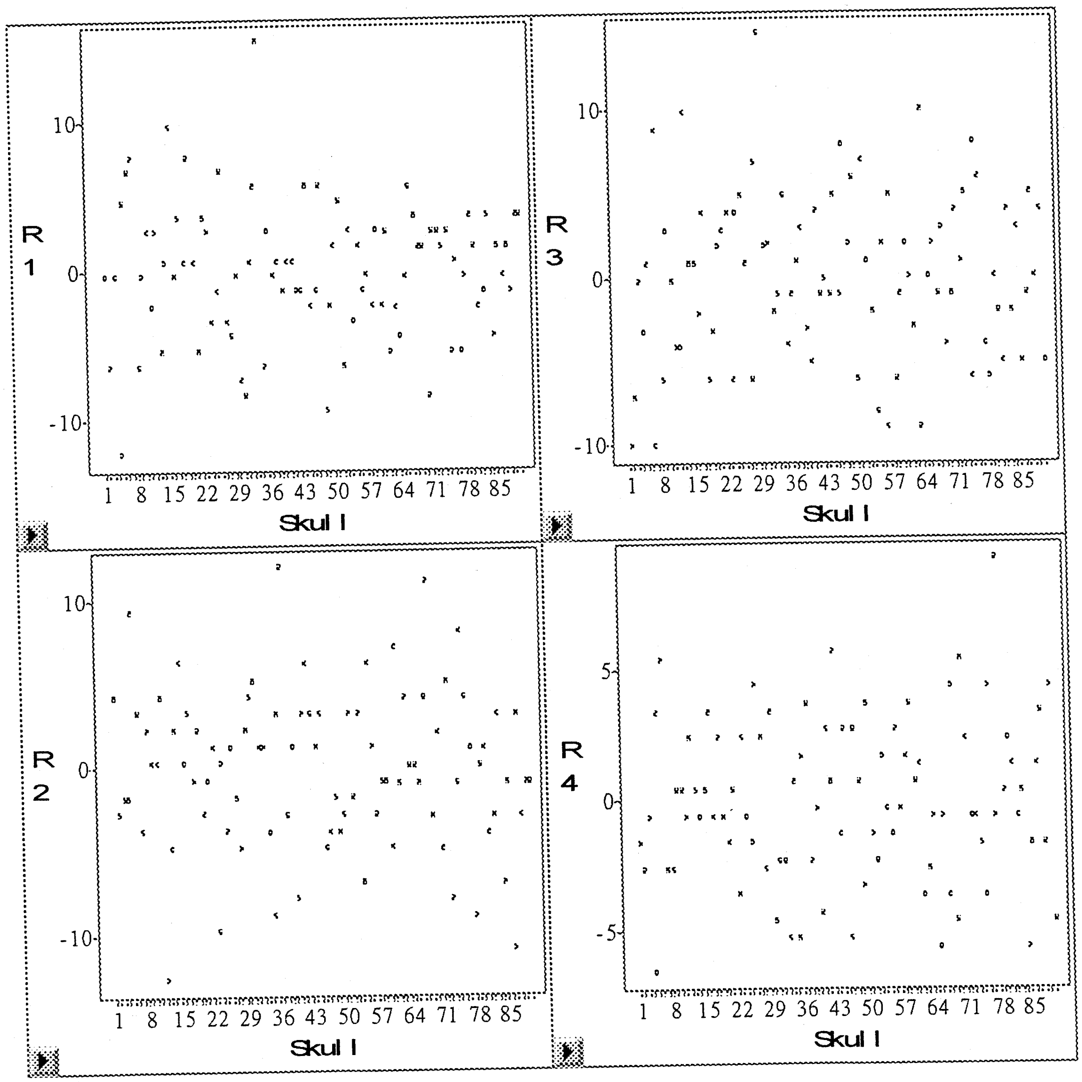
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2.0222222	1.0111111	0.10	0.9007
Error	87	840.2000000	9.6574713		

Time Period	MaxBreadth LSMEAN	LSMEAN Number
4000 B.C.	131.366667	1
3300 B.C.	132.366667	2
1850 B.C.	134.466667	3

Time Period	BasHeight LSMEAN	LSMEAN Number
4000 B.C.	133.600000	1
3300 B.C.	132.700000	2
1850 B.C.	133.800000	3

Time Period	BasLength LSMEAN	LSMEAN Number
4000 B.C.	99.1666667	1
3300 B.C.	99.0666667	2
1850 B.C.	96.0333333	3

Time Period	NasHeight LSMEAN	LSMEAN Number
4000 B.C.	50.5333333	1
3300 B.C.	50.2333333	2
1850 B.C.	50.5666667	3



6.25. (no transformation)

One Way MANOVA test: p value is smaller than 0.0001. The null hypothesis is rejected at 5% level.

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.11591099	18.98	10	98	<.0001
Pillai's Trace	1.20665556	15.21	10	100	<.0001
Hotelling-Lawley Trace	4.84442811	23.43	10	70.804	<.0001
Roy's Greatest Root	4.17841435	41.78	5	50	<.0001

From the SAS output, p value for all these five variable are smaller than 0.05, implying these five variables are responsible for any difference in the zones.

95% simultaneous confidence interval: (values from the SAS output)

$m = pg(g-1)/2 = 15$

$t_{53}(0.05/30) = 3.074$

X1:

$$\tau_{11} - \tau_{21} : (4.45 - 7.23) \pm 3.074\sqrt{\frac{187.58}{53}\left(\frac{1}{7} + \frac{1}{12}\right)} \Rightarrow -2.78 \pm 2.75$$

$$\tau_{11} - \tau_{31} : (4.45 - 3.23) \pm 3.074\sqrt{\frac{187.58}{53}\left(\frac{1}{7} + \frac{1}{37}\right)} \Rightarrow 1.22 \pm 2.38$$

$$\tau_{21} - \tau_{31} : (7.23 - 3.23) \pm 3.074\sqrt{\frac{187.58}{53}\left(\frac{1}{12} + \frac{1}{37}\right)} \Rightarrow 4.00 \pm 1.92$$

X2:

$$\tau_{12} - \tau_{22} : (33.09 - 22.25) \pm 3.074\sqrt{\frac{4221.16}{53}\left(\frac{1}{7} + \frac{1}{12}\right)} \Rightarrow -10.84 \pm 13.05$$

$$\tau_{12} - \tau_{32} : (33.09 - 43.57) \pm 3.074\sqrt{\frac{4221.16}{53}\left(\frac{1}{7} + \frac{1}{37}\right)} \Rightarrow -10.48 \pm 11.31$$

$$\tau_{22} - \tau_{32} : (22.25 - 43.57) \pm 3.074\sqrt{\frac{4221.16}{53}\left(\frac{1}{12} + \frac{1}{37}\right)} \Rightarrow -21.32 \pm 9.11$$

X3:

$\tau_{13}-\tau_{23}$: $(0.17-0.43)\pm 3.074\sqrt{\frac{4.44}{53}\left(\frac{1}{7}+\frac{1}{12}\right)} \Rightarrow -0.26\pm 0.42$

$\tau_{13}-\tau_{33}$: $(0.17-0.12)\pm 3.074\sqrt{\frac{4.44}{53}\left(\frac{1}{7}+\frac{1}{37}\right)} \Rightarrow 0.05\pm 0.37$

$\tau_{23}-\tau_{33}$: $(0.43-0.12)\pm 3.074\sqrt{\frac{4.44}{53}\left(\frac{1}{12}+\frac{1}{37}\right)} \Rightarrow 0.31\pm 0.30$

X4:

$\tau_{14}-\tau_{24}$: $(6.56-4.66)\pm 3.074\sqrt{\frac{57.04}{53}\left(\frac{1}{7}+\frac{1}{12}\right)} \Rightarrow 1.90\pm 1.52$

$\tau_{14}-\tau_{34}$: $(6.56-6.80)\pm 3.074\sqrt{\frac{57.04}{53}\left(\frac{1}{7}+\frac{1}{37}\right)} \Rightarrow -0.24\pm 1.31$

$\tau_{24}-\tau_{34}$: $(4.66-6.80)\pm 3.074\sqrt{\frac{57.04}{53}\left(\frac{1}{12}+\frac{1}{37}\right)} \Rightarrow -2.14\pm 1.06$

X5:

$\tau_{15}-\tau_{25}:(5.48-5.77)\pm 3.074\sqrt{\frac{338.02}{53}\left(\frac{1}{7}+\frac{1}{12}\right)} \Rightarrow -0.29\pm 3.69$

$\tau_{15}-\tau_{35}:(5.48-11.54)\pm 3.074\sqrt{\frac{338.02}{53}\left(\frac{1}{7}+\frac{1}{37}\right)} \Rightarrow -6.06\pm 3.20$

$\tau_{25}-\tau_{35}:(5.77-11.54)\pm 3.074\sqrt{\frac{338.02}{53}\left(\frac{1}{12}+\frac{1}{37}\right)} \Rightarrow -5.77\pm 2.58$

The intervals for X1 $(\tau_{11}-\tau_{21},\tau_{21}-\tau_{31})$,X2$(\tau_{21}-\tau_{31})$, 3$(\tau_{21}-\tau_{31})$,X4 $(\tau_{11}-\tau_{21},\tau_{21}-\tau_{31})$ and X5$(\tau_{11}-\tau_{31}\tau_{21}-\tau_{31})$ don't include 0.

SAS output:

Dependent Variable: x1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	135.6731502	67.8365751	19.17	<.0001
Error	53	187.5752427	3.5391555		
Corrected Total	55	323.2483929			

Dependent Variable: x2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3186.681172	1593.340586	20.01	<.0001

Error	53	4221.158113	79.644493		

Dependent Variable: x3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.98442037	0.49221018	5.88	0.0049
Error	53	4.43566535	0.08369180		
Corrected Total	55	5.42008571			

Dependent Variable: x4

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	48.8034220	24.4017110	22.67	<.0001
Error	53	57.0404334	1.0762346		
Corrected Total	55	105.8438554			

Dependent Variable: x5

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	209.2941996	104.6470998	16.41	<.0001
Error	53	338.0228861	6.3777903		
Corrected Total	55	547.3170857			

y	x1 LSMEAN	LSMEAN Number
SubMuli	4.44545455	1
Upper	7.22631579	2
Wilhelm	3.22857143	3

y	x2 LSMEAN	LSMEAN Number
SubMuli	33.0909091	1
Upper	22.2526316	2
Wilhelm	43.5714286	3

y	x3 LSMEAN	LSMEAN Number
SubMuli	0.17090909	1
Upper	0.43210526	2
Wilhelm	0.11714286	3

y	x4 LSMEAN	LSMEAN Number
SubMuli	6.56090909	1
Upper	4.65815789	2
Wilhelm	6.79571429	3

y	x5 LSMEAN	LSMEAN Number
SubMuli	5.4836364	1
Upper	5.7678947	2
Wilhelm	11.5400000	3

(with transformation for 6.25, x6=sqrt(x3), x7=log(x4))

Dependent Variable: x1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	70.8412407	35.4206204	10.11	0.0007
Error	24	84.0506111	3.5021088		
Corrected Total	26	154.8918519			

Dependent Variable: x2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1552.255556	776.127778	11.93	0.0003
Error	24	1561.744444	65.072685		
Corrected Total	26	3114.000000			

Dependent Variable: x5

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	128.2116830	64.1058415	11.36	0.0003
Error	24	135.4106800	5.6421117		
Corrected Total	26	263.6223630			

Dependent Variable: x6

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	69.4933889	34.7466944	8.63	0.0015
Error	24	96.6132778	4.0255532		
Corrected Total	26	166.1066667			

Dependent Variable: x7

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2010.758519	1005.379259	10.80	0.0005
Error	24	2233.484444	93.061852		

95% simultaneous confidence interval:

y	x1 LSMEAN	LSMEAN Number
SubMuli	4.86000000	1
Upper	7.42222222	2
Wilhelm	3.22500000	3

y	x1 LSMEAN	95% Confidence Limits	
SubMuli	4.860000	2.519208	7.200792
Upper	7.422222	6.188516	8.655928
Wilhelm	3.225000	0.607915	5.842085

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-2.562222	-5.602700	0.478256
1	3	1.635000	-2.399646	5.669646
2	3	4.197222	0.872585	7.521860

y	x2 LSMEAN	LSMEAN Number
SubMuli	33.2000000	1
Upper	21.9444444	2
Wilhelm	42.0000000	3

y	x2 LSMEAN	95% Confidence Limits	
SubMuli	33.200000	23.109854	43.290146
Upper	21.944444	16.626471	27.262418
Wilhelm	42.000000	30.718874	53.281126

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	11.255556	-1.850633	24.361744
1	3	-8.800000	-26.191619	8.591619
2	3	-20.055556	-34.386633	-5.724478

y	x5 LSMEAN	LSMEAN Number
SubMuli	3.8980000	1
Upper	6.0900000	2
Wilhelm	11.2800000	3

y	x5 LSMEAN	95% Confidence Limits	
SubMuli	3.898000	0.926889	6.869111
Upper	6.090000	4.524087	7.655913
Wilhelm	11.280000	7.958197	14.601803

Least Squares Means for Effect y

Difference	Simultaneous 95%

i	j	Between Means	Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-2.192000	-6.051205	1.667205
1	3	-7.382000	-12.503079	-2.260921
2	3	-5.190000	-9.409882	-0.970118

y	x6 LSMEAN	LSMEAN Number
SubMuli	3.92000000	1
Upper	7.21111111	2
Wilhelm	3.67500000	3

y	x6 LSMEAN	95% Confidence Limits	
SubMuli	3.920000	1.410363	6.429637
Upper	7.211111	5.888416	8.533806
Wilhelm	3.675000	0.869141	6.480859

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-3.291111	-6.550903	-0.031320
1	3	0.245000	-4.080670	4.570670
2	3	3.536111	-0.028337	7.100559

y	x7 LSMEAN	LSMEAN Number
SubMuli	30.2000000	1
Upper	21.6444444	2
Wilhelm	46.0000000	3

y	x7 LSMEAN	99% Confidence Limits	
SubMuli	30.200000	18.133428	42.266572
Upper	21.644444	15.284802	28.004086
Wilhelm	46.000000	32.509162	59.490838

