Exam Review

if rejected

1

$n = 4$

bivariate normal population

$\mu_0 = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$ $\quad X = \begin{pmatrix} 6 & 10 & 8 & 8 \\ 9 & 6 & 3 & 2 \end{pmatrix}$ $\quad$ data matrix $= \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \\ 8 & 2 \end{bmatrix}$

d

a) what is sampling distrib of T^2?

$$T^2 = n(\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

$H_0: \mu' = [9,5]$ $\quad \alpha = 0.05$

$H_a: \mu' \neq [9,5]$

$$\frac{(3)(2)}{2} F_{2,2}(0.05) = 3(19) = 57$$

b) calculate T^2

$$\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix} = \begin{bmatrix} \frac{6+10+8+8}{4} \\ \frac{9+6+3+2}{4} \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \end{bmatrix}$$

$$S = \frac{\sum (x - \bar{x})^2}{n-1}$$

$$S_{11} = \frac{(6-8)^2 + (10-8)^2 + (8-8)^2 + \cancel{(8-8)^2}}{3} = \frac{8}{3}$$

$$S_{12} = \frac{(6-8)(9-5) + (10-8)(6-5) + (8-8)(3-5) + \cancel{(8-8)}(2-5)}{3}$$

$S_{12} = \frac{-6}{3}$

$$S_{22} = \frac{(9-5)^2 + (6-5)^2 + (3-5)^2 + (2-5)^2}{3} = \frac{30}{3} = 10$$

Since T^2 $1.76 < 57$

fail to reject null hypot. conclude the test for equality of means vector is Not equal

$$S = \begin{bmatrix} 8/3 & -2 \\ -2 & 10 \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} 0.441 & 0.088 \\ 0.088 & 0.11767 \end{bmatrix}$$

$$T^2 = 4(8-9, 5-5) \begin{bmatrix} 0.441 & 0.088 \\ 0.088 & 0.117 \end{bmatrix} \begin{bmatrix} 8-5 \\ 5-5 \end{bmatrix} = 1.76 \cancel{>} 57$$

Exam Review

③ obtain 95% <u>confidence</u> <u>ellipse</u> for mean ? in wt from 2-3 yrs and 4-5 yrs

$$\bar{x} = \begin{bmatrix} 15.85 \\ 12.57 \end{bmatrix} \qquad S = \begin{bmatrix} 294.809 & 308.59 \\ 308.59 & 605.95 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 0.007 & -0.037 \\ -0.037 & 0.0035945 \end{bmatrix}$$

$n \to$ pairs of observation

eigenvalues and eigenvectors of S

$$\frac{\sqrt{\lambda_i}\,c}{\sqrt{n}} = \sqrt{\lambda_i}\sqrt{c}$$

$\lambda_1 = 795.972$ $\qquad e_1' = [0.52432,\ 0.8515]$

$\lambda_2 = 104.784$ $\qquad e_2' = [0.8515,\ -0.52432]$

$$n(\bar{x}-\mu)' S^{-1} (x-\mu) \le c^2 = \frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)$$

95% confidence ellipse $n=7$

$$7\left[15.85-\mu_1,\ 12.57-\mu_2\right] \begin{bmatrix} 0.007 & -0.037 \\ -0.037 & 0.0035 \end{bmatrix} \begin{bmatrix} 15.85-\mu_1 \\ 12.57-\mu_2 \end{bmatrix}$$

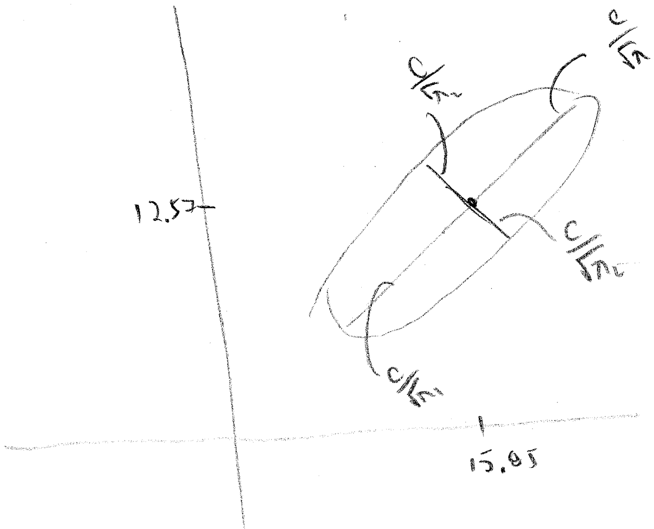
$$\le \frac{(2)(6)}{5} \underbrace{F_{2,5}(0.05)}_{=5.79}$$

$$\le 13.896$$

to ✓ if given mean is inside plug in μ_1 and μ_2

(3)

positive correlation



halflengths

major axis $105.17 = \sqrt{\lambda_1}\sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)} = \sqrt{795.97}\sqrt{13.996}$

minor axis

$38.15 = \quad = \sqrt{104.78}\sqrt{13.986}$

elongation is given by $= \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} = 2.75619$

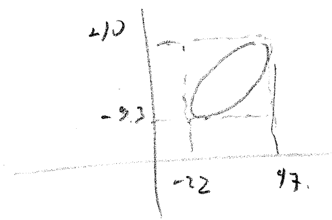
length of major axis is 2.77 times length of minor

semi(axis)

95%

a) T^2 confidence ellipse

$$\left(\bar{X}_1 \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{S_{11}}{n}} \right)$$



$$\left(\bar{X}_2 \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{S_{22}}{n}} \right)$$

$$15.85 \pm \sqrt{13.896} \sqrt{\frac{294.809}{7}} = (-8.329, 40.04)$$

$$12.57 \pm \sqrt{13.896} \sqrt{\frac{605.94}{7}} = (-22.099, 47.247)$$

$n=7$
$p=2$ $\quad t_6\left(\frac{0.05}{4}\right) = \text{tinv}((0.025 \times 2), 6) = 2.97$

95% T based

b) Bonferroni CI

$$\bar{X}_1 \pm t_{n-1}\left(\frac{\alpha}{2p}\right) \sqrt{\frac{S_{11}}{n}}$$

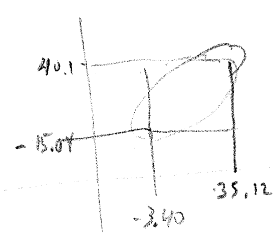
$$15.85 \pm 2.97 \sqrt{\frac{294}{7}}$$

$(-3.40, 35.12)$

$$\bar{X}_2 \pm t_{n-1}\left(\frac{\alpha}{2p}\right) \sqrt{\frac{S_{22}}{n}}$$

$$12.57 \pm 2.97 \sqrt{\frac{605.94}{7}}$$

$(-15.049, 40.1921)$



$p = 2$ # of diff. column (5)

$n = 12$ # times repeated.

