Multivariate Statistical Analysis Final Solution

January 13, 2012

1. Follow Result 6.1 (p. 275),

$$T^2 = n\bar{\mathbf{d}}'\mathbf{S}_d^{-1}\bar{\mathbf{d}} = 30[1 \quad 0 \quad -1]\frac{1}{162}\begin{bmatrix} 14 & 4 & -2 \\ 4 & 14 & 2 \\ -2 & 2 & 17 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \frac{175}{27} = 6.4815$$

$$\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha) = \frac{(30-1)\times 3}{30-3}F_{3,27}(0.05) = \frac{29\times 3}{27}\times 2.96 = 9.538$$

$$T^2 < \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$$

$\therefore$ Treatment difference does not exist for 5% significance level

95% Bonferroni individual mean differences :

$$\delta_i = \bar{d}_i \pm t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{s_{di}^2}{n}} = \bar{d}_i \pm t_{29}(0.00833)\sqrt{\frac{s_{di}^2}{n}}$$

$$\delta_1 = 1 \pm 2.541\sqrt{\frac{13}{30}} = 1 \pm 1.673, \text{ i.e., } (-0.673, 2.673)$$

$$\delta_2 = 0 \pm 2.541\sqrt{\frac{13}{30}} = 0 \pm 1.673, \text{ i.e., } (-1.673, 1.673)$$

$$\delta_3 = -1 \pm 2.541\sqrt{\frac{10}{30}} = -1 \pm 1.467, \text{ i.e., } (-2.467, 0.467)$$

(10%)

---

2. ,

$$\mathbf{S}_{\text{pooled}} = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$$

Equation (6 - 23) in p. 285,

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}\mathbf{S}_{\text{pooled}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = [-1 \quad -1 \quad 0]\frac{8}{1}\frac{1}{162}\begin{bmatrix} 14 & 4 & -2 \\ 4 & 14 & 2 \\ -2 & 3 & 17 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

$$= \frac{8}{162} \times 36 = 1.7778$$

Result 6.2 in p. 286,

$$\frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1}F_{p,n_1+n_2-p-1}(\alpha) = \frac{30\times 3}{28}F_{3,28}(0.05) = \frac{30\times 3}{28}\times 2.95 = 9.482$$

$$T^2 < \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1}F_{p,n_1+n_2-p-1}(\alpha)$$

$\therefore$ accept $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ at the 5% significance level

95% Bonferroni simultaneous confidence intervals for individual mean differences:

Follow the equation in p. 291,

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

$$\mu_{11} - \mu_{21}: \quad -1 \pm t_{30}(0.00833)\sqrt{\frac{1}{8}\times 13} = -1 \pm 3.233, \text{ i.e., } (-4.233, 2.233)$$

$$\mu_{12} - \mu_{22}: \quad -1 \pm t_{30}(0.00833)\sqrt{\frac{1}{8}\times 13} = 1 \pm 3.233, \text{ i.e., } (-2.233, 4.233)$$

$$\mu_{13} - \mu_{23}: \quad 0 \pm t_{30}(0.00833)\sqrt{\frac{1}{8}\times 10} = 0 \pm 2.835, \text{ i.e., } (-2.835, 2.835)$$

(10%)

----

3. Use ANOVA,

$$\bar{x}_A = \frac{22}{4} = 5.5, \quad \bar{x}_B = \frac{19}{4} = 4.75, \quad \bar{x}_C = \frac{21}{4} = 5.25, \quad \bar{x} = \frac{62}{12} = 5.1667$$

$$SS(between) = 4\left[(5.5-5.1667)^2 + (4.75-5.1667)^2 + (5.25-5.1667^2)\right] = 1.1667$$

$$\begin{aligned} SS(within) = {} & (7-5.5)^2 + (5-5.5)^2 + (6-5.5)^2 + (4-5.5)^2 + \\ & (6-4.75)^2 + (4-4.75)^2 + (4-4.75)^2 + (5-4.75)^2 + \\ & (5-5.25)^2 + (5-5.25)^2 + (5-5.25)^2 + (6-5.25)^2 = 8.5000 \end{aligned}$$

$$\begin{aligned} SS(total) = {} & (7-5.1667)^2 + (5-5.1667)^2 + (6-5.1667)^2 + (4-5.1667)^2 + \\ & (6-5.1667)^2 + (4-5.1667)^2 + (4-5.1667)^2 + (5-5.1667)^2 + \\ & (5-5.1667)^2 + (5-5.1667)^2 + (5-5.1667)^2 + (6-5.1667)^2 = 9.6667 \end{aligned}$$