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	8.555556	-7.117833	24.228944
1	3	-15.800000	-36.598235	4.998235
2	3	-24.355556	-41.493760	-7.217351

Exercise 8_14

Simple Statistics

	x1	x2	x3
Mean	4.640000000	45.40000000	9.965000000
StD	1.696870184	14.13465320	1.904641146

Covariance Matrix

		x1	x2	x3
x1	sweet rate	2.8793684	10.0100000	-1.8090526
x2	sodium	10.0100000	199.7884211	-5.6400000
x3	potassium	-1.8090526	-5.6400000	3.6276579

Total Variance 206.29544737

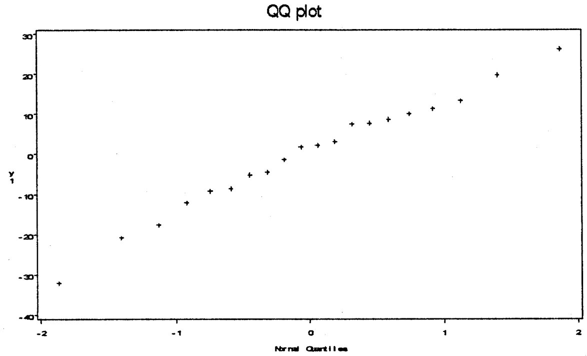
Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	200.462464	195.930874	0.9717	0.9717
2	4.531591	3.230198	0.0220	0.9937
3	1.301392		0.0063	1.0000

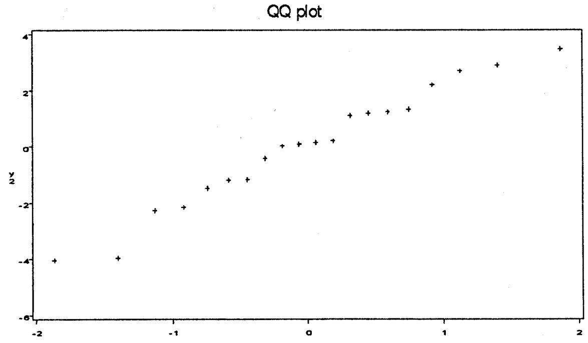
Eigenvectors

		Prin1	Prin2	Prin3
x1	sweet rate	0.050841	-.573704	0.817484
x2	sodium	0.998284	0.053020	-.024877
x3	potassium	-.029072	0.817345	0.575415

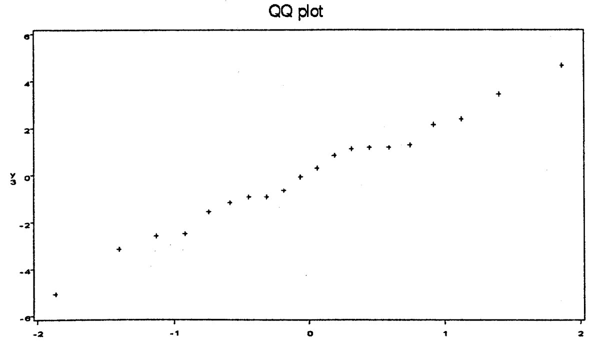
QQ plot



QQ plot



QQ plot



sas code:

```
data ex8_14;
 input x1-x3;
 label x1='sweet rate'
       x2='sodium'
       x3='potassium';
datalines;
 3.7  48.5   9.3
 5.7  65.1   8.0
 3.8  47.2  10.9
 3.2  53.2  12.0
 3.1  55.5   9.7
 4.6  36.1   7.9
 2.4  24.8  14.0
 7.2  33.1   7.6
 6.7  47.4   8.5
 5.4  54.1  11.3
 3.9  36.9  12.7
 4.5  58.8  12.3
 3.5  27.8   9.8
 4.5  40.2   8.4
 1.5  13.5  10.1
 8.5  56.4   7.1
 4.5  71.6   8.2
 6.5  52.8  10.9
 4.1  44.1  11.2
 5.5  40.9   9.4
  ;
  run;

  proc princomp cov data=ex8_14 out=result;
   var x1 x2 x3;
  run;

 data plotting;
   set ex8_14;
   y1=0.0508*(x1-4.64)+1.0*(x2-45.4)-0.029*(x3-9.965);
   y2=-0.57*(x1-4.64)+0.053*(x2-45.4)+0.817*(x3-9.965);
   y3=-0.817*(x1-4.64)-0.025*(x2-45.4)+0.575*(x3-9.965);
  run;

  proc capability data=plotting;
   qqplot y1;
   qqplot y2;
   qqplot y3;
  run;
```

8.14.

Pincipal component analysis:

The PRINCOMP Procedure

Observations	20
Variables	3

Simple Statistics

	x1	x2	x3
Mean	4.640000000	45.40000000	9.965000000
StD	1.696870184	14.13465320	1.904641146

Covariance Matrix

	x1	x2	x3
x1	2.8793684	10.0100000	-1.8090526
x2	10.0100000	199.7884211	-5.6400000
x3	-1.8090526	-5.6400000	3.6276579

Total Variance 206.29544737

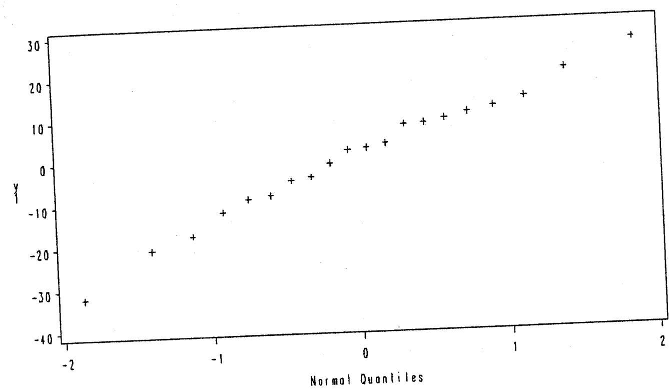
Eigenvalues of the Covariance Matrix

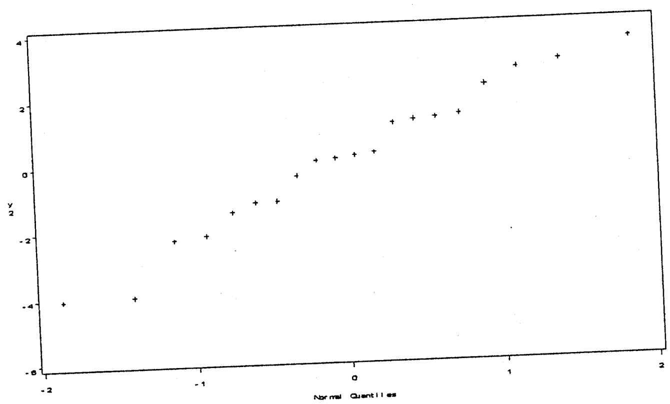
	Eigenvalue	Difference	Proportion	Cumulative
1	200.462464	195.930874	0.9717	0.9717
2	4.531591	3.230198	0.0220	0.9937
3	1.301392		0.0063	1.0000

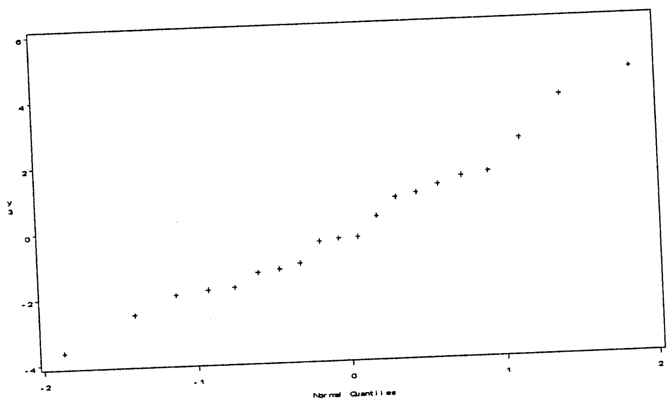
Eigenvectors

	Prin1	Prin2	Prin3
x1	0.050841	-.573704	0.817484
x2	0.998284	0.053020	-.024877
x3	-.029072	0.817345	0.575415

Q-Q plot: there are no suspect observations since all the observations are close to the line.







SAS Code:

```
(for 6.25)
Proc Format;
 Value BC 1="Wilhelm."
          2="SubMuli."
          3="Upper";
Run;

Data EgyptianSkulls;
 Skull+1;
 Input x1-x5 y $;
DataLines;
......
Proc Print Data=EgyptianSkulls Label;
```

```
x6=sqrt(x3);
x7=log(x4);
run;
Proc Print Data=EgyptianSkulls Label;
 Id Skull;
Run;

Proc IML;
 Reset Print Log FW=12;
 /* Specify the values */
 X={4000 3300 1850};
 /* Get orthogonal polynomials (rows) */
 P=(Orpol(X))`;
 /* Drop the coefficients into contrast statements */
Quit;

/*
 * Examine data for homogeneity of covariance matrices
 */
Proc Discrim Data=EgyptianSkulls Method=Normal Pool=Test;
 Class y;
 Var x1 x2 x5 x6 x7;
Run;

/*
 * MANOVA analysis
 * Use the ORDER=INTERNAL option so that contrast
 * coefficients correspond correctly with the Time Periods.
 */
Proc GLM Data=EgyptianSkulls Order=Internal;
 Class y;
 Model x1 x2  x5 x6 x7 = y;
 Contrast "Linear TimePeriod    " y 0.6125845862 0.1612064701 -
0.773791056;
 Contrast "Quadratic TimePeriod" y 0.5398210735 -0.80042435
0.2606032769;
 MANOVA H=y;
 LSMeans y / PDiff cl Adjust=Tukey Alpha=0.01; /* 0.05/5 */
 Output Out=Diagnostics R=R1-R5; /* output the residuals */
Run;
Quit;

/*
 * Invoke SAS/Insight for Interactive Data Analysis
 * to examine residuals for normality and outliers.
 */
Proc Insight Data=Diagnostics;
Run;

/*
 * Consider the profile analysis
 */
Proc GLM Data=EgyptianSkulls Order=Internal;
 Class y;
 Model x1 x2 x3 x4 x5 = y / NoUni;
 Contrast "Linear TimePeriod    " y 0.6125845862 0.1612064701 -
0.773791056;
```

```
 Contrast "Quadratic TimePeriod" y 0.5398210735 -0.80042435
0.2606032769;
 MANOVA H=y
        M=x1-x2,x2-x6,x6-x7,x7-x5;
 MANOVA H=y
        M=x1+x2+x6+x7+x5;
Run; (For 8.14)
data sweat;
input x1-x3;
Datalines;
...
;
proc princomp cov data=sweat out=result;
var x1 x2 x3;
run;

data sweatpc;
set sweat;
y1= ;
y2=;
y3=;
run;
proc capability data=sweatpc;
qqplot y1;
qqplot y2;
qqplot y3;
run;
```

√ 5.1 a) $\bar{\mathbf{x}} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$; $S = \begin{bmatrix} 8 & -10/3 \\ -10/3 & 2 \end{bmatrix}$

$T^2 = 150/11 = 13.64$

b) T^2 is $3F_{2,2}$ (see (5-5))

c) $H_0: \boldsymbol{\mu}' = [7, 11]$

$\alpha = .05$ so $F_{2,2}(.05) = 19.00$

Since $T^2 = 13.64 < 3F_{2,2}(.05) = 3(19) = 57$; do not reject H_0 at the $\alpha = .05$ level

√ 5.3 a) $T^2 = \dfrac{(n-1)\left|\sum_{j=1}^{n}(\mathbf{x}_j-\boldsymbol{\mu}_0)(\mathbf{x}_j-\boldsymbol{\mu}_0)'\right|}{\left|\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right|} - (n-1) = \dfrac{3(244)}{44} - 3 = 13.64$

b) $\Lambda = \left(\dfrac{\left|\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right|}{\left|\sum_{j=1}^{n}(\mathbf{x}_j-\boldsymbol{\mu}_0)(\mathbf{x}_j-\boldsymbol{\mu}_0)'\right|}\right)^{n/2} = \left(\dfrac{44}{244}\right)^2 = .0325$

Wilks' lambda $= \Lambda^{2/n} = \Lambda^{1/2} = \sqrt{.0325} = .1803$

√ 5.5 $H_0: \boldsymbol{\mu}' = [.55, .60]$; $T^2 = 1.17$

$\alpha = .05$; $F_{2,40}(.05) = 3.23$

Since $T^2 = 1.17 < \dfrac{2(41)}{40} F_{2,40}(.05) = 2.05(3.23) = 6.62$,

we do not reject H_0 at the $\alpha = .05$ level. The result is consistent with the 95% confidence ellipse for $\boldsymbol{\mu}$ pictured in Figure 5.1 since $\boldsymbol{\mu}' = [.55, .60]$ is <u>inside</u> the ellipse.

5.7

$$\overline{x} = \begin{bmatrix} 4.640 \\ 45.400 \\ 9.965 \end{bmatrix}, S = \begin{bmatrix} 2.879 & 10.010 & -1.810 \\ 10.010 & 199.788 & -5.640 \\ -1.810 & -5.640 & 3.628 \end{bmatrix}$$

(i)

95% simultaneous T^2 confidence intervals for μ_1, μ_2, μ_3:

$$\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) = \frac{(20-1)3}{20-3} F_{3,17}(0.05) = \frac{(19)3}{17}(3.20) = 10.73$$

$$4.640 - \sqrt{10.73}\sqrt{\frac{2.879}{20}} \le \mu_1 \le 4.640 + \sqrt{10.73}\sqrt{\frac{2.879}{20}} \qquad \Longrightarrow 3.40 \le \mu_1 \le 5.88$$

$$45.400 - \sqrt{10.73}\sqrt{\frac{199.788}{20}} \le \mu_2 \le 45.400 + \sqrt{10.73}\sqrt{\frac{199.788}{20}} \qquad \Longrightarrow 35.05 \le \mu_2 \le 55.75$$

$$9.965 - \sqrt{10.73}\sqrt{\frac{3.628}{20}} \le \mu_3 \le 9.965 + \sqrt{10.73}\sqrt{\frac{3.628}{20}} \qquad \Longrightarrow 8.57 \le \mu_3 \le 11.36$$

(ii)

95% simultaneous Bonferroni confidence intervals for μ_1, μ_2, μ_3:

$$t_{n-1}\left(\frac{\alpha}{2p}\right) = t_{20-1}\left(\frac{0.05}{2 \times 3}\right) = t_{19}(0.0083) = 2.625$$

$$4.640 - 2.625\sqrt{\frac{2.879}{20}} \le \mu_1 \le 4.640 + 2.625\sqrt{\frac{2.879}{20}} \qquad \Longrightarrow \overset{3.64}{\cancel{4.03}} \le \mu_1 \le \overset{5.64}{\cancel{5.25}}$$

$$45.400 - 2.625\sqrt{\frac{199.788}{20}} \le \mu_2 \le 45.400 + 2.625\sqrt{\frac{199.788}{20}} \qquad \Longrightarrow \overset{37.10}{\cancel{40.28}} \le \mu_2 \le \overset{53.70}{\cancel{50.52}}$$

$$9.965 - 2.625\sqrt{\frac{3.628}{20}} \le \mu_3 \le 9.965 + 2.625\sqrt{\frac{3.628}{20}} \qquad \Longrightarrow \overset{8.85}{\cancel{9.27}} \le \mu_3 \le \overset{11.08}{\cancel{10.66}}$$

(iii)

Compare the intervals in part(i) and part(ii), we see that Bonferroni intervals provide more precise estimates than the T^2-intervals.