2-factor MANOVA

Spec. g = levels of factor $1 = 3$

time b = levels of factor $2 = 3$

To test interaction compute

df int $= (b-1)(g-1) = 4$

Λ^* = wilks lambda of interaction $= 0.0870$

For $(g-1)(b-1) = 1$

$$F = \left(\frac{1-\lambda^*}{\lambda^*}\right) \cdot \frac{\frac{(gb(n-1)-p+1)}{2}}{\frac{|(g-1)(b-1)-p|+1}{2}} =$$

F_{v_1, v_2}

$3 = v_1 = |(g-1)(b-1)-p|+1$

$26 = v_2 = gb(n-1)-p+1$

$F_{3,26}(0.05) = 2.98$

$$F = \frac{1-0.0870}{0.0870}\left(\frac{\frac{(3)(3)(11)-2+1}{2}}{3/2}\right)$$

> 2.98 reject Null hyp

$\exists$ interaction effect

Test each factor same way

Spec.

$H_0 \quad \tau_1 = \tau_2 = 0$

Time

$H_0 \quad \beta_1 = \beta_2 = 0$

(4) find porportion of total pop variance explained by 1st component

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix} \qquad p=4$$

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

$$\begin{vmatrix} 1-\lambda & \rho & \rho & \rho \\ \rho & 1-\lambda & \rho & \rho \\ \rho & \rho & 1-\lambda & \rho \\ \rho & \rho & \rho & 1-\lambda \end{vmatrix} = 0 \rightarrow \begin{vmatrix} 1-\lambda+3\rho & \rho & \rho & \rho \\ \vdots & 1-\lambda & \rho & \rho \\ \vdots & \rho & 1-\lambda & \rho \\ \vdots & \rho & \rho & 1-\lambda \end{vmatrix} = 0$$

$$(1-\lambda+3\rho)(1-\lambda-\rho)^3 = 0$$

$$\lambda_1 = 1+3\rho \qquad \lambda_2 = \lambda_3 = \lambda_4 = 1-\rho$$

if $\rho > 0$ then

$1+3\rho > 1-\rho$

Total of variance explained by $\dfrac{1+3\rho}{1+3\rho+3(1-\rho)} = \dfrac{1+3\rho}{4}$

if $\rho < 0$ then

$1-\rho > 1+3\rho$

$$\rightarrow \frac{3(1-\rho)}{1+3\rho+3(1-\rho)} = \frac{3-3\rho}{4}$$

pg 440–441

Ex 8.1/2 obtaining from Σ

8.6 " " from ρ

8.1 $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

suppose to be positive

$\lambda_1 = 5.828$ $\quad e_1' = \begin{bmatrix} -0.382683 & 0 & 0.92388 \\ 0.9238789 & 0 & 0.38269 \\ 0 & 1 & 0 \end{bmatrix}$

$\lambda_2 = 2$

$\lambda_3 = 0.17$

$\qquad\qquad\qquad\qquad e_1^T \qquad e_2^T \qquad e_3^T$

$$\rho_{Y_1,X_1} = \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}}$$

$$= \frac{0.383\sqrt{5.828}}{\sqrt{1}} = 0.925$$

$$\rho_{Y_1,X_2} \quad \frac{e_{12}\sqrt{\lambda_2}}{\sqrt{\sigma_{22}}} = \frac{-0.923\sqrt{2}}{\sqrt{5}} = -0.998$$

$$\rho_{Y_2X_1} = \rho_{Y_2X_2} = 0$$

$$\rho_{Y_2X_3} = \frac{\sqrt{\lambda_2}}{\sqrt{\sigma_{33}}} = \frac{\sqrt{2}}{\sqrt{2}} = 1 \leftarrow \text{yes}$$

8.2 $\Sigma = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix}$ $\rho = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$

① $\lambda_1 = 100.16$ $e_1' = [0.40, 0.999]$

$\lambda_2 = 0.84$ $e_2' = [0.999, -0.40]$

$(1 + (1-p)\rho)$

② $\lambda_1 = 1 + \rho = 1.4$ $e_1' = [0.707, 0.707]$

$1-\lambda$ $\lambda_2 = 1 - \rho = 0.6$ $e_2 = [0.707, -0.707]$

respective pc are

$\Sigma =$ $Y_1 = 0.40 X_1 + 0.999 X_2$

$Y_2 = 0.999 X_1 - 0.40 X_2$

$\rho:$ $Y_1 = 0.707 Z_1 + 0.707 Z_2 = 0.707\left(\frac{X_1 - \mu}{\sqrt{S_{11}}}\right) + 0.707\left(\frac{X_2 - \mu}{\underset{=10}{\sqrt{S_{22}}}}\right)$

$Y_2 = 0.707 Z_1 - 0.707 Z_2$

1st principal component $\rightarrow$ $\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{100.16}{101} = 0.992$

if want 1st prin comp of total (standarized) pop variance

$\frac{\lambda_1}{p} = \frac{1.4}{2} = 0.7$

5.1 a) find T^2

$$X = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$$

$$T^2 = n(\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0)$$

$n = 4 \quad p = 2$

$$\mu_0' = [7, 11]$$

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$$

$$s_{11} = \frac{(2-6)^2 + (8-6)^2 + (6-6)^2 + (8-6)^2}{3} = 8$$

$$s_{22} = \frac{(12-10)^2 + (9-10)^2 + (9-10)^2 + (10-10)^2}{3} = 2$$

$$s_{12} = \frac{(2-6)(12-10) + (8-6)(9-10) + (6-6)(9-10) + (8-6)(10-10)}{3} = \frac{-10}{3}$$

$$S = \begin{bmatrix} 8 & \frac{-10}{3} \\ \frac{-10}{3} & 2 \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} 9/22 & 15/22 \\ 15/22 & 18/11 \end{bmatrix}$$

$$T^2 = 4 \, [6-7, 10-11] \begin{bmatrix} 9/22 & 15/22 \\ 15/22 & 18/11 \end{bmatrix} \begin{bmatrix} 6-7 \\ 10-11 \end{bmatrix}$$

$$T^2 = 13.64$$

5.1

$n=4$

$p=2$

b) Specify dist. of T^2

$$\frac{(n-1)p}{n-p} F_{p,n-p} = \frac{6}{2} F_{2,2}$$

c) $\alpha = 0.05$

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

$T^2 = 13.64 < 3(19)$ do not reject H_0

5.2 verify T^2 remains unchanged if by Cx_j

$$C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \qquad X = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}$$

$$CX' = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 6 & 10 & 8 \\ 9 & 6 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 6-9 & 10-6 & 8-3 \\ 6+9 & 10+6 & 8+3 \end{bmatrix} = \begin{bmatrix} -3 & 4 & 5 \\ 15 & 16 & 11 \end{bmatrix}$$

$$X = CX \sim \text{New data matrix} = \begin{bmatrix} -3 & 15 \\ 4 & 16 \\ 5 & 11 \end{bmatrix}$$

Pg 216

if $\bar{y} = C\bar{x}$, $S_y = CSC'$ $\mu_y = C\mu$ $\mu_{y,0} = C\mu_0$

$$T^2 = n(\bar{y} - \mu_{y_0})' S_y^{-1} (\bar{y} - \mu_{y,0}) = 4 \begin{bmatrix} 2-4, 14-14 \end{bmatrix} \begin{bmatrix} 7/108 & 5/108 \\ 5/108 & 19/108 \end{bmatrix} \begin{bmatrix} 2-4 \\ 14-14 \end{bmatrix}$$

$$T^2 = 7/9$$

from original

$$\bar{x} = \begin{bmatrix} 8 \\ 6 \end{bmatrix} \qquad S = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix} \qquad \mu_0 = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$$

shows T is invariant under transformation C_x

$$\bar{y} = C\bar{x} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 8 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 14 \end{bmatrix}$$

$$S_y = CSC' = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 19 & -5 \\ -5 & 7 \end{bmatrix}$$

$$S_y^{-1} = \begin{bmatrix} 7/108 & 5/108 \\ 5/108 & 19/108 \end{bmatrix} \qquad \mu_{y_0} = C\mu_0 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \\ 14 \end{bmatrix}$$

