UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2016/2017 Academic Session

June 2017

MSG 466 - Multivariate Analysis
[Analisis Multivariat]

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of TWENTY-ONE pages of printed material before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi DUA PULUH SATU muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]

Instructions: Answer **all eight** [8] questions.

*[**Arahan:** Jawab **semua lapan** [8] soalan.]*

In the event of any discrepancies, the English version shall be used.

[Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi Bahasa Inggeris hendaklah diguna pakai].

<u>Question 1</u>

(a) What are the three properties that must be satisfied by a distance measure, $d\,(P, Q)$ between two points $P = (x_1,\ x_2, \ldots, x_p)$ and $Q = (y_1,\ y_2, \ldots, y_p)$?

(b) The distance from the point $P = (x_1, x_2)$ to the point $Q = (y_1, y_2)$ is defined as

$$d(P,Q) = \max\left(\,|x_1 - y_1|,\ |x_2 - y_2|\,\right).$$

(i) Does this measure of distance satisfy the three conditions that have been stated in (a) above? Justify your answer.

(ii) Compute the distance from the point $P = (1, -3)$ to the point $Q = (-2, -1)$.

(iii) Plot the locus of points whose distance from $P = (1, -3)$ is 2.

[20 marks]

<u>Soalan 1</u>

(a) Apakah tiga sifat yang mesti dipenuhi bagi suatu ukuran jarak, $d\,(P, Q)$ antara dua titik $P = (x_1,\ x_2, \ldots, x_p)$ dan $Q = (y_1,\ y_2, \ldots, y_p)$?

(b) Jarak dari titik $P = (x_1, x_2)$ ke titik $Q = (y_1, y_2)$ ditakrifkan sebagai

$$d(P,Q) = \max\left(\,|x_1 - y_1|,\ |x_2 - y_2|\,\right).$$

(i) Adakah ukuran jarak ini memenuhi ketiga-tiga syarat yang dinyatakan dalam (a) di atas? Tentusahkan jawapan anda.

(ii) Kira jarak antara titik $P = (1, -3)$ ke titik $Q = (-2, -1)$.

(iii) Lakar lokus titik-titik yang jaraknya dari $P = (1, -3)$ adalah 2.

[20 markah]

<u>Question 2</u>

(a) Suppose $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Based on a random sample of size 30, the sample mean vector and the sample variance-covariance matrix are obtained as below:

$$\bar{\mathbf{x}} = \begin{bmatrix} 9 \\ 4 \end{bmatrix} \text{ and } \mathbf{S} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

(i) Is $\mathbf{S}$ a positive definite matrix?

(ii) Determine the generalized sample variance and total sample variance.

(iii) Find the maximum likelihood estimates of the 2×1 mean vector $\boldsymbol{\mu}$ and the 2×2 variance-covariance matrix $\boldsymbol{\Sigma}$.

(b) Let **X** be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\mathbf{X}=\begin{bmatrix}X_1\\X_2\\X_3\end{bmatrix},\ \boldsymbol{\mu}=\begin{bmatrix}4\\-3\\2\end{bmatrix} \text{ and } \boldsymbol{\Sigma}=\begin{bmatrix}9&0&-2\\0&4&1\\-2&1&1\end{bmatrix}.$$

(i) Find the distributions of $\begin{bmatrix}X_1\\X_3\end{bmatrix}$.

(ii) Determine the conditional distribution of X_2 given that $X_1=x_2$ and $X_3=x_3$.

[30 marks]

Soalan 2

(a) *Andaikan* $\mathbf{X}:\ \mathrm{N}_2(\boldsymbol{\mu},\boldsymbol{\Sigma})$. *Berdasarkan suatu sampel rawak bersaiz 30, vektor min sampel dan matriks varians-kovarians sampel diperoleh seperti di bawah:*

$$\bar{\mathbf{x}}=\begin{bmatrix}9\\4\end{bmatrix} \textit{ dan } \mathbf{S}=\begin{bmatrix}3&1\\1&2\end{bmatrix}$$

(i) *Adakah* **S** *matriks tentu positif?*

(ii) *Tentukan varians sampel teritlak dan varians sampel keseluruhan.*

(iii) *Cari anggaran kebolehjadian maksimum bagi vektor min* 2×1, $\boldsymbol{\mu}$ *dan matriks varians-kovarians* 2×2, $\boldsymbol{\Sigma}$.