√ 5.12 Initial estimates are

$$\tilde{\mu} = \begin{bmatrix} 4 \\ 6 \\ 2 \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} 0.5 & 0.0 & 0.5 \\ & 2.0 & 0.0 \\ & & 1.5 \end{bmatrix}.$$

The first revised estimates are

$$\tilde{\mu} = \begin{bmatrix} 4.0833 \\ 6.0000 \\ 2.2500 \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} 0.6042 & 0.1667 & 0.8125 \\ & 2.500 & 0.0 \\ & & 1.9375 \end{bmatrix}.$$

5.13 The χ^2 distribution with 3 degrees of freedom.

5.14 Length of one-at-a time t-interval / Length of Bonferroni interval $= t_{n-1}(\alpha/2)/t_{n-1}(\alpha/2m)$.

n	m = 2	m = 4	m = 10
15	0.8546	0.7489	0.6449
25	0.8632	0.7644	0.6678
50	0.8691	0.7749	0.6836
100	0.8718	0.7799	0.6910
∞	0.8745	0.7847	0.6983

√ 5.15

(a).

$$\begin{aligned} E(X_{ij}) &= (1)p_i + (0)(1-p_i) = p_i. \\ Var(X_{ij}) &= (1-p_i)^2 p_i + (0-p_i)^2(1-p_i) = p_i(1-p_i) \end{aligned}$$

(b). $Cov(X_{ij}, X_{kj}) = E(X_{ij}X_{ik}) - E(X_{ij})E(X_{kj}) = 0 - p_i p_k = -p_i p_k$.

/ 5.16

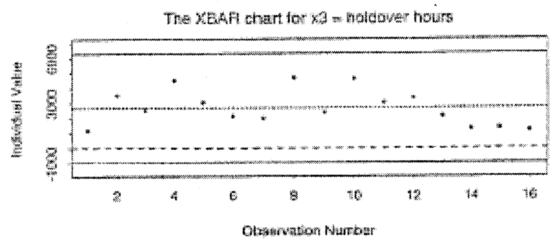
(a). Using $\hat{p}_j \pm \sqrt{\chi^2_4(0.05)}\sqrt{\hat{p}_j(1-\hat{p}_j)/n}$, the 95 % confidence intervals for p_1, p_2, p_3, p_4, p_5 are (0.221, 0.370), (0.258, 0.412), (0.098, 0.217), (0.029, 0.112), (0.084, 0.198) respectively.

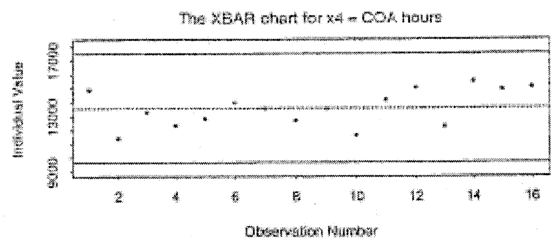
(b). Using $\hat{p}_1 - \hat{p}_2 \pm \sqrt{\chi^2_4(0.05)}\sqrt{(\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2) - 2\hat{p}_1\hat{p}_2)/n}$, the 95 % confidence interval for $p_1 - p_2$ is $(-0.118, 0.0394)$. There is no significant difference in two proportions.

Cf book is slightly larger.

5.24 Individual $\overline{X}$ charts for the Madison, Wisconsin, Police Department data

	xbar	s	LCL	UCL	
LegalOT	3557.8	606.5	1738.1	5377.4	
ExtraOT	1478.4	1182.8	-2070.0	5026.9	use LCL = 0
Holdover	2676.9	1207.7	-946.2	6300.0	use LCL = 0
COA	13563.6	1303.2	9654.0	17473.2	
MeetOT	800.0	474.0	-622.1	2222.1	use LCL = 0

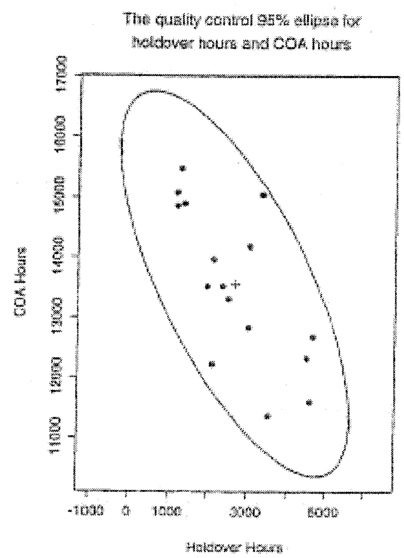


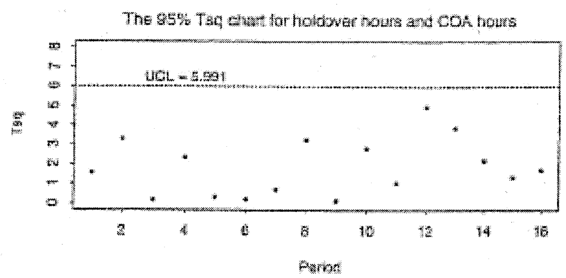


Both holdover and COA hours are stable and in control.

5.25 Quality ellipse and T^2 chart for the holdover and COA overtime hours. All points are in control. The quality control 95% ellipse is

$$1.37\times10^{-6}(x_3-2677)^2+1.18\times10^{-6}(x_4-13564)^2$$
$$+1.80\times10^{-6}(x_3-2677)(x_4-13564)=5.99.$$

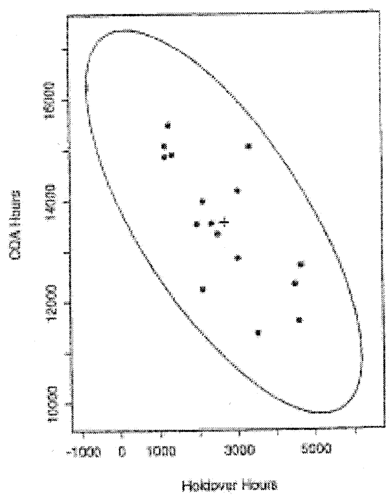




5.27 The 95% prediction ellipse for x_3 = holdover hours and x_4 = COA hours is

$$1.37\times 10^{-6}(x_3-2677)^2+1.18\times 10^{-6}(x_4-13564)^2$$
$$+1.80\times 10^{-6}(x_3-2677)(x_4-13564)=8.51.$$

The 95% control ellipse for future holdover hours and COA hours



6.8 a) For first variable:

observation	=	mean	+	treatment effect	+	residual

$$\begin{bmatrix} 6 & 5 & 8 & 4 & 7 \\ 3 & 1 & 2 & & \\ 2 & 5 & 3 & 2 & \end{bmatrix} = \begin{bmatrix} 4 & 4 & 4 & 4 & 4 \\ 4 & 4 & 4 & & \\ 4 & 4 & 4 & 4 & \end{bmatrix} + \begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ -2 & -2 & -2 & & \\ -1 & -1 & -1 & -1 & \end{bmatrix} + \begin{bmatrix} 0 & -1 & 2 & -2 & 1 \\ 1 & -1 & 0 & & \\ -1 & 2 & 0 & -1 & \end{bmatrix}$$

$SS_{obs} = 246$ $\quad SS_{mean} = 192$ $\quad SS_{tr} = 36$ $\quad SS_{res} = 18$

For second variable:

$$\begin{bmatrix} 7 & 9 & 6 & 9 & 9 \\ 3 & 6 & 3 & & \\ 3 & 1 & 1 & 3 & \end{bmatrix} = \begin{bmatrix} 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & & \\ 5 & 5 & 5 & 5 & \end{bmatrix} + \begin{bmatrix} 3 & 3 & 3 & 3 & 3 \\ -1 & -1 & -1 & & \\ -3 & -3 & -3 & -3 & \end{bmatrix} + \begin{bmatrix} -1 & 1 & -2 & 1 & 1 \\ -1 & 2 & -1 & & \\ 1 & -1 & -1 & 1 & \end{bmatrix}$$

$SS_{obs} = 402$ $\quad SS_{mean} = 300$ $\quad SS_{tr} = 84$ $\quad SS_{res} = 18$

Cross product contributions:

275 $\quad$ 240 $\quad$ 48 $\quad$ -13

b) MANOVA table:

Source of Variation	SSP	d.f.
Treatment	$B = \begin{bmatrix} 36 & 48 \\ 48 & 84 \end{bmatrix}$	$3 - 1 = 2$
Residual	$W = \begin{bmatrix} 18 & -13 \\ -13 & 18 \end{bmatrix}$	$5 + 3 + 4 - 3 = 9$
Total (corrected)	$\begin{bmatrix} 54 & 35 \\ 35 & 102 \end{bmatrix}$	11

c) $\Lambda^* = \frac{|W|}{|B+W|} = \frac{155}{4283} = .0362$

Using Table 6.3 with $p = 2$ and $g = 3$

$$\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right)\left(\frac{\Sigma n_\ell - g - 1}{g-1}\right) = 17.02 .$$

Since $F_{4,16}(.01) = 4.77$ we conclude that treatment differences exist at $\alpha = .01$ level.

Alternatively, using Bartlett's procedure,

$$-\left(n - 1 - \frac{(p+g)}{2}\right) \ln \Lambda^* = -(12 - 1 - \tfrac{5}{2})\ln(.0362) = 28.209$$

Since $\chi_4^2(.01) = 13.28$ we again conclude treatment differences exist at $\alpha = .01$ level.

6.11 $L(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \Sigma) = L(\underset{\sim}{\mu}_1, \Sigma)L(\underset{\sim}{\mu}_2, \Sigma)$

$$= \left[\frac{1}{(2\pi)^{\frac{(n_1+n_2)p}{2}} |\Sigma|^{\frac{n_1+n_2}{2}}}\right] \exp\left\{-\frac{1}{2}\left(\operatorname{tr} \Sigma^{-1}[(n_1-1)S_1 + (n_2-1)S_2]\right.\right.$$

$$\left.\left. + n_1(\bar{\underset{\sim}{x}}_1 - \underset{\sim}{\mu}_1)' \Sigma^{-1}(\bar{\underset{\sim}{x}}_1 - \underset{\sim}{\mu}_1) + n_2(\bar{\underset{\sim}{x}}_2 - \underset{\sim}{\mu}_2)' \Sigma^{-1}(\bar{\underset{\sim}{x}}_2 - \underset{\sim}{\mu}_2)\right)\right\}$$

using (4-16) and (4-17). The likelihood is maximized with respect to $\underset{\sim}{\mu}_1$ and $\underset{\sim}{\mu}_2$ at $\hat{\underset{\sim}{\mu}}_1 = \bar{\underset{\sim}{x}}_1$ and $\hat{\underset{\sim}{\mu}}_2 = \bar{\underset{\sim}{x}}_2$ respectively and with respect to Σ at

$$\hat{\Sigma} = \frac{1}{n_1+n_2}[(n_1-1)S_1 + (n_2-2)S_2] = \left(\frac{n_1+n_2-2}{n_1+n_2}\right) S_{\text{pooled}}$$

(For the maximization with respect to Σ see Result 4.10 with $b = \frac{n_1+n_2}{2}$ and $B = (n_1-1)S_1 + (n_2-2)S_2$)

6.13 a) and b) For first variable:

$$\text{Observation} = \text{mean} + \begin{matrix}\text{factor 1}\\ \text{effect}\end{matrix} + \begin{matrix}\text{factor 2}\\ \text{effect}\end{matrix} + \text{residual}$$

$$\begin{bmatrix} 6 & 4 & 8 & 2 \\ 3 & -3 & 4 & -4 \\ -3 & -4 & 3 & -4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 4 & 4 & 4 \\ -1 & -1 & -1 & -1 \\ -3 & -3 & -3 & -3 \end{bmatrix} + \begin{bmatrix} 1 & -2 & 4 & -3 \\ 1 & -2 & 4 & -3 \\ 1 & -2 & 4 & -3 \end{bmatrix} + \begin{bmatrix} 0 & 1 & -1 & 0 \\ 2 & -1 & 0 & -1 \\ -2 & 0 & 1 & 1 \end{bmatrix}$$

$SS_{tot} = 220$ $SS_{mean} = 12$ $SS_{fac\,1} = 104$ $SS_{fac\,2} = 90$ $SS_{res} = 14$

For second variable:

$$\begin{bmatrix} 8 & 6 & 12 & 6 \\ 8 & 2 & 3 & 3 \\ 2 & -5 & -3 & -6 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \end{bmatrix} + \begin{bmatrix} 5 & 5 & 5 & 5 \\ 1 & 1 & 1 & 1 \\ -6 & -6 & -6 & -6 \end{bmatrix} + \begin{bmatrix} 3 & -2 & 1 & -2 \\ 3 & -2 & 1 & -2 \\ 3 & -2 & 1 & -2 \end{bmatrix} + \begin{bmatrix} -3 & 0 & 3 & 0 \\ 1 & 0 & -2 & 1 \\ 2 & 0 & -1 & -1 \end{bmatrix}$$

$SS_{tot} = 440$ $SS_{mean} = 108$ $SS_{fac\,1} = 248$ $SS_{fac\,2} = 54$ $SS_{res} = 30$

Sum of cross products:

$$SCP_{tot} = SCP_{mean} + SCP_{fac\,1} + SCP_{fac\,2} + SCP_{res}$$

$$227 = 36 + 148 + 51 - 8$$

c) MANOVA table:

Source of Variation	SSP	d.f.
Factor 1	$\begin{bmatrix} 104 & 148 \\ 148 & 248 \end{bmatrix}$	$g-1 = 3-1 = 2$
Factor 2	$\begin{bmatrix} 90 & 51 \\ 51 & 54 \end{bmatrix}$	$b-1 = 4-1 = 3$
Residual	$\begin{bmatrix} 14 & -8 \\ -8 & 30 \end{bmatrix}$	$(g-1)(b-1) = 6$
Total (Corrected)	$\begin{bmatrix} 208 & 191 \\ 191 & 332 \end{bmatrix}$	$gb-1 = 11$

d) We reject $H_0: \underset{\sim}{\tau}_1 = \underset{\sim}{\tau}_2 = \underset{\sim}{\tau}_3 = \underset{\sim}{0}$ at $\alpha = .05$ level since

$$-\left[(g-1)(b-1) - \left(\frac{p+1-(g-1)}{2}\right)\right]\ln\Lambda^* = -\left[6 - \frac{3-2}{2}\right]\ln\left(\frac{|SS_{res}|}{|SSP_{fac\,1} + SSP_{res}|}\right)$$

$$= -5.5\ln\left(\frac{356}{13204}\right) = 19.87 > \chi^2_4(.05) = 9.49$$

and conclude there are factor 1 effects.