5.31

$$X = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix} \quad \mu_0 = \begin{bmatrix} 7 \\ 11 \end{bmatrix} \quad \bar{x} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$$

expression to find T^2

a) $$T^2 = \frac{(n-1)\,|\hat{\Sigma}_0|}{|\hat{\Sigma}|} - (n-1) = \frac{(n-1)\left|\sum (x_j-\mu_0)(x_j-\mu_0)'\right|}{\left|\sum (x_j-\bar{x})(x_j-\bar{x})'\right|} - (n-1)$$

$$A = \sum (x_j-\mu_0)(x_j-\mu_0)' = \begin{pmatrix} 2-7 \\ 12-11 \end{pmatrix}(2-7,\ 12-11) + \begin{pmatrix} 8-7 \\ 9-11 \end{pmatrix}(8-7,\ 9-11) + \begin{pmatrix} -1 \\ -2 \end{pmatrix}(-1,-2) + \begin{pmatrix} 1 \\ -1 \end{pmatrix}(1,-1)$$

$$= \begin{bmatrix} 25 & -5 \\ -5 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 28 & -6 \\ -6 & 10 \end{bmatrix}$$

$$|A| = 244$$

$$F = \sum (x_j-\bar{x})(x_j-\bar{x})' = \begin{pmatrix} 2-6 \\ 12-10 \end{pmatrix}(2-6,\ 12-10) + \begin{pmatrix} 2 \\ -1 \end{pmatrix}(2,-1) + \begin{pmatrix} 0 \\ -1 \end{pmatrix}(0,-1) + \begin{pmatrix} 2 \\ 0 \end{pmatrix}(2,0)$$

$$= \begin{bmatrix} 16 & -8 \\ -8 & 4 \end{bmatrix} + \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

$$F = \begin{bmatrix} 24 & -10 \\ -10 & 6 \end{bmatrix}$$

$$|F| = 44$$

$$T^2 = \frac{3(244)}{44} - 3 = 13.64$$

5.3| evaluate Λ $n=4$

evaluat Wilks lambda

$$\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}\right)^{n/2} = \left(\frac{|\Sigma(x_j-\bar{x})(x_j-\bar{x})'|}{|\Sigma(x_j-\mu_0)(x_j-\mu_0)'|}\right) \leq c_\alpha$$

$$\Lambda = \left(\frac{44}{244}\right)^2 = 0.0325$$

wilk's lambda

$$\Lambda^{2/n} = \Lambda^{1/2} = (0.0325)^{\frac{1}{2}} = 0.1803$$

5.1

(b) T^2 has the distribution of ; $\frac{(n-1)p}{(n-p)} F_{p,n-p}$

$$= \frac{(4-1)2}{(4-2)} F_{2,4-2}$$

$$= 3F_{2,2}$$

(c) $3F_{2,2}(0.05) = 3(19.00) = 57$

Since $T^2 = 13.6 \not> 57$; we fail to reject H_0 at the 5% level of significance $\therefore \mu = \mu_0$.

5.2

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

$$X = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}$$

$$CX' = \underset{2\times 2}{\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}} \underset{2\times 3}{\begin{bmatrix} 6 & 10 & 8 \\ 9 & 6 & 3 \end{bmatrix}}$$

$$CX' = \begin{bmatrix} 6-9 & 10-6 & 8-3 \\ 6+9 & 10+6 & 8+3 \end{bmatrix} = \begin{bmatrix} -3 & 4 & 5 \\ 15 & 16 & 11 \end{bmatrix}$$

$$X = CX \sim \text{New data Matrix} = \begin{bmatrix} -3 & 15 \\ 4 & 16 \\ 5 & 11 \end{bmatrix}$$

From Page 216 in text

$\bar{y} = C\bar{x}$, $S_y = CSC'$, $\mu_y = C\mu \Rightarrow \mu_{y,0} = C\mu_0$

$$T^2 = n(\bar{y} - \mu_{y,0})' S_y^{-1} (\bar{y} - \mu_{y,0})$$

From Original Matrix

$$\bar{x} = \begin{bmatrix} 8 \\ 6 \end{bmatrix} \quad S = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix} \quad \mu_0 = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$$

5.1(a) $T^2 = n(\bar{x}-\mu_0)' S^{-1}(\bar{x}-\mu_0)$ $\quad X = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$

$n = 4,\ p = 2$

$$\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} \dfrac{2+8+6+8}{4} \\ \dfrac{12+9+9+10}{4} \end{bmatrix} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$$

$$s_{11} = \frac{\sum_{i=1}^{3}(x_1-\bar{x}_1)^2}{n-1} = \frac{(2-6)^2+(8-6)^2+(6-6)^2+(8-6)^2}{3} = \frac{24}{3} = 8$$

$$s_{12} = \frac{\sum_{i=1}^{3}(x_1-\bar{x}_1)(x_2-\bar{x}_2)}{n-1} = \frac{(2-6)(12-10)+(8-6)(9-10)+(6-6)(9-10)+(8-6)(10-10)}{3} = -3.33$$

$$s_{22} = \frac{(12-10)^2+(9-10)^2+(9-10)^2+(10-10)^2}{3} = \frac{6}{3} = 2$$

$$S = \begin{bmatrix} 8 & -3.33 \\ -3.33 & 2 \end{bmatrix} ; \quad S^{-1} = \frac{1}{8(2)-(-3.33)(-3.33)} \begin{bmatrix} 2 & 3.33 \\ 3.33 & 8 \end{bmatrix}$$

$$S^{-1} = \frac{1}{4.91} \begin{bmatrix} 2 & 3.33 \\ 3.33 & 8 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 0.407 & 0.678 \\ 0.678 & 1.629 \end{bmatrix}$$

$$T^2 = 4\,[6-7,\ 10-11] \begin{bmatrix} 0.407 & 0.678 \\ 0.678 & 1.629 \end{bmatrix} \begin{bmatrix} 6-7 \\ 10-11 \end{bmatrix}$$

$$T^2 = [-4, -4] \begin{bmatrix} 0.407 & 0.678 \\ 0.678 & 1.629 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$T^2 = [-4, -4] \begin{bmatrix} -1.085 \\ -2.307 \end{bmatrix}$$

$$T^2 = 13.57 \sim 13.6$$

5.2 Continued

From SAS output

$$\bar{y} = C\bar{x} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 8 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 14 \end{bmatrix}$$

$$S_y = CSC' = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 19 & -5 \\ -5 & 7 \end{bmatrix}$$

$$S_y^{-1} = \begin{bmatrix} 0.065 & 0.046 \\ 0.046 & 0.176 \end{bmatrix}$$

$$\mu_{y,0} = C\mu_0 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \\ 14 \end{bmatrix}$$

$$T^2 = n(\bar{y} - \mu_{y,0})' S_y^{-1} (\bar{y} - \mu_{y,0})$$

$$T^2 = 3\,(2-4,\ 14-14) \begin{bmatrix} 0.065 & 0.046 \\ 0.046 & 0.176 \end{bmatrix} \begin{bmatrix} 2-4 \\ 4-14 \end{bmatrix}$$

$T^2 = 0.777 = \frac{7}{9}$ which shows that T is invariant under the transformation Cx.