(b) *Biar* ***X*** *sebagai* $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *yang mana*

$$\mathbf{X}=\begin{bmatrix}X_1\\X_2\\X_3\end{bmatrix},\ \boldsymbol{\mu}=\begin{bmatrix}4\\-3\\2\end{bmatrix} \textit{ dan } \boldsymbol{\Sigma}=\begin{bmatrix}9&0&-2\\0&4&1\\-2&1&1\end{bmatrix}.$$

(i) *Cari taburan bagi* $\begin{bmatrix}X_1\\X_3\end{bmatrix}$.

(ii) *Tentukan taburan bersyarat bagi* X_2 *diberi bahawa* $X_1=x_2$ *dan* $X_3=x_3$.

[30 markah]

<u>Question 3</u>

Let $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ be mutually independent and identically distributed 3×1 random vectors from $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = (3, 1, 2)' \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 3 \\ 0 & 3 & 9 \end{pmatrix}.$$

(a) Find the marginal distribution for each of the random vectors

$$\mathbf{V}_1 = \frac{1}{3}\mathbf{X}_1 + \frac{1}{3}\mathbf{X}_2 + \frac{1}{3}\mathbf{X}_3$$

and

$$\mathbf{V}_2 = \mathbf{X}_1 - 2\mathbf{X}_2 + \mathbf{X}_3.$$

(b) Are $\mathbf{V}_1$ and $\mathbf{V}_2$ independent? Justify your answer.

(c) Specify the joint distribution of $\mathbf{V}_1$ and $\mathbf{V}_2$.

[25 marks]

<u>Soalan 3</u>

Biar $\mathbf{X}_1$, $\mathbf{X}_2$ *dan* $\mathbf{X}_3$ *sebagai vektor rawak* 3×1 *bertaburan saling tak bersandar dan secaman daripada* $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *yang mana*

$$\boldsymbol{\mu} = (3, 1, 2)' \textit{ dan } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 3 \\ 0 & 3 & 9 \end{pmatrix}.$$

(a) Cari taburan sut bagi setiap vektor rawak

$$\mathbf{V}_1 = \frac{1}{3}\mathbf{X}_1 + \frac{1}{3}\mathbf{X}_2 + \frac{1}{3}\mathbf{X}_3$$

dan

$$\mathbf{V}_2 = \mathbf{X}_1 - 2\mathbf{X}_2 + \mathbf{X}_3.$$

(b) Adakah $\mathbf{V}_1$ *dan* $\mathbf{V}_2$ *tak bersandar? Tentusahkan jawapan anda.*

(c) Perincikan taburan tercantum $\mathbf{V}_1$ *dan* $\mathbf{V}_2$.

[25 markah]

Question 4

Mandible measurements were made on eight male modern dogs from Thailand. Two of the measurements are length of first molar (in mm) and breadth of first molar (in mm), which are assumed to have a bivariate normal distribution. The sample mean vector and the sample variance-covariance matrix are as follows:

$$\bar{\mathbf{x}}_m = \begin{bmatrix} 19.8 \\ 7.8 \end{bmatrix} \text{ and } \mathbf{S}_m = \begin{bmatrix} 0.79 & 0.19 \\ 0.19 & 0.08 \end{bmatrix}$$

(a) Test $H_o: \boldsymbol{\mu}_m = (18.5, 8.5)'$ at $\alpha = 0.05$. State any assumptions that you have made and give your conclusion from this test.

(b) Similar measurements are also made for eight female dogs of the same species. The sample mean vector and the sample variance-covariance matrix are given by

$$\bar{\mathbf{x}}_f = \begin{bmatrix} 19.0 \\ 7.5 \end{bmatrix} \text{ and } \mathbf{S}_f = \begin{bmatrix} 0.57 & 0.26 \\ 0.26 & 0.37 \end{bmatrix}.$$

Is there evidence of differences between the mandible size of the male and the female dogs? Perform an appropriate test and state your assumptions.

[20 marks]

Soalan 4

Ukuran-ukuran rahang dibuat pada lapan ekor anjing jantan moden daripada Thailand. Dua daripada ukuran adalah panjang molar pertama (dalam mm) dan lebar molar pertama (dalam mm), yang dianggap mempunyai taburan normal bivariat. Vektor min sampel dan matriks varians-kovarians sampel adalah seperti berikut:

$$\bar{\mathbf{x}}_m = \begin{bmatrix} 19.8 \\ 7.8 \end{bmatrix} \textit{ dan } \mathbf{S}_m = \begin{bmatrix} 0.79 & 0.19 \\ 0.19 & 0.08 \end{bmatrix}$$

(a) Uji $H_o: \boldsymbol{\mu}_m = (18.5, 8.5)'$ pada $\alpha = 0.05$. Nyatakan sebarang andaian yang telah anda buat dan berikan kesimpulan daripada ujian ini.