We also reject $H_0: \underset{\sim}{\beta}_1 = \underset{\sim}{\beta}_2 = \underset{\sim}{\beta}_3 = \underset{\sim}{\beta}_4 = \underset{\sim}{0}$ at the $\alpha = .05$ level

since

$$-[(g-1)(b-1) - (\frac{p+1-(b-1)}{2})\ell n\Lambda^* = -[6-\frac{3-3}{2}]\ell n\left(\frac{|SSP_{res}|}{|SSP_{fac\ 2}+SSP_{res}|}\right)$$

$$= -6\,\ell n\left(\frac{356}{6887}\right) = 17.77 > \chi^2_6(.05) = 12.59$$

and conclude there are factor 2 effects.

6.14 b) MANOVA Table:

Source of Variation	SSP	d.f.
Factor 1	$\begin{bmatrix} 496 & 184 \\ 184 & 208 \end{bmatrix}$	2
Factor 2	$\begin{bmatrix} 36 & 24 \\ 24 & 36 \end{bmatrix}$	3
Interaction	$\begin{bmatrix} 32 & 0 \\ 0 & 44 \end{bmatrix}$	6
Residual	$\begin{bmatrix} 312 & -84 \\ -84 & 400 \end{bmatrix}$	12
Total (Corrected)	$\begin{bmatrix} 876 & 124 \\ 124 & 688 \end{bmatrix}$	23

c) Since $-[gb(n-1)-(p+1-(g-1)(b-1))/2]\ell n\Lambda^* = -13.5\ell n\left(\frac{|SSP_{res}|}{|SSP_{int}+SSP_{res}|}\right)$

$= -13.5\ell n(.808) = 2.88 < \chi^2_{12}(.05) = 21.03$ we do not reject

$H_0: \underset{\sim}{\gamma}_{11} = \underset{\sim}{\gamma}_{12} = \cdots = \underset{\sim}{\gamma}_{34} = \underset{\sim}{0}$ (no interaction effects) at the

$\alpha = .05$ level.

Since

$$-[gb(n-1)-(p+1-(g-1))/2]\ln\Lambda^* = -11.5\ln\left(\frac{|SSP_{res}|}{|SSP_{fac\,1} + SSP_{res}|}\right)$$

$= -11.5\ln(.2447) = 16.19 > \chi_4^2(.05) = 9.49$ we <u>reject</u>

$H_0: \underset{\sim}{\tau}_1 = \underset{\sim}{\tau}_2 = \underset{\sim}{\tau}_3 = \underset{\sim}{0}$ (no factor 1 effects) at the $\alpha = .05$

level.

Since

$$-[gb(n-1)-(p+1-(b-1))/2]\ln\Lambda^* = -12\ln\left(\frac{|SSP_{res}|}{|SSP_{fac\,2} + SSP_{res}|}\right)$$

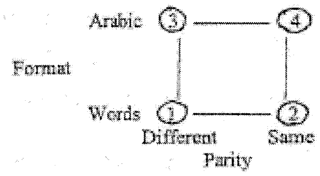
$= -12\ln(.7949) = 2.76 < \chi_6^2(.05) = 12.59$ we <u>do not reject</u>

$H_0: \underset{\sim}{\beta}_1 = \underset{\sim}{\beta}_2 = \underset{\sim}{\beta}_3 = \underset{\sim}{\beta}_4 = \underset{\sim}{0}$ (no factor 2 effects) at the

$\alpha = .05$ level.

(d)

7
6.17 a)



<u>Effects</u>	<u>Contrast</u>
Parity main:	$(\mu_2+\mu_4)-(\mu_1+\mu_3)$
Format main:	$(\mu_3+\mu_4)-(\mu_1+\mu_2)$
Interaction:	$(\mu_2+\mu_3)-(\mu_1+\mu_4)$

Contrast matrix:

$$C=\begin{pmatrix}-1 & 1 & -1 & 1\\ -1 & -1 & 1 & 1\\ -1 & 1 & 1 & -1\end{pmatrix}$$

Since $T^2 = 153.7 > \frac{31(3)}{29}(2.93) = 9.40$, reject $H_0 : C\mu = 0$ (no treatment effects) at the 5 % level.

b) 95% simultaneous T^2 intervals for the contrasts:

Parity main effect: $-206.4 \pm \sqrt{9.40}\sqrt{\frac{19,577.5}{32}} \rightarrow (-282.2, -130.6)$

Format main effect: $-307 \pm \sqrt{9.40}\sqrt{\frac{40,269.3}{32}} \rightarrow (-415.8, -198.2)$

Interaction effect: $22.4 \pm \sqrt{9.40}\sqrt{\frac{10,007.3}{32}} \rightarrow (-31.8, 76.6)$

No interaction effect. Parity effect—"different" responses slower than "same" responses. Format effect—"words" slower than "Arabic".

c) The M model of numerical cognition is supported by this experiment.

d) The multivariate normal model is a reasonable population model for the scores corresponding to the parity contrast, the format contrast and the interaction contrast.

6.24 Wilks' lambda: $\Lambda^* = .8301$. Since $g = 3$, $\left(\frac{90-4-2}{4}\right)\left(\frac{1-\sqrt{.8301}}{\sqrt{.8301}}\right) = 2.049$ is an F value with 8 and 168 degrees of freedom. Since p-value $= P(F > 2.049) = .044$, we would just reject the null hypothesis $H_0 : \underset{\sim}{\tau}_1 = \underset{\sim}{\tau}_2 = \underset{\sim}{\tau}_3 = \underset{\sim}{0}$ at the 5% level implying there is a time period effect.

F statistics and p-values for ANOVA's:

	$\underline{F}$	$\underline{p\text{-value}}$
MaxBrth:	3.66	.030
BasHght:	0.47	.629
BasLgth:	3.84	.025
NasHght:	0.10	.901

Any differences over time periods are probably due to changes in maximum breath of skull (MaxBrth) and basialveolar length of skull (BasLgth).

95% Bonferroni simultaneous intervals: $m = pg(g-1)/2 = 12$,
$t_{87}(.05/24) = 2.94$

<u>BasBrth</u>

$\tau_{11} - \tau_{21}:\ -1 \pm 2.94\sqrt{\frac{1785.4}{87}\left(\frac{1}{30}+\frac{1}{30}\right)} \rightarrow -1 \pm 3.44$

$\tau_{11} - \tau_{31}:\ -3.1 \pm 3.44$

$\tau_{21} - \tau_{31}:\ -2.1 \pm 3.44$

<u>BasHght</u>

$\tau_{12} - \tau_{22}:\ 0.9 \pm 2.94\sqrt{\frac{1924.3}{87}\left(\frac{1}{30}+\frac{1}{30}\right)} \rightarrow 0.9 \pm 3.57$

$\tau_{12} - \tau_{32}:\ -0.2 \pm 3.57$

$\tau_{22} - \tau_{32}:\ -1.1 \pm 3.57$

<u>BasLgth</u>

$\tau_{13} - \tau_{23}:\ 0.10 \pm 2.94\sqrt{\frac{2153}{87}\left(\frac{1}{30}+\frac{1}{30}\right)} \rightarrow 0.10 \pm 3.78$

$\tau_{13} - \tau_{33}:\ 3.14 \pm 3.78$

$\tau_{23} - \tau_{33}:\ 3.03 \pm 3.78$

NasHgth

$\tau_{14} - \tau_{24}:\ 0.30 \pm 2.94\sqrt{\frac{840.2}{87}\left(\frac{1}{30}+\frac{1}{30}\right)} \rightarrow 0.30 \pm 2.36$

$\tau_{14} - \tau_{34}:\ -0.03 \pm 2.36$

$\tau_{24} - \tau_{34}:\ -0.33 \pm 2.36$

All the simultaneous intervals include 0. Evidence for changes in skull size over time is marginal. If changes exist, then these changes might be in maximum breath and basialveolar length of skull from time periods 1 to 3.

The usual MANOVA assumptions appear to be satisfied for these data.

6.26 To test for parallelism, consider H_{01}: $C\underset{\sim}{\mu}_1 = C\underset{\sim}{\mu}_2$ with C given by (6-61).

$$C(\bar{\underset{\sim}{x}}_1 - \bar{\underset{\sim}{x}}_2) = \begin{bmatrix} -.413 \\ -.167 \\ -.036 \end{bmatrix}; \quad (CS_{pooles}C')^{-1} = \begin{bmatrix} 1.674 & .947 & .616 \\ & 2.014 & 1.144 \\ & & 2.341 \end{bmatrix}$$

$T^2 = 9.58 > c^2 = 8.0$, we reject H_0 at the $\alpha = .05$ level. The excess electrical usage of the test group was much lower than that of the control group for the 11 A.M., 1 P.M. and 3 P.M. hours. The similar 9 A.M. usage for the two groups contradicts the parallelism hypothesis.

6.31 (a) Two-factor MANOVA of peanuts data

E = Error SS&CP Matrix

	X1	X2	X3
X1	104.205	49.365	76.48
X2	49.365	352.105	121.995
X3	76.48	121.995	94.835

H = Type III SS&CP Matrix for FACTOR1 (Location)

	X1	X2	X3
X1	0.7008333333	-10.6575	7.1291666667
X2	-10.6575	162.0675	-108.4125
X3	7.1291666667	-108.4125	72.520833333

Manova Test Criteria and Exact F Statistics for
the Hypothesis of no Overall FACTOR1 Effect
H = Type III SS&CP Matrix for FACTOR1 E = Error SS&CP Matrix

S=1 M=0.5 N=1

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.10651620	11.1843	3	4	0.0205
Pillai's Trace	0.89348380	11.1843	3	4	0.0205
Hotelling-Lawley Trace	8.38824348	11.1843	3	4	0.0205
Roy's Greatest Root	8.38824348	11.1843	3	4	0.0205

H = Type III SS&CP Matrix for FACTOR2 (Variety)

	X1	X2	X3
X1	196.115	365.1825	42.6275
X2	365.1825	1089.015	414.655
X3	42.6275	414.655	284.10166667

Manova Test Criteria and F Approximations for
the Hypothesis of no Overall FACTOR2 Effect
H = Type III SS&CP Matrix for FACTOR2 E = Error SS&CP Matrix

S=2 M=0 N=1

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.01244417	10.6191	6	8	0.0019
Pillai's Trace	1.70910921	9.7924	6	10	0.0011
Hotelling-Lawley Trace	21.37567504	10.6878	6	6	0.0055
Roy's Greatest Root	18.18761127	30.3127	3	5	0.0012

H = Type III SS&CP Matrix for FACTOR1*FACTOR2

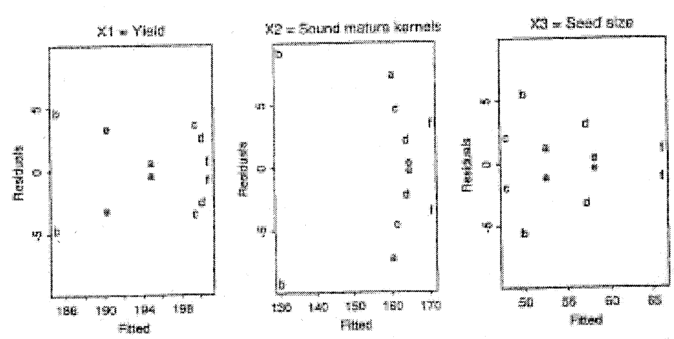
	X1	X2	X3
X1	205.10166667	363.6675	107.78583333
X2	363.6675	780.695	254.22
X3	107.78583333	254.22	85.951666667

```
Manova Test Criteria and F Approximations for
the Hypothesis of no Overall FACTOR1*FACTOR2 Effect
H = Type III SS&CP Matrix for FACTOR1*FACTOR2   E = Error SS&CP Matrix

S=2    M=0    N=1
Statistic                        Value          F     Num DF    Den DF  Pr > F
Wilks' Lambda               0.07429984     3.5582        6         8  0.0508
Pillai's Trace              1.29086073     3.0339        6        10  0.0587
Hotelling-Lawley Trace      7.54429038     3.7721        6         6  0.0655
Roy's Greatest Root         6.82409388    11.3735        3         5  0.0113
```

(b) Residual analysis. The residuals for X_2 at location 2 and variety 5 look large.