5.3

(a) $$T^2 = \frac{(n-1)\,|\hat{\Sigma}_0|}{|\hat{\Sigma}|} - (n-1)$$

$\downarrow$

$$T^2 = \frac{(n-1)\left|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right|}{\left|\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right|} - (n-1)$$

Since $S = \frac{1}{n-1}\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'$

$\therefore (n-1)S = \sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'$

~~From~~ ~~Ques~~ 5.1 $\Rightarrow S = \begin{pmatrix} 8 & -3.33 \\ -3.33 & 2 \end{pmatrix}$

$\therefore (n-1)S = 3\begin{pmatrix} 8 & -3.33 \\ -3.33 & 2 \end{pmatrix} = \begin{pmatrix} 24 & -10 \\ -10 & 6 \end{pmatrix}$

$$\therefore \left|\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right| = \left|\begin{pmatrix} 24 & -10 \\ -10 & 6 \end{pmatrix}\right|$$

$$= (24\times 6) - (-10\times -10)$$

$$= 144 - 100$$

$$= 44$$

$$X = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix} \qquad \mu_0 = [7, 11]$$

$$\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)' = \left(\begin{bmatrix} 2 \\ 12 \end{bmatrix} - \begin{bmatrix} 7 \\ 11 \end{bmatrix}\right)\left(\begin{bmatrix} 2 \\ 12 \end{bmatrix} - \begin{bmatrix} 7 \\ 11 \end{bmatrix}\right)' +$$

$$\left[\begin{pmatrix} 8 \\ 9 \end{pmatrix} - \begin{pmatrix} 7 \\ 11 \end{pmatrix}\right]\left[\begin{pmatrix} 8 \\ 9 \end{pmatrix} - \begin{pmatrix} 7 \\ 11 \end{pmatrix}\right]' + \left[\begin{pmatrix} 6 \\ 9 \end{pmatrix} - \begin{pmatrix} 7 \\ 11 \end{pmatrix}\right]\left[\begin{pmatrix} 6 \\ 9 \end{pmatrix} - \begin{pmatrix} 7 \\ 11 \end{pmatrix}\right]' +$$

$$\left[\begin{pmatrix} 8 \\ 10 \end{pmatrix} - \begin{pmatrix} 7 \\ 11 \end{pmatrix}\right]\left[\begin{pmatrix} 8 \\ 10 \end{pmatrix} - \begin{pmatrix} 7 \\ 11 \end{pmatrix}\right]'$$

5.3 Continued

(a) $$\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)' = \begin{pmatrix}-5\\1\end{pmatrix}(-5\ \ 1) + \begin{pmatrix}1\\-2\end{pmatrix}(1\ \ -2) + \begin{pmatrix}-1\\-2\end{pmatrix}(-1\ \ -2) + \begin{pmatrix}1\\-1\end{pmatrix}(1\ \ -1)$$

$$= \begin{bmatrix}25 & -5\\-5 & 1\end{bmatrix} + \begin{bmatrix}1 & -2\\-2 & 4\end{bmatrix} + \begin{bmatrix}1 & 2\\2 & 4\end{bmatrix} + \begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix}$$

$$= \begin{bmatrix}28 & -6\\-6 & 10\end{bmatrix}$$

$$\left|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right| = \left|\begin{bmatrix}28 & -6\\-6 & 10\end{bmatrix}\right|$$

$$= 28(10) - (-6)(-6)$$

$$= 280 - 36$$

$$= 244$$

5.3(b) $$\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}\right)^{n/2} = \left(\frac{\left|\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right|}{\left|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right|}\right)^{n/2}$$

$\therefore$ From 5.3(a) $\left|\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'\right| = 44$

& $\left|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right| = 244$

$n = 4$

$$\therefore \Lambda = \left(\frac{44}{244}\right)^{4/2} = (0.18033)^2 = 0.0325$$

Wilks' lambda $= \Lambda^{2/n} = (0.0325)^{2/4} = (0.0325)^{1/2} = 0.1803$

5.18

$n = 87$

$p = 3$

$T^2 = 223.3108$

$$\frac{(n-1)p}{(n-p)} F_{p,n-p}(0.05) = \frac{(86)(3)}{85} F_{3,85}(0.05)$$

$$(2.68)$$

$$\doteq 3.035$$

$T^2 \doteq 223.31 > 3.035$ so reject null hyp

6.11

joint 95% confidence region

	CL		SL	
	x_1	x_2	x_1	x_2
1	6	27	25	15

d_1	-19	...
d_2	12	...

$$\bar{d} = \begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix} \qquad S_d = \begin{bmatrix} 199 & 88.38 \\ 88.38 & 418.61 \end{bmatrix}$$

$$\left(\bar{d} - 0\right)' S_d^{-1} (\bar{d} - 0) = 1.24$$

$n = 11 \quad p = 2$

$$\frac{(n-1)p}{n(n-p)} \underset{F_{2,9}}{F_{p,n-p}}(0.05) = \frac{20}{99}(4.26) = 0.8606$$

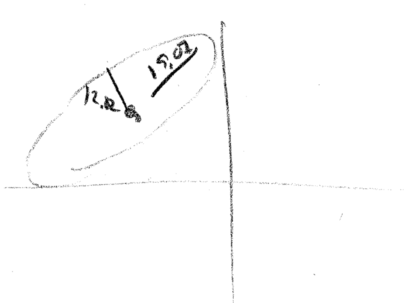
λ_1 449.7 $\quad e_1' = [\quad]$

λ_2 168.08 $\quad e_2' = [\quad]$

Major axis

$$\sqrt{\lambda_1}\sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}} = \sqrt{449.78}\sqrt{.8606} = 19.67$$

$$\sqrt{168.08}\sqrt{0.8608} = 12.02$$



6.31 remove $n=8$

$$\bar{X} = \begin{bmatrix} -12 \\ 8.6 \end{bmatrix} \qquad S = \begin{bmatrix} 136.44 & -52.44 \\ -52.44 & 198.2667 \end{bmatrix}$$

$n=10 \quad p=2$

$$(\bar{d}-\delta)\, S_d^{-1}\, (\bar{d}-\delta) = 1.14$$

$$\frac{(n-1)p}{n(n-p)} F_{p,n-p}(0.05) = \frac{(9)(2)}{(8)}(4.46) = 1.0035$$

$F_{2,8}$

$n(\bar{d}-\delta)S_d^{-1}(\bar{d}-\delta)$

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

Reject null hypot.

use T^2

joint 95% confidence region

% confidence region for δ consists of all δ st

$$(\bar{d}-\delta)' S^{-1} (\bar{d}-\delta) \le \frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha)$$

?