$df(between) = 3 - 1 = 2$

$df(within) = 3(4-1) = 9$

$df(total) = 3 \times 4 - 1 = 11 = 2 + 9$

ANOVA Table:

Source	Sum of Squares	df	Mean Square	F
between	1.1667	2	0.5833	0.6176
within	8.5000	9	0.9444	
total	9.6667	11		

$F = 0.6176 < F_{2,9}(0.05) = 4.26$, accept the hypothesis

(10%)

---

4. For first variable:

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

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

For second variable:

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

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

Cross product contribution

$SCP_{obs} = 275, \quad SS_{mean} = 240, \quad SS_{treat} = 48, \quad SS_{res} = -13$

MANOVA Table

Treatment $\mathbf{B} = \begin{bmatrix} 36 & 48 \\ 48 & 84 \end{bmatrix}$ d.f. = 3 – 1 = 2

Residual $\mathbf{W} = \begin{bmatrix} 18 & -13 \\ -13 & 18 \end{bmatrix}$ d.f. = 5 + 3 + 4 -3 = 9

Total (corrected) $\mathbf{B} + \mathbf{W} = \begin{bmatrix} 54 & 35 \\ 35 & 102 \end{bmatrix}$ d.f. = 11

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{155}{4283} = 0.0362$$

Using Table 6.3 with p = 2 and g = 3

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

Since $F_{4,16}(0.01) = 4.77$, we conclude that the mean vectors are different at 0.01

level

(10%)

---

5. Use Result 7.1 (p. 364) Result 7.1, p. 364, Textbook

$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix}$, $\mathbf{y}_1' = [4 \quad 2 \quad 3 \quad 1 \quad -1]$, $\mathbf{y}_2' = [2 \quad 3 \quad 5 \quad 7 \quad 8]$

$\mathbf{Z}'\mathbf{y}_1 = \begin{bmatrix} 9 \\ -11 \end{bmatrix}$, $\mathbf{Z}'\mathbf{y}_2 = \begin{bmatrix} 25 \\ 16 \end{bmatrix}$, $\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$, $(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix}$

$\hat{\boldsymbol{\beta}}_{(1)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_1 = \begin{bmatrix} 1.8 \\ -1.1 \end{bmatrix}$, $\hat{\boldsymbol{\beta}}_{(2)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_2 = \begin{bmatrix} 5 \\ 1.6 \end{bmatrix}$

$\hat{\mathbf{y}}_1 = \mathbf{Z}\hat{\boldsymbol{\beta}}_{(1)} = [4 \quad 2.9 \quad 1.8 \quad 0.7 \quad -0.4]'$, $\hat{\mathbf{y}}_2 = \mathbf{Z}\hat{\boldsymbol{\beta}}_{(2)} = [1.8 \quad 3.4 \quad 5 \quad 6.6 \quad 8.2]'$

$\hat{\boldsymbol{\varepsilon}}_{(1)} = \mathbf{y}_1 - \hat{\mathbf{y}}_1 = [0 \quad -0.9 \quad 1.2 \quad 0.3 \quad -0.6]'$,

$\hat{\boldsymbol{\varepsilon}}_{(2)} = \mathbf{y}_2 - \hat{\mathbf{y}}_2 = [0.2 \quad -0.4 \quad 0 \quad 0.4 \quad -0.2]'$

$\hat{\boldsymbol{\varepsilon}}_{(1)}'\hat{\boldsymbol{\varepsilon}}_{(1)} = 2.7$, $\hat{\boldsymbol{\varepsilon}}_{(2)}'\hat{\boldsymbol{\varepsilon}}_{(2)} = 0.4$

(10%)