CODE	FACTOR1	FACTOR2	PRED1	RES1	PRED2	RES2	PRED3	RES3
a	1	5	194.80	0.50	160.40	-7.30	52.55	-1.15
a	1	5	194.80	-0.50	160.40	7.30	52.55	1.15
b	2	5	185.05	4.65	130.30	9.20	49.95	5.55
b	2	5	185.05	-4.65	130.30	-9.20	49.95	-5.55
c	1	6	199.45	3.55	161.40	-4.60	47.80	2.00
c	1	6	199.45	-3.55	161.40	4.60	47.80	-2.00
d	2	6	200.15	2.55	163.95	2.15	57.25	3.15
d	2	6	200.15	-2.55	163.95	-2.15	57.25	-3.15
e	1	8	190.25	3.25	164.80	-0.30	58.20	-0.40
e	1	8	190.25	-3.25	164.80	0.30	58.20	0.40
f	2	8	200.75	0.75	170.30	-3.50	66.10	-1.10
f	2	8	200.75	-0.75	170.30	3.50	66.10	1.10

Plot of residuals versus fitted values



(c) Univariate two-factor ANOVA. With $\alpha = 0.05$, from Wilk's lambda test in part (a), the interaction term can be dropped.

```
Dependent Variable: X1
                 R-Square          C.V.        Root MSE                 X1 Mean
                 0.388870      3.187484         6.21798                 195.075

Source                 DF   Type III SS     Mean Square    F Value     Pr > F
FACTOR1                 1      0.700833        0.700833       0.02     0.8962
FACTOR2                 2    196.115000       98.057500       2.54     0.1403

Dependent Variable: X2
                 R-Square          C.V.        Root MSE                 X2 Mean
                 0.524809      7.506437         11.8996                 158.525

Source                 DF   Type III SS     Mean Square    F Value     Pr > F
FACTOR1                 1     162.06750       162.06750       1.14     0.3159
FACTOR2                 2    1089.01500       544.50750       3.85     0.0676

Dependent Variable: X3
                 R-Square          C.V.        Root MSE                 X3 Mean
                 0.663596      8.595035         4.75377                 55.3083

Source                 DF   Type III SS     Mean Square    F Value     Pr > F
FACTOR1                 1     72.520833       72.520833       3.21     0.1110
FACTOR2                 2    284.101667      142.050833       6.29     0.0229
```

(d) Bonferroni simultaneous comparisons of variety.
Only varieties 5 and 8 differ, and they differ only on X_3.

```
Bonferroni (Dunn) T tests for variable: X1
Alpha= 0.05  Confidence= 0.95  df= 8  MSE= 38.66333
Critical Value of T= 3.01576
Minimum Significant Difference= 13.26
Comparisons significant at the 0.05 level are indicated by '***'.

               Simultaneous                Simultaneous
                   Lower     Difference       Upper
   FACTOR2      Confidence    Between      Confidence
  Comparison       Limit       Means          Limit
6     - 8         -8.960       4.300         17.560
6     - 5         -3.385       9.875         23.135
8     - 6        -17.560      -4.300          8.960
8     - 5         -7.685       5.575         18.835
5     - 6        -23.135      -9.875          3.385
5     - 8        -18.835      -5.575          7.685
```

```
Bonferroni (Dunn) T tests for variable: X2
Alpha= 0.05  Confidence= 0.95  df= 8  MSE= 141.6
Critical Value of T= 3.01576
Minimum Significant Difference= 25.375
Comparisons significant at the 0.05 level are indicated by '***'.
```

FACTOR2 Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
8 - 6	-20.500	4.875	30.250	
8 - 5	-3.175	22.200	47.575	
6 - 8	-30.250	-4.875	20.500	
6 - 5	-8.050	17.325	42.700	
5 - 8	-47.575	-22.200	3.175	
5 - 6	-42.700	-17.325	8.050	

```
Bonferroni (Dunn) T tests for variable: X3
Alpha= 0.05  Confidence= 0.95  df= 8  MSE= 22.59833
Critical Value of T= 3.01576
Minimum Significant Difference= 10.137
Comparisons significant at the 0.05 level are indicated by '***'.
```

FACTOR2 Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
8 - 6	-0.512	9.625	19.762	
8 - 5	0.763	10.900	21.037	***
6 - 8	-19.762	-9.625	0.512	
6 - 5	-8.862	1.275	11.412	
5 - 8	-21.037	-10.900	-0.763	***
5 - 6	-11.412	-1.275	8.862	

6.35 Fitting a quadratic growth curve to calcium measurements on the dominant ulna, treating all 31 subjects as a single group.

```
XBAR              MLE of beta          [B'Sp^(-1)B]^(-1)
70.7839           71.6039              92.2789 -5.9783  0.0799
71.9323            3.8673              -5.9783  9.3020 -2.9033
71.8065           -1.9404               0.0799 -2.9033  1.0760
64.6548

S                                      W = (n-1)*S
94.5441 90.7962 80.0081 78.0676        2836.322 2723.886 2400.243 2342.027
90.7962 93.6616 78.9965 77.7725        2723.886 2809.848 2369.894 2333.175
80.0081 78.9965 77.1546 70.0366        2400.243 2369.894 2314.639 2101.099
78.0676 77.7725 70.0366 75.9319        2342.027 2333.175 2101.099 2277.957

Estimated covariance matrix            W2
 3.1894 -0.2066  0.0028                2857.167 2764.522 2394.410 2369.674
-0.2066  0.3215 -0.1003                2764.522 2889.063 2358.522 2387.070
 0.0028 -0.1003  0.0372                2394.410 2358.522 2316.271 2093.362
                                       2369.674 2387.070 2093.362 2314.625

Lambda = |W|/|W2| = 0.7653
```

Since, with $\alpha = 0.01$, $-\left[n - \frac{1}{2}(p - q + 1)\right]\log(\Lambda) = 7.893 > \chi^2_{4-2-1}(0.01) = 6.635$, we reject the null hypothesis of a quadratic fit at $\alpha = 0.01$.

Multivariate HW

JAIME FRADE ①

19

6.1

joint 95% confidence region

$$\bar{d} = \begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix} \qquad S_d = \begin{bmatrix} 199.26 & 88.38 \\ 88.38 & 418.61 \end{bmatrix} \qquad \delta' = [0,\ 0]$$

$$(\bar{d} - \delta)' S_d^{-1} (\bar{d} - \delta) = 1.24$$

$\alpha = 0.05$ $\qquad n = 11 \quad p = 2$

$$\frac{(n-1)p}{(n-p)} F_{p,n-p,}(0.05)$$

$$\frac{(10)(2)}{(11)(9)} F_{2,9}(0.05) \sim \frac{20}{(11)(9)}(4.26) = 0.8606$$

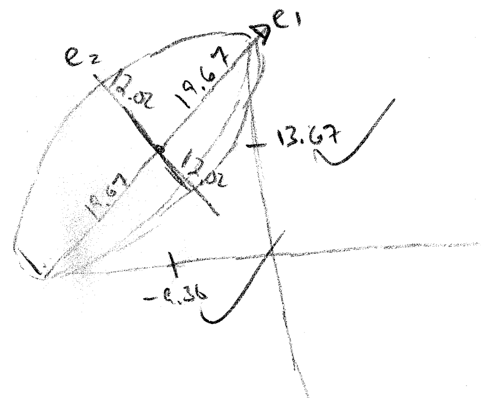
The result is consistent

$\lambda_1 = 449.78822$ $\qquad e_1' = [0.3326,\ 0.94303]$

$\lambda_2 = 168.08178$ $\qquad e_2' = [0.94303,\ -0.33268]$

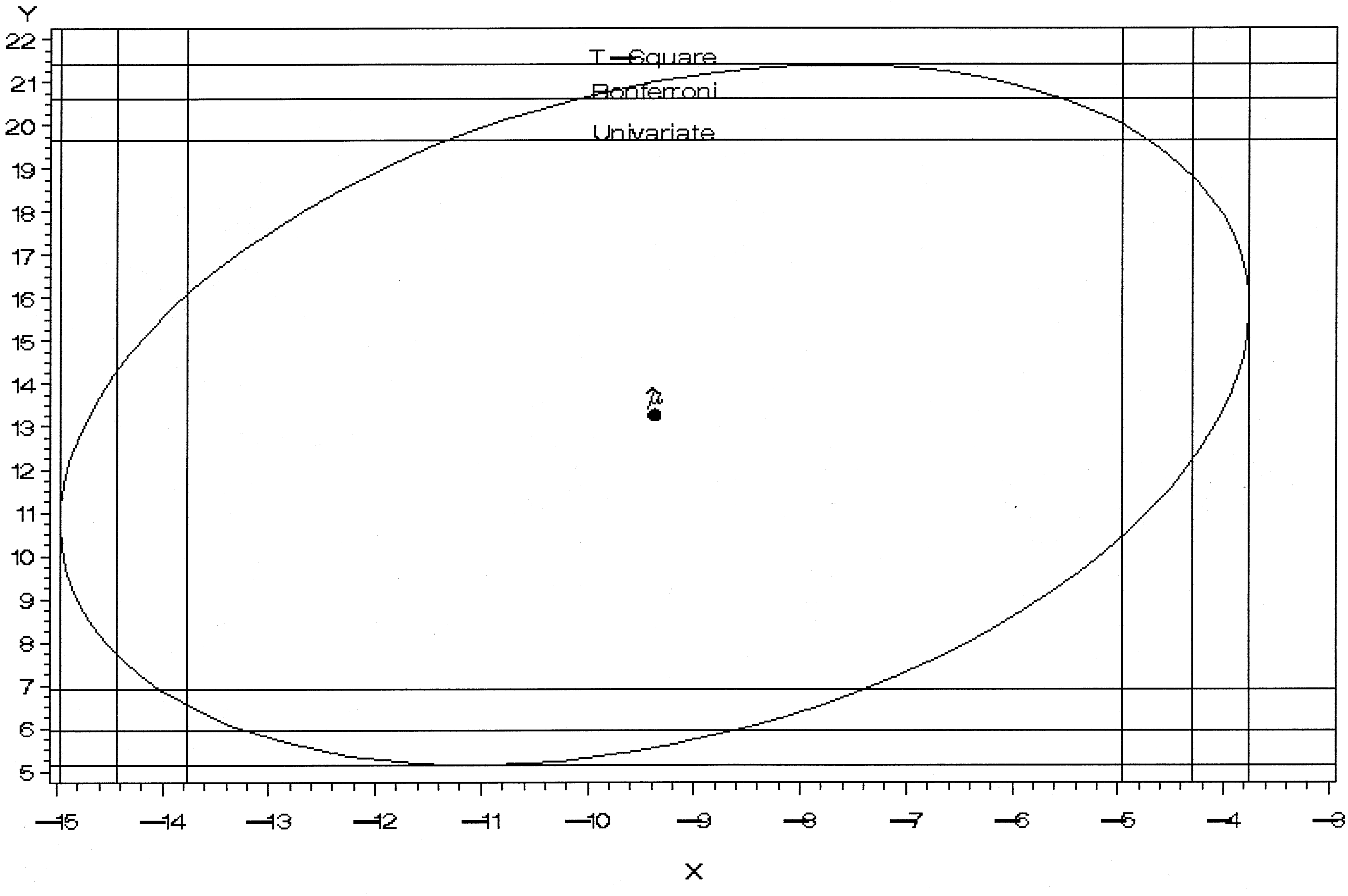
major axis $\sqrt{449.78}\sqrt{0.8606} = 19.67$