Simultaneous

$$\bar{d} \pm \sqrt{\frac{(n-1)p}{n-p} F_{n,p}(\alpha)} \sqrt{\frac{s_d^2}{n}}$$

$\sqrt{\chi_p^2(\alpha)}$ $n-p$ large

Bonferroni

$(-21.518, -2.08)$

$(-3.35, 20.55)$

$$d_i \pm t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{S_{d_i}^2}{n}}$$

6.4

$$\bar{d} = \begin{bmatrix} -0.558 \\ 0.2956 \end{bmatrix} \qquad S = \begin{bmatrix} 0.456081 & -0.073566 \\ -0.073566 & 0.1838 \end{bmatrix}$$

$(\bar{d} - \delta)' S_d^{-1} (\bar{d} - \delta) = 0.92$, $\frac{p(n-1)}{n-p} F_{p,n-p}(.05)$

$n(\qquad) = 10.21$ $\qquad \frac{(2)(10)}{9} F_{2,9}(.05) = 9.47$

$10.21 > 9.47$ Reject H_0

T^2 int.

$(-1.18, 0.06)$

$(-0.102, 0.69)$

Bon

$(-1.09, -0.02)$

$(-0.044, 0.636)$

6.6

Treatment 2 $\begin{bmatrix}3\\3\end{bmatrix}$ $\begin{bmatrix}1\\6\end{bmatrix}$ $\begin{bmatrix}2\\3\end{bmatrix}$ $n_1 = 3$

Treat 3 $\begin{bmatrix}2\\3\end{bmatrix}$ $\begin{bmatrix}5\\1\end{bmatrix}$ $\begin{bmatrix}3\\1\end{bmatrix}$ $\begin{bmatrix}2\\3\end{bmatrix}$ $n_2 = 4$

TREAT 2

$\bar{x}_1 = \begin{bmatrix}2\\4\end{bmatrix}$ $S_1 = \begin{bmatrix}1 & -3/2\\ -3/2 & 3\end{bmatrix}$

$\bar{x}_1 - \bar{x}_2 = \begin{bmatrix}-1\\2\end{bmatrix}$

Treat 3

$\bar{x}_2 = \begin{bmatrix}3\\2\end{bmatrix}$ $S_2 = \begin{bmatrix}2 & -4/3\\ -4/3 & 4/3\end{bmatrix}$

$$S_{pooled} = \frac{n_1 - 1}{n_1 + n_2 - 2} S_1 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2$$

$$= \frac{2}{5}\begin{bmatrix}1 & -3/2\\ -3/2 & 3\end{bmatrix} + \frac{3}{5}\begin{bmatrix}2 & -4/3\\ -4/3 & 4/3\end{bmatrix}$$

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

6.6

confidence ellip centered at $[-1, 2]'$

$n_1 = 3$

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

$n_4 = 4$

$p = 2$

$|S_{pool} - \lambda I| = 0$

$\lambda_1 = 3.214$

$\lambda_2 = 0.3857$

$$e_1 = \begin{bmatrix} -0.6552 \\ 0.7554 \end{bmatrix} \qquad e_2 = \begin{bmatrix} 6.75544 \\ 0.05520 \end{bmatrix}$$

$F_{2,4}(0.05) = 6.94$

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right) c^2 = \left(\frac{1}{3} + \frac{1}{4}\right) \frac{n_1+n_2-2}{\underbrace{n_1+n_2-p-1}_{v_2}} F_{p,v_2}(\alpha)$$

$$\left(\frac{1}{3} + \frac{1}{4}\right) \frac{5}{4} (6.94) = 5.06$$

confid ellip extends

$$\sqrt{\lambda_i} \sqrt{\frac{1}{n_1} + \frac{1}{n_2} c^2} = \sqrt{\lambda_i} \sqrt{5.06}$$

6.6

part b) $H_0\ \mu_1 - \mu_2 = 0$ $\alpha = 0.01$

$$T^2 = [-1, 2] \left[\left(\frac{1}{3} + \frac{1}{4}\right) \begin{bmatrix} 1.6 & -1.4 \\ -1.4 & 2 \end{bmatrix} \right]^{-1} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$T^2 = 3.88$

$$\frac{(n_1 + n_2 - 2)\, p}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1}(\alpha) = \frac{(5)(2)}{4}(18) = 45$$

Do not reject $H_0\ \mu_1 = \mu_2 = 0$

Jaime Frade

Problem 6.24

F-Statistics and p-values are given in the output

	F-Statistic	p-value	Mean	SSE
MaxBreadth	3.66	0.0298	132.7333	1785.400000
BasHeight	0.47	0.6293	133.3667	1924.300000
BasLength	3.84	0.0251	98.08889	2153.000000
NasHeight	0.10	0.9007	50.44444	840.2000000

Differences over time periods may be due to `MaxBreadth and BasLength`. From the residual plots, the data seems normal. The usual MANOVA assumptions are satistfied

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

S=2 M=0.5 N=41

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.83010268	2.05	8	168	0.0436

```
95% simultaneous confidence intervals (values in SAS)
m = (pg(g-1))/2 = 12
t(df=87, 0.05/24) = 2.94
```

```
MaxBreath
```

Least Squares Means for Effect TimePeriod

i	j	Difference Between Means	Simultaneous 98.75% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-1.000000	-4.442312	2.442312
1	3	-3.100000	-6.542312	0.342312
2	3	-2.100000	-5.542312	1.342312

Jaime Frade

BasHeight

Least Squares Means for Effect TimePeriod

i	j	Difference Between Means	Simultaneous 98.75% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.900000	-2.673706	4.473706
1	3	-0.200000	-3.773706	3.373706
2	3	-1.100000	-4.673706	2.473706

BasLength

Least Squares Means for Effect TimePeriod

i	j	Difference Between Means	Simultaneous 98.75% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.100000	-3.680110	3.880110
1	3	3.133333	-0.646777	6.913443
2	3	3.033333	-0.746777	6.813443

NasHeight

Least Squares Means for Effect TimePeriod

i	j	Difference Between Means	Simultaneous 98.75% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.300000	-2.061423	2.661423
1	3	-0.033333	-2.394756	2.328089
2	3	-0.333333	-2.694756	2.028089

OUTPUT

The SAS System

The GLM Procedure

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		
Corrected Total	89	1935.600000			

R-Square	Coeff Var	Root MSE	MaxBreadth Mean
0.077599	3.412936	4.530104	132.7333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TimePeriod	2	150.2000000	75.1000000	3.66	0.0298

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TimePeriod	2	150.2000000	75.1000000	3.66	0.0298

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Linear TimePeriod	1	150.1983368	150.1983368	7.32	0.0082
Quadratic TimePeriod	1	0.0016632	0.0016632	0.00	0.9928

The SAS System

The GLM Procedure

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		
Corrected Total	89	1944.900000			

R-Square	Coeff Var	Root MSE	BasHeight Mean
0.010592	3.526383	4.703019	133.3667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TimePeriod	2	20.60000000	10.30000000	0.47	0.6293

Source	DF	Type III SS	Mean Square	F Value	Pr > F

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TimePeriod	2	20.60000000	10.30000000	0.47	0.6293

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Linear TimePeriod	1	2.69719335	2.69719335	0.12	0.7278
Quadratic TimePeriod	1	17.90280665	17.90280665	0.81	0.3708

The SAS System

The GLM Procedure

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		
Corrected Total	89	2343.288889			

R-Square	Coeff Var	Root MSE	BasLength Mean
0.081206	5.071572	4.974648	98.08889

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TimePeriod	2	190.2888889	95.1444444	3.84	0.0251

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TimePeriod	2	190.2888889	95.1444444	3.84	0.0251