---

6. Eq. (7.-40), p. 399, Textbook

100(1-α)% confidence region for $E([Y_1, Y_2]' \mid \mathbf{z}_0) = \boldsymbol{\beta}'\mathbf{z}_0$ is provided by the inequality

$$(\boldsymbol{\beta}'\mathbf{z}_0 - \hat{\boldsymbol{\beta}}'\mathbf{z}_0)'\left(\frac{n}{n-r-1}\hat{\boldsymbol{\Sigma}}\right)^{-1}(\boldsymbol{\beta}'\mathbf{z}_0 - \hat{\boldsymbol{\beta}}'\mathbf{z}_0) \le \mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0\left[\left(\frac{m(n-r-1)}{n-r-m}\right)F_{m,n-r-m}(\alpha)\right]$$

$\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}_{(1)} \quad \hat{\boldsymbol{\beta}}_{(2)}] = \begin{bmatrix} 1.8 & 5 \\ -1.1 & 1.6 \end{bmatrix}$, $\mathbf{z}_0 = \begin{bmatrix} 1 \\ -0.5 \end{bmatrix}$,

$$\hat{\boldsymbol{\beta}}'\mathbf{z}_0 = \begin{bmatrix} 1.8 & -1.1 \\ 5 & 1.6 \end{bmatrix}\begin{bmatrix} 1 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 2.35 \\ 4.2 \end{bmatrix}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \frac{1}{5}\begin{bmatrix} 0 & -0.9 & 1.2 & 0.3 & -0.6 \\ 0.2 & -0.4 & 0 & 0.4 & -0.2 \end{bmatrix}\begin{bmatrix} 0 & 0.2 \\ -0.9 & -0.4 \\ 1.2 & 0 \\ 0.3 & 0.4 \\ -0.6 & -0.2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 2.7 & 0.6 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.54 & 0.12 \\ 0.12 & 0.08 \end{bmatrix}$$

$$\mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0 = [1 \quad -0.5]\begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix}\begin{bmatrix} 1 \\ -0.5 \end{bmatrix} = [1 \quad -0.5]\begin{bmatrix} 0.2 \\ -0.05 \end{bmatrix} = 0.225$$

90% confidence region for $E([Y_1, Y_2] \mid \mathbf{z}_0) = \boldsymbol{\beta}'\mathbf{z}_0$ is

$$(\boldsymbol{\beta}'\mathbf{z}_0 - \begin{bmatrix} 2.35 \\ 4.2 \end{bmatrix})' \left(\frac{5}{5-1-1} \begin{bmatrix} 0.54 & 0.12 \\ 0.12 & 0.08 \end{bmatrix} \right)^{-1} (\boldsymbol{\beta}'\mathbf{z}_0 - \begin{bmatrix} 2.35 \\ 4.2 \end{bmatrix}) \le 0.225 \left[\left(\frac{2(5-1-1)}{5-1-2} \right) F_{2,2}(0.1) \right]$$

i.e.,

$$(\boldsymbol{\beta}'\mathbf{z}_0 - \begin{bmatrix} 2.35 \\ 4.2 \end{bmatrix})' \begin{bmatrix} 1.6667 & -2.5 \\ -2.5 & 11.25 \end{bmatrix} (\boldsymbol{\beta}'\mathbf{z}_0 - \begin{bmatrix} 2.35 \\ 4.2 \end{bmatrix}) \le 0.225 * 3 * 9.00 = 6.075$$

Eq. (7-42), p. 399, Textbook

100(1-α)% prediction region for $\mathbf{Y}_0 = [Y_1, Y_2]'$ is provided by the inequality

$$(\mathbf{Y}_0 - \hat{\boldsymbol{\beta}}'\mathbf{z}_0)' \left(\frac{n}{n-r-1} \hat{\boldsymbol{\Sigma}} \right)^{-1} (\mathbf{Y}_0 - \hat{\boldsymbol{\beta}}'\mathbf{z}_0) \le \left(1 + \mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0\right) \left[\left(\frac{m(n-r-1)}{n-r-m} \right) F_{m,n-r-m}(\alpha) \right]$$