minor axis $\sqrt{168.08}\sqrt{0.8606} = 12.02$

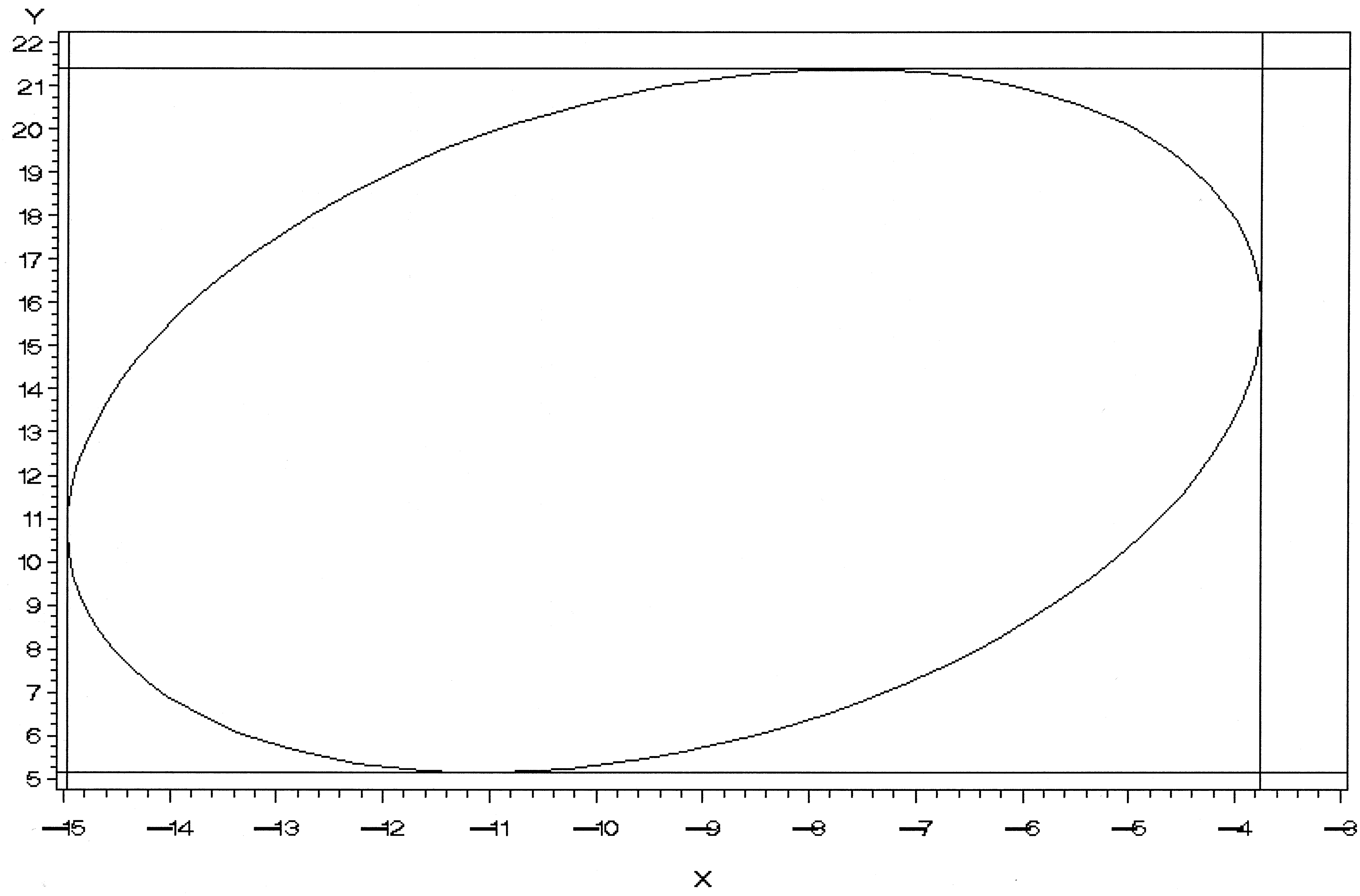


Reject H_0

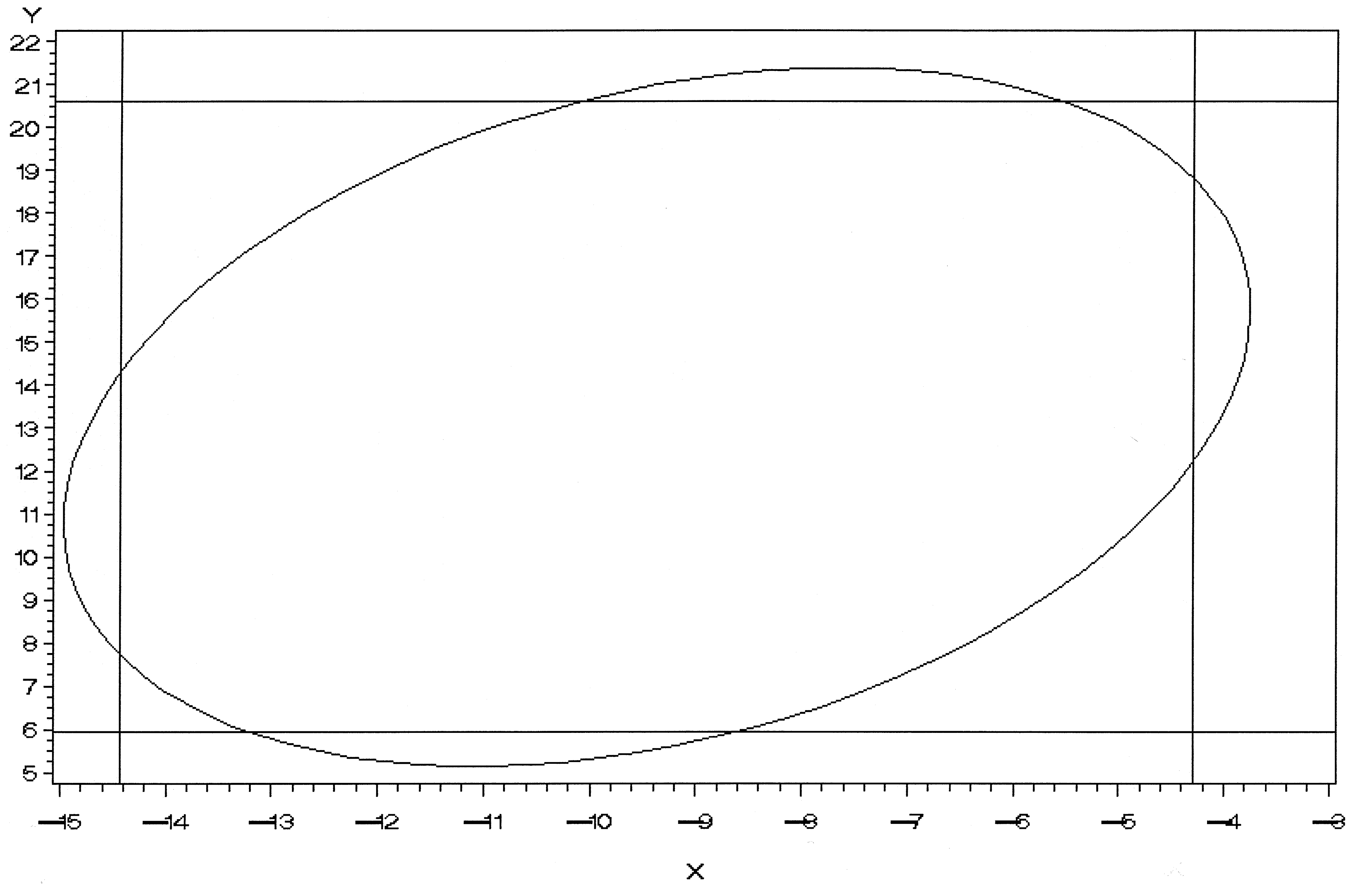
Bivariate Normal 95% Confidence Ellipse
With Various 95% Univariate Confidence Intervals



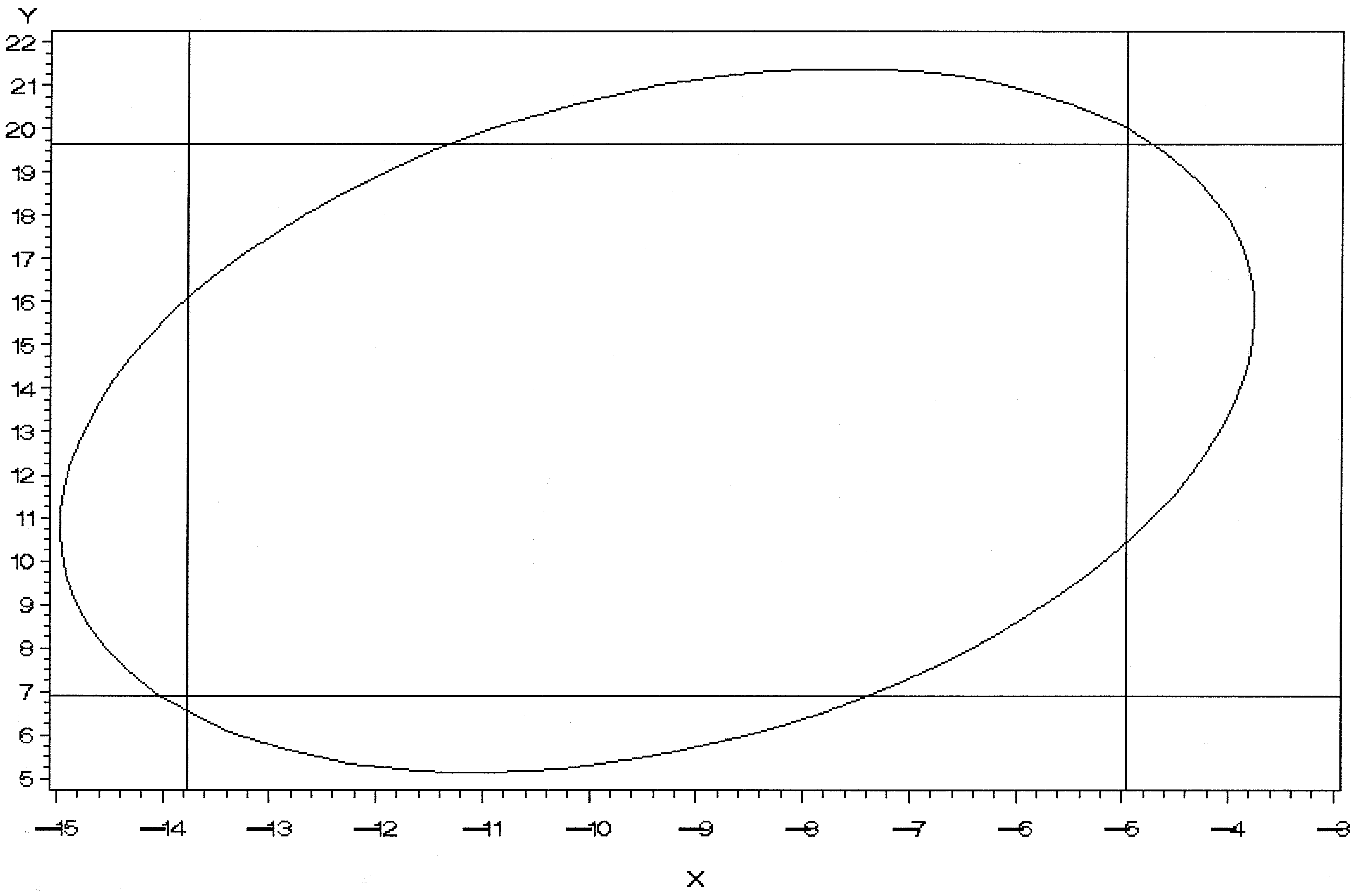
Bivariate Normal 95% Confidence Ellipse
With 95% T-Square Confidence Intervals



Bivariate Normal 95% Confidence Ellipse
With 95% Bonferroni Confidence Intervals



Bivariate Normal 95% Confidence Ellipse
With Usual 95% Univariate Confidence Intervals



6.3

FRADR ②

$$\bar{d} = \begin{bmatrix} -12 \\ 8.6 \end{bmatrix} \qquad S_d = \begin{bmatrix} 136.44 & -52.44 \\ -52.44 & 198.26 \end{bmatrix}$$