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Linear TimePeriod	1	174.0152807	174.0152807	7.03	0.0095
Quadratic TimePeriod	1	16.2736082	16.2736082	0.66	0.4196

The SAS System

The GLM Procedure

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		
Corrected Total	89	842.2222222			

R-Square	Coeff Var	Root MSE	NasHeight Mean

R-Square	Coeff Var	Root MSE	NasHeight Mean
0.002401	6.160534	3.107647	50.44444

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TimePeriod	2	2.02222222	1.01111111	0.10	0.9007

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TimePeriod	2	2.02222222	1.01111111	0.10	0.9007

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Linear TimePeriod	1	0.16496882	0.16496882	0.02	0.8963
Quadratic TimePeriod	1	1.85725341	1.85725341	0.19	0.6621

The SAS System

The GLM Procedure
Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for TimePeriod
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1			
		MaxBreadth	**BasHeight**	**BasLength**	**NasHeight**
0.18696913	**92.61**	-0.01731044	-0.00316643	0.01599700	0.00492103
0.01491287	**7.39**	-0.01243028	0.01722156	-0.00745836	0.01318513
0.00000000	**0.00**	0.00150700	-0.00942315	0.00123927	0.03287969
0.00000000	**0.00**	0.01189934	0.01182951	0.01252783	0.00000000

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

S=2 M=0.5 N=41

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

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

```
Time          MaxBreadth      LSMEAN
```

Jaime Frade

```
        Period               LSMEAN        Number

        4000 B.C.         131.366667          1
        3300 B.C.         132.366667          2
        1850 B.C.         134.466667          3

                           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

                          Variable:  R1

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

        Time                BasHeight       LSMEAN
        Period                 LSMEAN       Number

        4000 B.C.          133.600000          1
        3300 B.C.          132.700000          2
        1850 B.C.          133.800000          3

                    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
                          Variable:  R2

                          Normal Probability Plot
  12.5+                                                     *
      |                                                 *  ++
      |                                                  ++
      |                                               *++
      |                                             *++
   7.5+                                            *+
      |                                         +**
      |                                       +**
      |                                     ***
   2.5+                                 ****
      |                               **+
      |                             ***
      |                           **+
      |                         ***
      |                       **+
  -2.5+                     ***
      |                   ***+
      |                 **++
      |                 ++
      |               +**
  -7.5+             ++*
      |           ++**
      |         ++*
      |       ++*
      |++
 -12.5+*
       +----+----+----+----+----+----+----+----+----+----+
           -2       -1         0        +1        +2
         Time                BasLength      LSMEAN
         Period                LSMEAN       Number
         4000 B.C.          99.1666667         1
         3300 B.C.          99.0666667         2
         1850 B.C.          96.0333333         3

                     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

                          Variable:  R3
                          Normal Probability Plot
  14.5+                                                     *
      |                                                     +
      |                                                   ++
      |                                                 ++
   9.5+                                              *+*
      |                                            *+
      |                                          **
      |                                        +**
      |                                      +**
   4.5+                                    ***
      |                                  ***
      |                                **
      |                              ***
      |                           +***
  -0.5+                         +**
      |                       ***
      |                     +**
```

```
       |               ++**
       |             ++**
  -5.5+            ++**
       |        *****
       |         +*
       |       ++*
       |     +* *
 -10.5+*   +*
      +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

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

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

```
                  Variable:  R4
                Normal Probability Plot
  9.5+                                                  *
      |                                               +++
      |                                            +++
      |                                         *+* *
      |                                       ***
      |                                    ***+
      |                                 ****
  1.5+                               **+
      |                          +****
      |                       ****
      |                     ***
      |                  ***
      |              +***
      |           +****
      |    * *+* *
 -6.5+*   +++
      +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

CODE

```
Options PS=55 LS=78 PageNo=1;
/*
 * The ODS HTML generates SAS Output in HTML
 * format that makes tables that drop into
 * word processors pretty easily. You may need
 * to change the file destination, or you can
 * delete the statement entirely if you like.
 */
ODS HTML File="C:\Documents and Settings\Jaime\Desktop\Skulls.rtf" Style=Minimal;
Proc Format;
 Value BC 1="4000 B.C."
          2="3300 B.C."
          3="1850 B.C.";
Run;
Data EgyptianSkulls;
 Skull+1;
 Input MaxBreadth BasHeight BasLength NasHeight TimePeriod;
 LMaxBreadth=Log(MaxBreadth);
 LBasHeight=Log(BasHeight);
 LBasLength=Log(BasLength);
 LNasHeight=Log(NasHeight);
 Label MaxBreadth="Maximum Breadth of Skull (mm)"
       BasHeight ="Basibregmatic Height of Skull (mm)"
         BasLength ="Basialveolar Length of Skull (mm)"
         NasHeight ="Nasal Height of Skull (mm)"
         TimePeriod="Time Period";
 Format TimePeriod BC.;
DataLines;
********
;
Proc Print Data=EgyptianSkulls Label;
 Id Skull;
Run;
```

Jaime Frade

```
/*
 * MANOVA analysis
 * Use the ORDER=INTERNAL option so that contrast
 * coefficients correspond correctly with the Time Periods.
 */
Proc GLM Data=EgyptianSkulls Order=Internal;
 Class TimePeriod;
 Model MaxBreadth BasHeight BasLength NasHeight = TimePeriod;
MANOVA H=TimePeriod;
 LSMeans TimePeriod / PDiff Adjust=Tukey Alpha=0.0125; /* 0.05/4 */
 Output Out=Diagnostics R=R1-R4; /* output the residuals */
means TimePeriod;
LSMeans TimePeriod / PDiff CL Adj=Bon Alpha=0.0125;
Run;
Quit;

/*
 * Invoke SAS/Insight for Interactive Data Analysis
 * to examine residuals for normality and outliers.
 */
Proc univariate Data=Diagnostics plot normal;
var R1-R4;
Run;

ODS HTML Close;
```

Problem 6.25

F-Statistics and p-values are given in the output

	F-Statistic	p-value	Mean	SSE
vanadium	19.17	<.0001	6.180357	187.5752427
iron	20.01	<.0001	27.04643	4221.158113
beryllium	5.88	<.0001	0.341429	4.43566535
SathydroCarb	22.67	<.0001	5.299107	57.0404334
AromhydroCarb	16.41	<.0001	6.433571	338.0228861

Differences over time periods may be due to each variable. Using no transformations, one-way MANOVA resulted in p-value < 0.0001. Null hypothesis is rejected. From SAS output, the p-value of all five variables are smaller than 0.0001, which implies these variables are responsible for any difference in the zones.