Thus, 90% prediction region for $\mathbf{Y}_0 = [Y_1, Y_2]'$ is

$$(\mathbf{Y}_0 - \begin{bmatrix} 2.35 \\ 4.2 \end{bmatrix})' \begin{bmatrix} 1.6667 & -2.5 \\ -2.5 & 11.25 \end{bmatrix} (\mathbf{Y}_0 - \begin{bmatrix} 2.35 \\ 4.2 \end{bmatrix}) \le 1.225 * 3 * 9.00 = 33.075$$

(10%)

---

7. For use of the result in p. 442, Textbook, eigenvalues and the first eigenvector have to be found.

$$\begin{vmatrix} 8-\lambda & 0 & -2 \\ 0 & 6-\lambda & 0 \\ -2 & 0 & 5-\lambda \end{vmatrix} = 0 \Rightarrow (6-\lambda) = 0 \text{ or } \begin{vmatrix} 8-\lambda & -2 \\ -2 & 5-\lambda \end{vmatrix} = \lambda^2 - 13\lambda + 36 = 0, \quad , \quad ,$$

$$\therefore \hat{\lambda}_1 = 9, \hat{\lambda}_2 = 6, \hat{\lambda}_3 = 4$$

$$\hat{\lambda}_1 = 9, \begin{bmatrix} -1 & 0 & -2 \\ 0 & -3 & 0 \\ -2 & 0 & -4 \end{bmatrix} \begin{bmatrix} \hat{e}_{11} \\ \hat{e}_{12} \\ \hat{e}_{13} \end{bmatrix} = 0 \Rightarrow \hat{\mathbf{e}}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$$

$$Y_1 = \hat{\mathbf{e}}_1'\mathbf{x} = \frac{1}{\sqrt{5}}(2x_1 - x_3) \approx 0.8944x_1 - 0.4472x_2$$

$$\text{proportion of the total sample variance due to } Y_1 = \frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3} = \frac{9}{19} \approx 0.4737$$

Eq. (8-33) in p. 456, Textbook: $\dfrac{\hat{\lambda}_i}{1 + z(\alpha/2)\sqrt{2/n}} \le \lambda_i \le \dfrac{\hat{\lambda}_i}{1 - z(\alpha/2)\sqrt{2/n}}$.

Hnece the 95% confidence interval for the first eigenvalue is

$$\left(\frac{\hat{\lambda}_1}{1+z(0.025)\sqrt{2/200}}, \frac{\hat{\lambda}_1}{1-z(0.025)\sqrt{2/200}}\right) \approx \left(\frac{9}{1.196}, \frac{9}{0.804}\right) \approx (7.5251, 11.1940).$$

(10%)

----

8. From Equations (9-12) and (9-13) in p. 489

$$\mathbf{L} = \sqrt{\lambda_1}\mathbf{e}_1 = \sqrt{\frac{9}{5}}\begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$$

$$\mathbf{\Sigma} \approx \mathbf{LL}'+\mathbf{\Psi} = \frac{9}{5}\begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}\begin{bmatrix} 2 & 0 & -1 \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{bmatrix} = \begin{bmatrix} 7.2+\psi_1 & 0 & -3.6 \\ 0 & \psi_2 & 0 \\ -3.6 & 0 & 1.8+\psi_3 \end{bmatrix},.$$

$$\therefore \psi_1 = 8-7.2 = 0.8, \psi_2 = 6, \psi_3 = 5-1.8 = 3.2$$

$$\mathbf{\Psi} = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3.2 \end{bmatrix}, \text{residual} = \mathbf{\Sigma} - \mathbf{LL}' - \mathbf{\Psi} = \begin{bmatrix} 0 & 0 & 1.6 \\ 0 & 0 & 0 \\ 1.6 & 0 & 0 \end{bmatrix}$$

(10%)

----

9. Use (9-27) in p. 496, the communality for the skull length is

$$\hat{h}_1^2 = \hat{\ell}_{11}^2 + \hat{\ell}_{12}^2 = 0.602^2 + 0.200^2 = 0.4024$$

The proportion of variance explained by the first factor, according to (9-28) is

$$\frac{\hat{\ell}_{11}^2 + \hat{\ell}_{21}^2 + \cdots + \hat{\ell}_{61}^2}{6} = \frac{4.0011}{6} = 0.6669$$

(10%)

----

10. Dependent on your opinions. (10%)