$$(\bar{d} - \delta)' S_d^{-1} (\bar{d} - \delta) = \cancel{1.8}$$

sample 8 removed

$n = \cancel{11}$ 10 $p = 2$

$$\frac{(n-1)p}{(n-p)} F_{p,n-p}(0.05) = 0.\cancel{x} \quad 1.004$$ So Reject

$\delta_1 = (-16.63, \cancel{-7}.3607)$

$\delta_2 = (3.007, \cancel{14}.192)$

Bonferroni CI

$(-16.1931, -7.806)$ × $(-22, -2)$

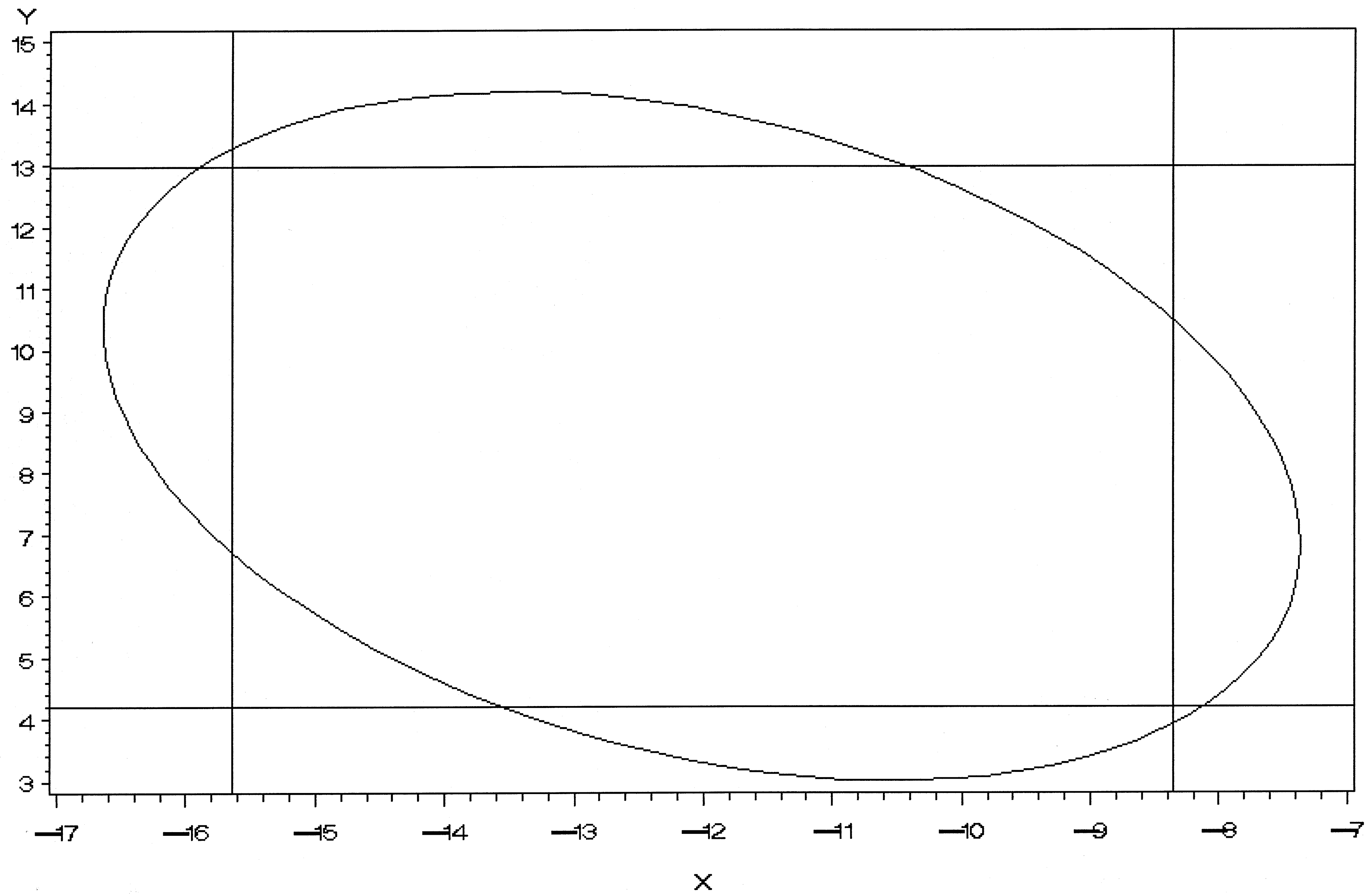
and

$(3.5447, 13.6552)$ × $(-3.4, 20.6)$

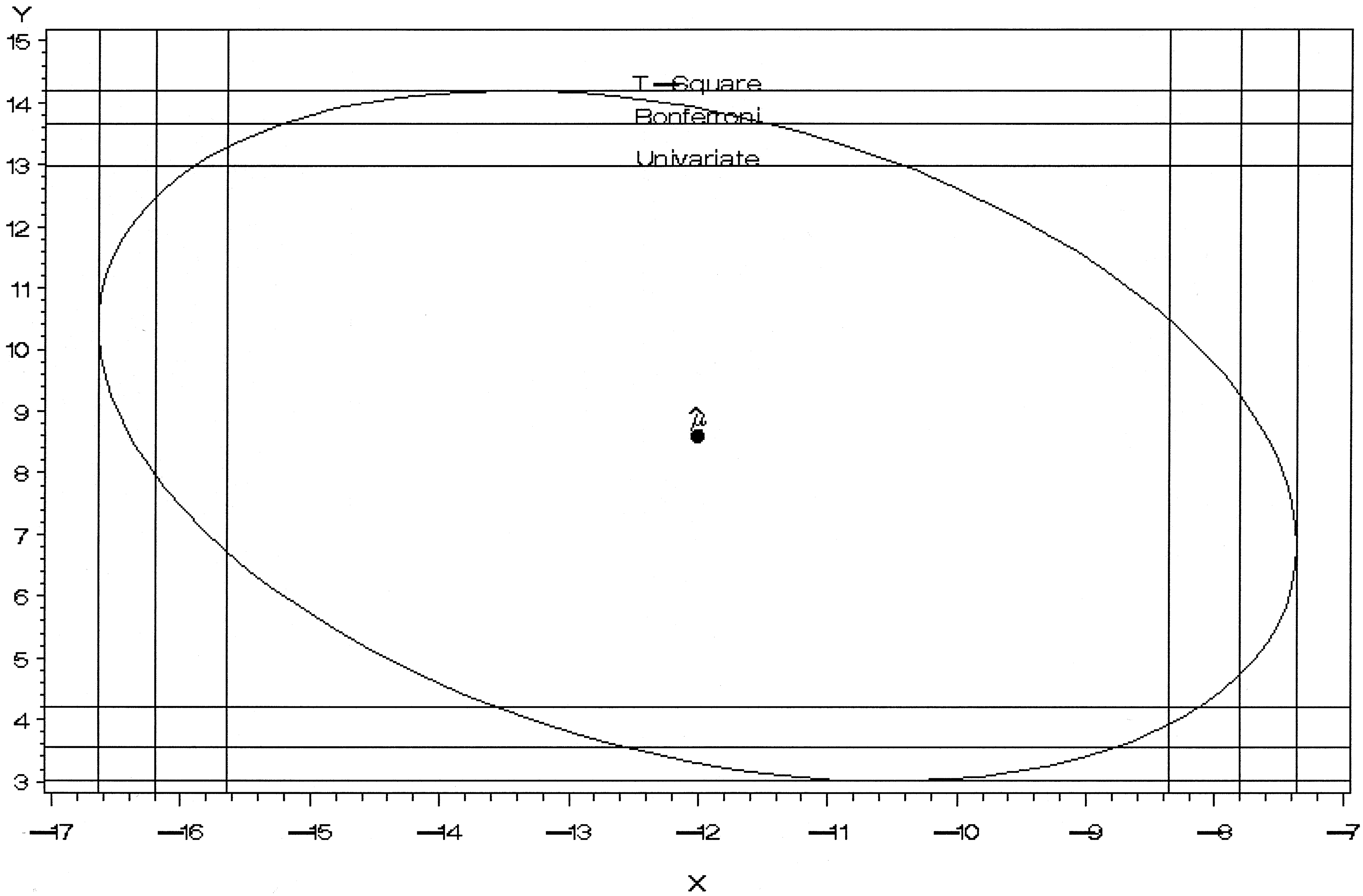
Yes, outlier makes difference

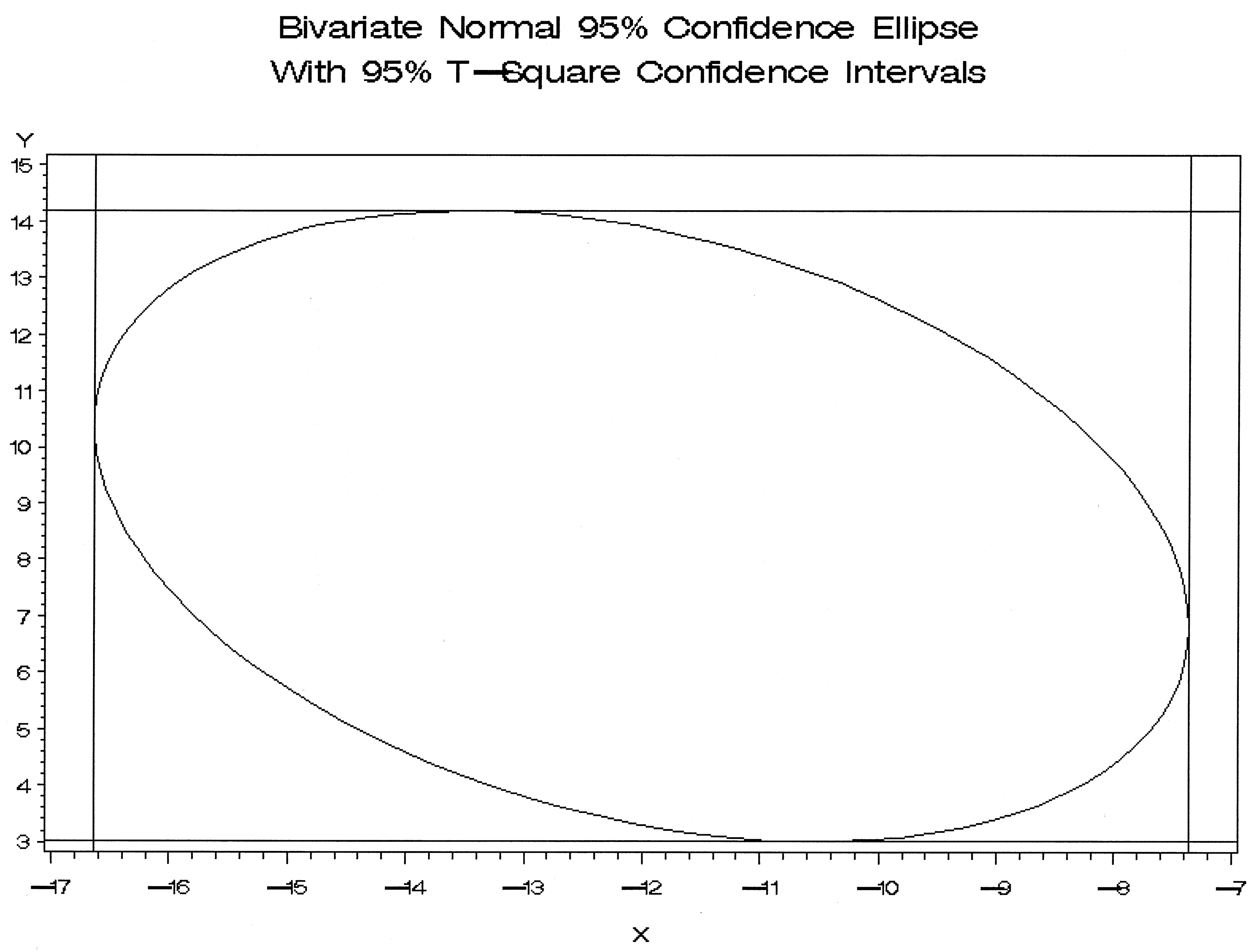
Reject $H_0: \delta = 0$ b/c CI doesn't include zero.

Bivariate Normal 95% Confidence Ellipse
With Usual 95% Univariate Confidence Intervals

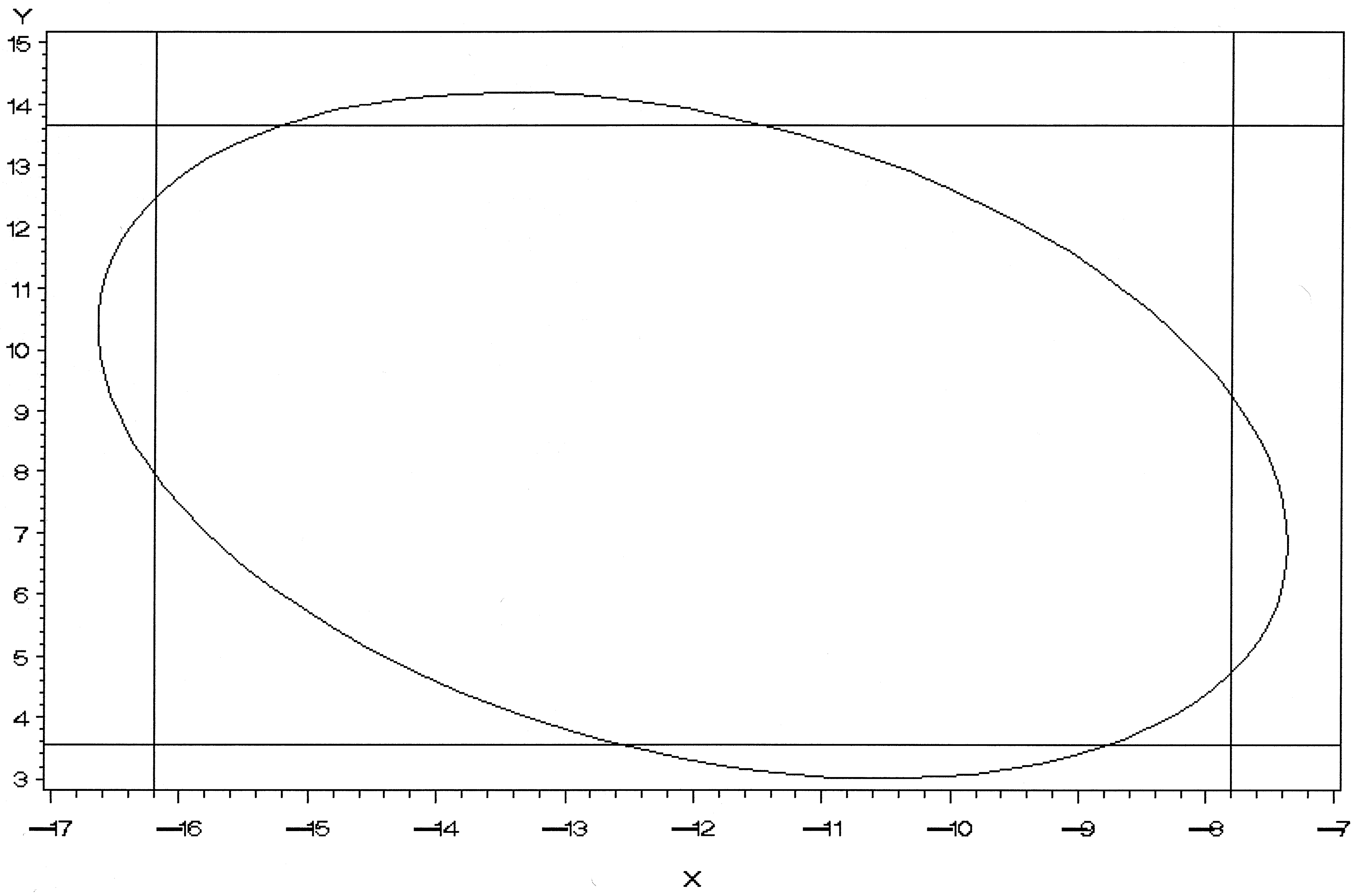


Bivariate Normal 95% Confidence Ellipse With Various 95% Univariate Confidence Intervals




Bivariate Normal 95% Confidence Ellipse
With 95% T-Square Confidence Intervals
Y
15
14
13
12
11
10
9
8
7
6
5
4
3
-17
-16
-15
-14
-13
-12
-11
-10
-9
-8
-7
X

Bivariate Normal 95% Confidence Ellipse With 95% Bonferroni Confidence Intervals



6.4

$$\bar{d} = \begin{bmatrix} -0.588 \\ 0.226 \end{bmatrix} \qquad S_{\bar{d}} = \begin{bmatrix} 0.525944 & 1.296265 \\ 1.296265 & 0.548775 \end{bmatrix}$$