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

S=2 M=1 N=23.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.11591099	18.98	10	98	<.0001

```
95% simultaneous confidence intervals (values in SAS)
m = (pg(g-1))/2 = 115
t(df=87, 0.05/30) = 3.074
```

```
The intervals that do not include zero are listed below.
```

OUTPUT

The SAS System

The GLM Procedure

Class Level Information		
Class	Levels	Values
Zones	3	SubMuli Upper Wilhelm

Number of Observations Read	56
Number of Observations Used	56

The SAS System

The GLM Procedure

Dependent Variable: vanadium vanadium

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			

The SAS System

The GLM Procedure

Dependent Variable: iron iron

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		
Corrected Total	55	7407.839286			

The SAS System

The GLM Procedure

Dependent Variable: beryllium beryllium

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			

The SAS System

The GLM Procedure

Dependent Variable: SathydroCarb Saturated hydrocarbons

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			

The SAS System

The GLM Procedure

Dependent Variable: AromhydroCarb Aromatic hydrocabons

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			

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

S=2 M=1 N=23.5

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

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

The SAS System

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Zones	vanadium LSMEAN	LSMEAN Number
SubMuli	4.44545455	1
Upper	7.22631579	2
Wilhelm	3.22857143	3

Least Squares Means for Effect Zones

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-2.780861	-4.373321	-1.188402
1	3	1.216883	-1.031910	3.465676
2	3	3.997744	2.084705	5.910784

Zones	iron LSMEAN	LSMEAN Number
SubMuli	33.0909091	1
Upper	22.2526316	2
Wilhelm	43.5714286	3

Least Squares Means for Effect Zones

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	10.838278	3.283938	18.392617
1	3	-10.480519	-21.148384	0.187345
2	3	-21.318797	-30.393910	-12.243684

Zones	beryllium LSMEAN	LSMEAN Number
SubMuli	0.17090909	1
Upper	0.43210526	2
Wilhelm	0.11714286	3

Least Squares Means for Effect Zones

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-0.254154	-0.474152	-0.034157
1	3	0.040663	-0.270007	0.351332
2	3	0.294817	0.030532	0.559102

Zones	SathydroCarb LSMEAN	LSMEAN Number
SubMuli	6.56090909	1
Upper	4.65815789	2
Wilhelm	6.79571429	3

Least Squares Means for Effect Zones

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.346943	0.185685	0.508200
1	3	-0.043061	-0.270781	0.184659
2	3	-0.390004	-0.583724	-0.196283

Zones	AromhydroCarb LSMEAN	LSMEAN Number
SubMuli	5.4836364	1
Upper	5.7678947	2
Wilhelm	11.5400000	3

Least Squares Means for Effect Zones

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-0.284258	-2.421993	1.853476
1	3	-6.056364	-9.075166	-3.037561
2	3	-5.772105	-8.340190	-3.204021

CODE

```
Options PS=55 LS=78 PageNo=1;
/*
 * The ODS HTML generates SAS Output in HTML
 * format that makes tables that drop into
 * word processors pretty easily. You may need
 * to change the file destination, or you can
 * delete the statement entirely if you like.
 */
ODS HTML File="C:\Documents and Settings\Jaime\Desktop\crudeoil.doc" Style=Minimal;
Data Crudeoilsample;
infile "E:\FSU-FALL07-School\Fall07-school\sta5707\test2\Hwchap6-8\T11-7.dat";
input vanadium iron beryllium SathydroCarb AromhydroCarb Zones$;
 beryllium1=sqrt(beryllium);
 SathydroCarb1=log(SathydroCarb);
 Label vanadium="vanadium"
       iron ="iron"
         beryllium ="beryllium"
         SathydroCarb ="Saturated hydrocarbons"
         AromhydroCarb = "Aromatic hydrocabons"
         Zones="Zones of sandstone";
Proc Print Data=Crudeoilsample Label;
 Id crudeoil;
Run;
/*
 * MANOVA analysis
 * Use the ORDER=INTERNAL option so that contrast
 * coefficients correspond correctly with the Time Periods.
 */
Proc GLM Data=Crudeoilsample;
 Class Zones;
 Model vanadium iron beryllium1 SathydroCarb1 AromhydroCarb = Zones;
 MANOVA H=Zones;
 LSMeans Zones / PDiff Adjust=Tukey Alpha=0.0125; /* 0.05/4 */
 Output Out=Diagnostics R=R1-R4; /* output the residuals */
means Zones;
LSMeans Zones / PDiff CL Adj=Bon Alpha=0.05;
Run;
Quit;

/*
 * Invoke SAS/Insight for Interactive Data Analysis
 * to examine residuals for normality and outliers.
 */
Proc univariate Data=Diagnostics plot normal;
var R1-R4;
Run;


ODS HTML Close;
```

Problem 8.14

Summary

The first principal component, which explains 97.17% of the total variance. (see pg 446). Consequently, sampe variation Is summarized very well by one principal component and a reduction from 20 observations on 3 observations to 20 observations on 1 principal component is reasonable. Adding two principal components can increase to 99.37%.

$$\hat{y}_1 = 0.050841(Sweatrate) - 0.998284(sodium) - 0.029072(potassium)$$
$$\hat{y}_2 = -.573704(Sweatrate) + 0.053020(sodium) + 0.817345(potassium)$$
$$\hat{y}_3 = 0.817484(Sweatrate) - 0.024877(sodium) + 0.575415(potassium)$$

The PRINCOMP Procedure

Observations	20
Variables	3

Simple Statistics

	sweatrate	sodium	potassium
Mean	4.640000000	45.40000000	9.965000000
StD	1.696870184	14.13465320	1.904641146

Covariance Matrix

		sweatrate	sodium	potassium
sweatrate	sweatrate	2.8793684	10.0100000	-1.8090526
sodium	sodium	10.0100000	199.7884211	-5.6400000
potassium	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
sweatrate	sweatrate	0.050841	-.573704	0.817484
sodium	sodium	0.998284	0.053020	-.024877
potassium	potassium	-.029072	0.817345	0.575415

Problem 8.14

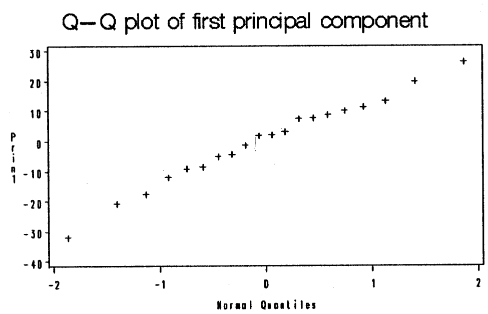
The FACTOR Procedure
Initial Factor Method: Principal Components

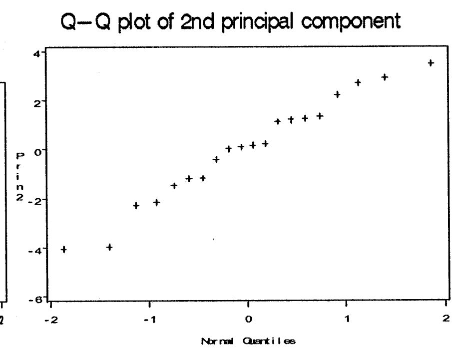
The FACTOR Procedure
Initial Factor Method: Principal Components

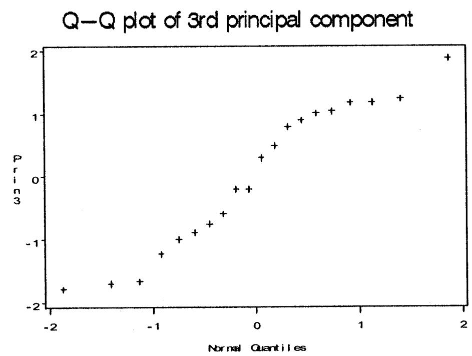
```
    Scree Plot of Eigenvalues
      3.0 ^
          ,
          ,                 1
          ,
          ,
          ,
          ,
      2.5 ^
          ,
          ,
          ,
          ,
          ,
          ,
      2.0 ^
          ,
    E     ,                          2
    i     ,
    g     ,
    e     ,
    n     ,                                   3
    v 1.5 ^
    a     ,
    l     ,
    u     ,
    e     ,
    s     ,
          ,
      1.0 ^
          ,
          ,
          ,                                            4
          ,
          ,
          ,
      0.5 ^
          ,
          ,
          ,
          ,
          ,
          ,
      0.0 ^                                                     5        6        7
          Šƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ
                  0       1       2       3       4       5        6        7

                                        Number
```

Problem 8.14







Plots do not seem to indicate outliers.

Obs	individual	sweatrate	sodium	potassium	Prin1	Prin2	Prin3
1	15	1.5	13.5	10.1	-32.0088	0.22042	-1.69566
2	7	2.4	24.8	14.0	-20.7958	3.49086	1.00309
3	13	3.5	27.8	9.8	-17.6230	-0.41400	-0.58905
4	8	7.2	33.1	7.6	-12.0800	-4.05385	1.03788
5	6	4.6	36.1	7.9	-9.2260	-2.15796	-0.98958
6	11	3.9	36.9	12.7	-8.6025	2.20931	1.18027
7	14	4.5	40.2	8.4	-5.1527	-1.47453	-0.88561
8	20	5.5	40.9	9.4	-4.4321	-1.19378	0.48987
9	19	4.1	44.1	11.2	-1.3611	1.25029	0.30154
10	3	3.8	47.2	10.9	1.7270	1.34157	-0.19345
11	9	6.7	47.4	8.5	2.1439	-2.27320	0.79128
12	1	3.7	48.5	9.3	3.0662	0.16011	-1.22820
13	18	6.5	52.8	10.9	7.4547	0.08948	1.87445
14	4	3.2	53.2	12.0	7.6542	2.90299	-0.20024
15	10	5.4	54.1	11.3	8.6849	1.11642	1.17304
16	5	3.1	55.5	9.7	10.0121	1.20241	-1.66266
17	16	8.5	56.4	7.1	11.2607	-3.97297	1.23328
18	12	4.5	58.8	12.3	13.3020	2.69929	0.89580
19	2	5.7	65.1	8.0	19.7772	-1.16971	-0.75423
20	17	4.5	71.6	8.2	26.1992	0.02684	-1.78182

Problem 8.14

CODE

```
data sweat;
infile 'C:\Documents and Settings\jfrade\Desktop\T5-1.dat';
input individual sweatrate sodium potassium;
ods rtf file='C:\Documents and Settings\jfrade\Desktop\PCA_output1.rtf';
proc princomp cov data=sweat out=result;
var sweatrate sodium potassium;
proc factor rotate=varimax score scree plot nfactors=3;
proc factor method=ml rotate=hk score scree plot nfactors=3 ;
ods rtf close;run;quit;
ods rtf file='C:\Documents and Settings\jfrade\Desktop\PCA_graph2.rtf';
proc capability data=result;
var Prin1;
qqplot;
title 'Q-Q plot of first principal component';
run;
ods rtf close;run;quit;
```

	prop	
a) $\lambda_1 = 0.00036$	0.6016	proportion explained = 85.74%
$\lambda_2 = 0.000079$	0.1326	$\frac{(\lambda_1+\lambda_2+\lambda_3)}{\sum \lambda_i}$
$\lambda_3 = 0.000074$	0.1232	

(b) No $\Rightarrow$ 4 PC would do a better job $\rightarrow$ 94.25%

7) $\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix} =$

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

Since we know that the corr. matrix is given by

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

$p = 4$

then $\lambda_1 = 1 + 3\rho$

$\{\lambda_2 = \lambda_3 = \lambda_4 = 1 - \rho$

If $\Rightarrow \rho > 0$

then $1 + 3\rho > 1 - \rho$

$\downarrow$

$\therefore$ Total of sam variance explained by $PC_1 = \frac{1+3\rho}{1+3\rho + 3(1-\rho)}$

$= \frac{1+3\rho}{4}$

If $\rho < 0$ then

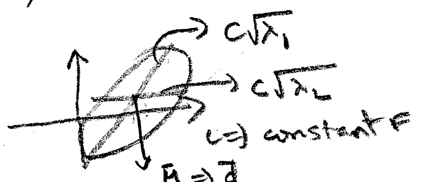
$1 - \rho > 1 + 3\rho$

$\Rightarrow \frac{3(1-\rho)}{4} = \frac{3-3\rho}{4}$

(2) (a) 95% T² C.I. = (−8.35, 40.04) NB
(−22.10, 47.24)

2 Files ⇒ 2Bconf. or EX.5.2
can work
↓
Input meandiff &
Cov-matrix

(b) 95% Bon C.I. = (−3.41, 35.12)
(−15.05, 40.19)



		F-value	P-value
(3) Factor 1 ⇒	Species Effect	15.46	< 0.0001
Factor 2 ⇒	Time Effect	42.52	< 0.0001
Factor 1 × factor 2 ⇒	Interactn effect	0.74	0.5741

The Species & time effects were significant but the interaction effect was not significant on the spectral reflectance

Overall MANOVA ⇒ The overall MANOVA was significant with Wilk's lambd

$\Lambda^* = 0.0688$; p-value < 0.0001

(5) (a) Only one $\lambda_1 = 89.9$, proportion of variance explain by 1st PC = 0.9801 ~ 98.10%

(b) How ? ⇒ proc sort data = track... ; ⇒ Something wrong in his code.
↓
do you look at the output.

by prin1; ⇒
run;

$\lambda_1 = 6.62 \to 0.8278$
$\lambda_2 = 0.878 \to .1098$
93.75
⇒ 2 PC.

Review Test 2

① (a) Page 212

$$T^2 = n(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0) > \frac{(n-1)}{(n-p)} p F_{p,n-p}(\alpha)$$

(a) Distn of $T^2 = \frac{(n-1)}{(n-p)} p F_{p,n-p}$; $p = 2, n = 4$

$$= \frac{3 \cdot 2}{2} F_{2,2}(0.05)$$

$$= 3F_{2,2}(0.05)$$

$$= 3 \cdot 19$$

$$= 57$$

(b) $X = \begin{pmatrix} 6 & 10 & 8 & 8 \\ 9 & 6 & 3 & 2 \end{pmatrix}$ Data matrix $= \begin{pmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \\ 8 & 2 \end{pmatrix}$ $\bar{X} = \begin{pmatrix} 8 \\ 5 \end{pmatrix}$

$\mu_0^T = (9, 5)$ * see SAS

$S = \begin{bmatrix} 2.67 & -2.00 \\ -2.00 & 10.00 \end{bmatrix}$ $S^{-1} = \begin{bmatrix} 0.441 & 0.088 \\ 0.088 & 0.118 \end{bmatrix}$

$$T^2 = 4(8-9, 5-5) \begin{bmatrix} 0.441 & 0.088 \\ 0.088 & 0.118 \end{bmatrix} \begin{bmatrix} 8-9 \\ 5-5 \end{bmatrix}$$

$T^2 = 1.76$

(c) $H_0: \mu = \mu_0$

$H_A: \mu \neq \mu_0$

Since $T^2 = 1.76 < 57$ then we fail to reject the null hypothesis and conclude that the test for equality of means vector is not equal.