$$(\bar{d} - \delta)' S_{d}^{-1} (\bar{d} - \delta) = 10.21$$

$T^2 = 6.508$

$$10.27 > 9.47$$

Reject $H_0: \delta = 0$

Simultaneous confidence intervals

$(1.85, 2.42) \times (-1.18, 0.07)$

$(2.566, 3.155) \times (-0.1, 0.69)$

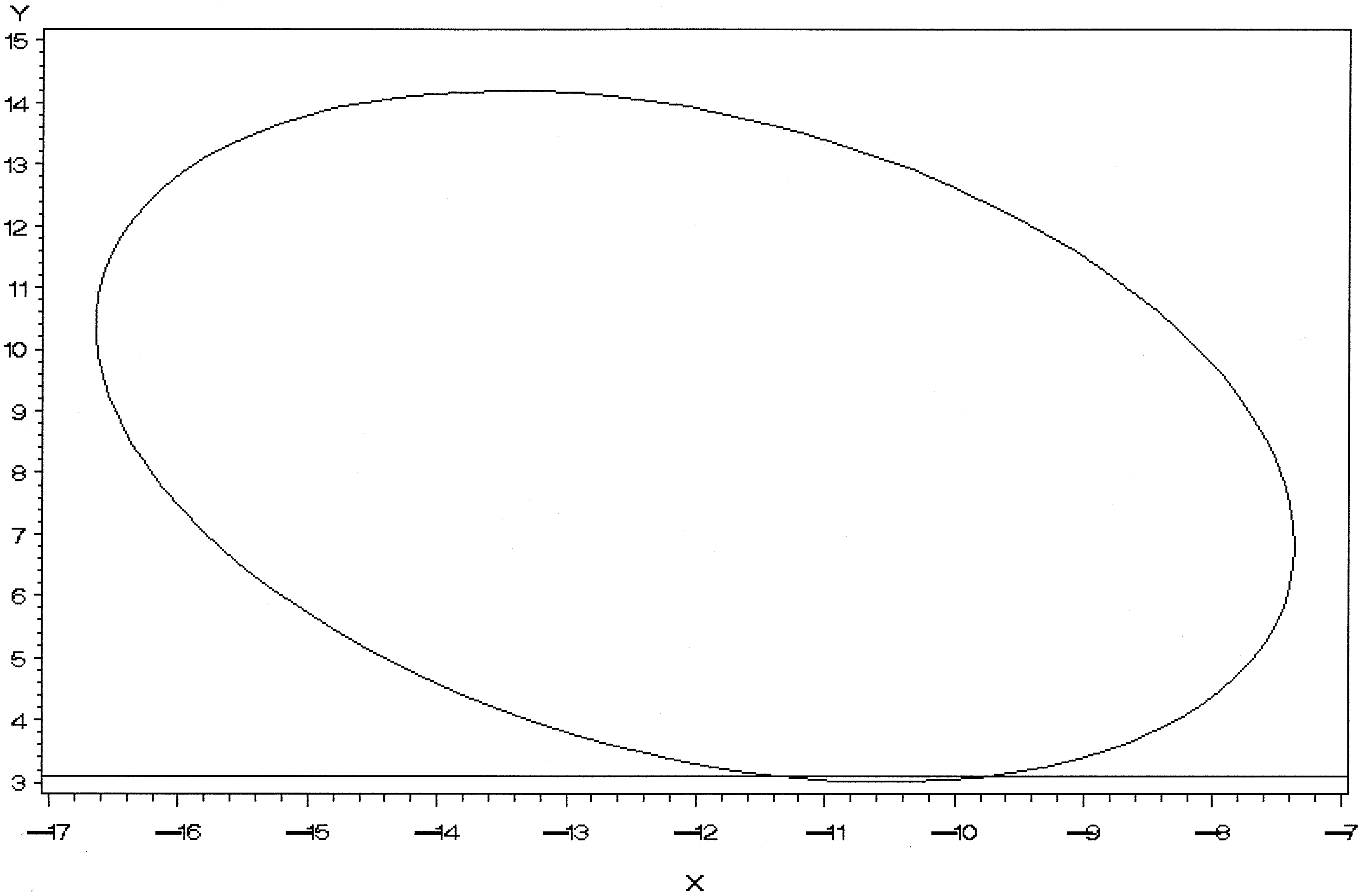
Bonferroni

$(1.87, 2.40)$

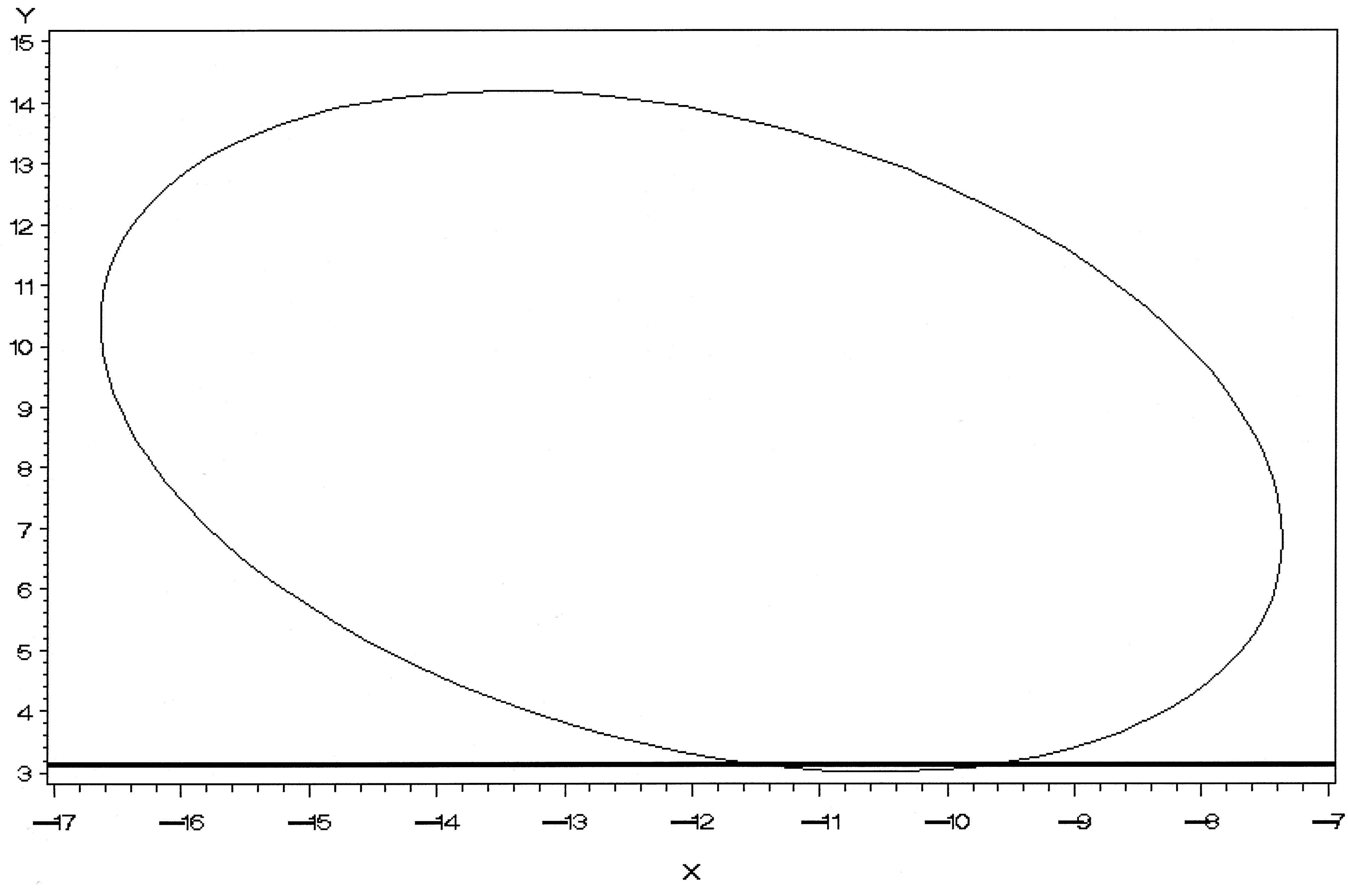
$(2.59, 3.12)$

Reject Ho

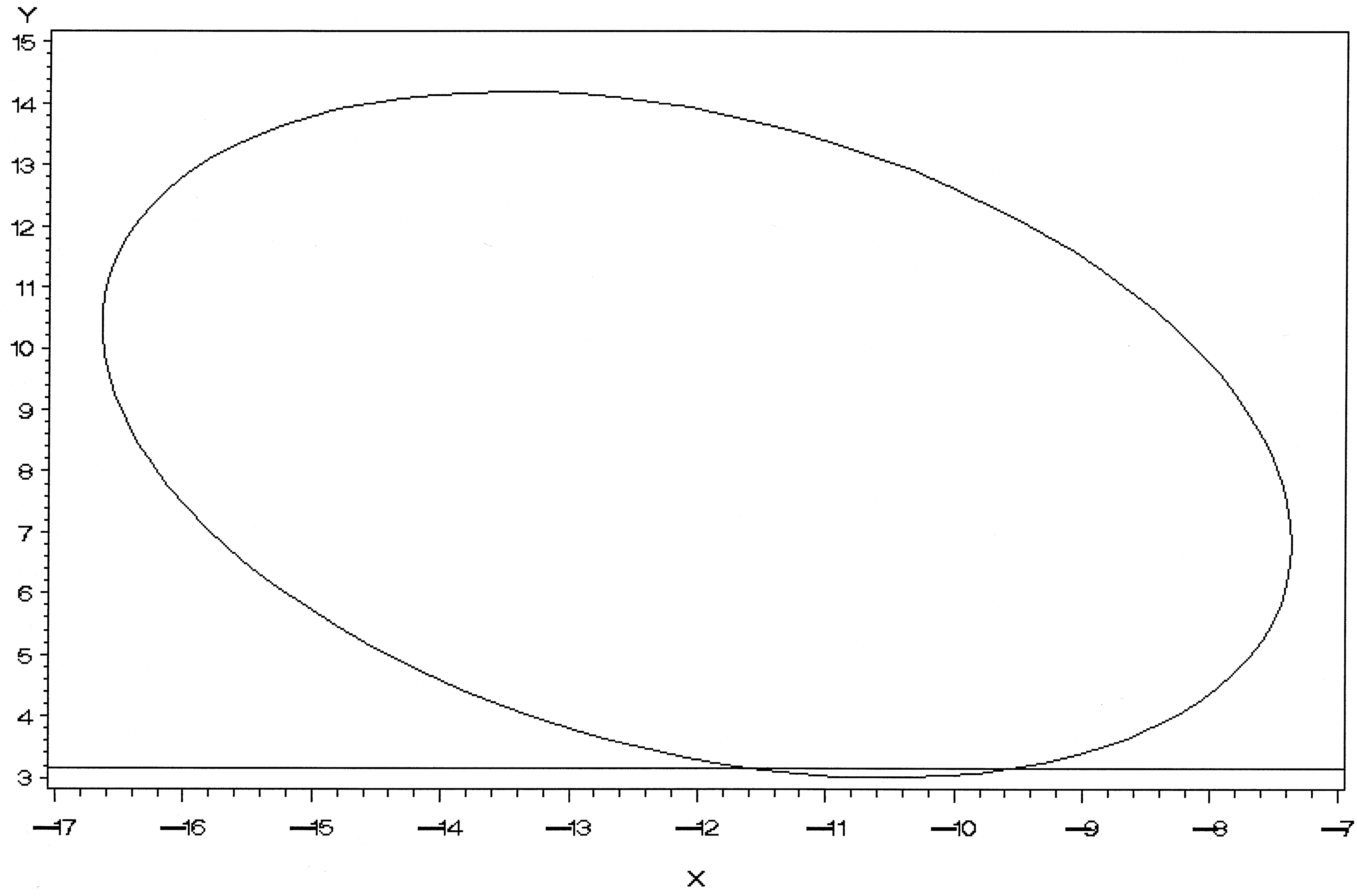
Bivariate Normal 95% Confidence Ellipse
With Usual 95% Univariate Confidence Intervals



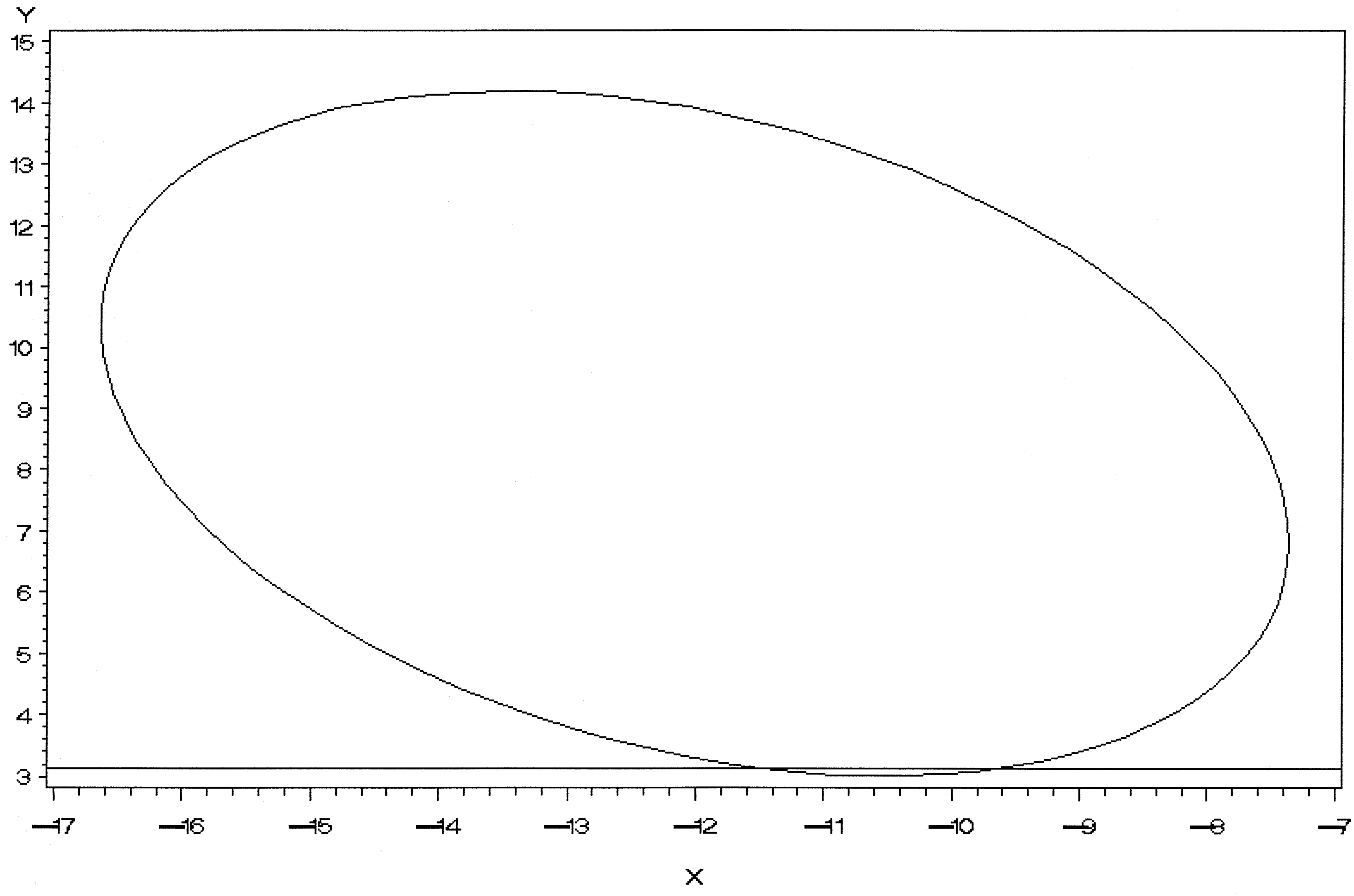
Bivariate Normal 95% Confidence Ellipse
With Various 95% Univariate Confidence Intervals



Bivariate Normal 95% Confidence Ellipse
With 95% T-Square Confidence Intervals



Bivariate Normal 95% Confidence Ellipse
With 95% Bonferroni Confidence Intervals



6.6 a)

Treatment 2

Sample mean vector $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$

Sample covariance $\begin{bmatrix} 1 & -3/2 \\ -3/2 & 3 \end{bmatrix}$

Treatment 3

sample mean vector $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$

Sample covariance $\begin{bmatrix} 2 & -4/3 \\ -4/3 & 4/3 \end{bmatrix}$

Spooled = $\begin{bmatrix} 1.6 & -1.4 \\ -1.4 & 2 \end{bmatrix}$