(b) Ukuran yang sama juga telah dibuat pada lapan ekor anjing betina daripada spesies yang sama. Vektor min sampel dan matriks varians-kovarians sampelnya diberi oleh

$$\bar{\mathbf{x}}_f = \begin{bmatrix} 19.0 \\ 7.5 \end{bmatrix} \textit{ dan } \mathbf{S}_f = \begin{bmatrix} 0.57 & 0.26 \\ 0.26 & 0.37 \end{bmatrix}.$$

Adakah terdapat bukti perbezaan saiz rahang antara anjing jantan dan anjing betina? Lakukan ujian yang bersesuaian dan nyatakan andaian-andaian anda.

[20 markah]

Question 5

The table below displays the edited data of four measurements made on 40 male skulls from the area of Thebes in Egypt. There are two samples of 20 skulls each from the early predynastic period and the Roman period. For each skull, the measurements (in mm) are:

X_1 = maximum breadth
X_2 = basibregmatic height
X_3 = basialveolar length
X_4 = nasal height

Skull	X_1	X_2	X_3	X_4
1	131	138	89	49
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
20	132	131	101	49
21	137	123	91	50
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
40	145	129	89	47

Skulls 1 to 20 are from the early predynastic period while skulls 21 to 40 are from the Roman period. Statistical analyses using MINITAB have been performed for this dataset and the output is displayed in Appendix A.

(a) Based on the output, in your opinion, what analyses have been performed?

(b) How are the four measurements related?

(c) Do the skulls from the early predynastic and Roman periods have statistically significant difference for their mean values of the variables? Discuss your answer.

(d) Suppose, a skull is just found from the area and the four measurements on the skull are (126, 133, 102, 51). Using the information in the output, determine whether the skull is most likely from the early predynastic or Roman period. Discuss your answer.

[25 marks]

Soalan 5

Jadual di bawah mempamerkan data yang telah disunting bagi empat ukuran yang dibuat pada 40 tengkorak lelaki daripada kawasan Thebes di Mesir. Terdapat 2 sampel dengan 20 tengkorak setiap satunya daripada zaman awal pradinastik dan zaman Roman. Bagi setiap tengkorak, ukuran-ukurannya (dalam mm) adalah:

X_1 = *lebar maksimum*
X_2 = *tinggi basibregmatik*
X_3 = *panjang basialveolar*
X_4 = *tinggi hidung*

Tengkorak	X_1	X_2	X_3	X_4
1	*131*	*138*	*89*	*49*
⋮	⋮	⋮	⋮	⋮
20	*132*	*131*	*101*	*49*
21	*137*	*123*	*91*	*50*
⋮	⋮	⋮	⋮	⋮
40	*145*	*129*	*89*	*47*

Tengkorak 1 ke 20 adalah daripada zaman awal pradinastik sementara tengkorak 21 ke 40 adalah daripada zaman Roman. Analisis statistik menggunakan MINITAB telah dijalankan pada set data ini dan outputnya dipamerkan dalam Lampiran A.

(a) *Berdasarkan output, pada pendapat anda, analisis apa yang telah dijalankan?*

(b) *Bagaimana keempat-empat ukuran ini berkait?*

(c) *Adakah tengkorak pada zaman awal pradinastik dan tengkorak pada zaman Roman mempunyai perbezaan bererti secara statistik bagi nilai min pembolehubah-pembolehubah mereka? Bincangkan jawapan anda.*

(d) *Andaikan suatu tengkorak baru sahaja dijumpai daripada kawasan itu dan empat ukuran pada tengkorak adalah (126, 133, 102, 51). Menggunakan maklumat daripada output, tentukan sama ada tengkorak itu berkemungkinan besar daripada zaman awal pradinastik atau daripada zaman Roman. Bincangkan jawapan anda.*

[25 markah]

Question 6

(a) Answer the following questions on Factor Analysis:

(i) What is the purpose of the analysis?

(ii) How does it differ from Principal Component Analysis?

(iii) How are the variables grouped?

(iv) What is the purpose of factor rotation?

(b) A set of data consists of 130 observations generated by scores on a psychological test administered to Peruvian teenagers (see table below). The scores were accumulated into five subscale scores labeled independence (indep), support (supp), benevolence (benev), conformity (conform) and leadership (leader).

Indep	Supp	Benev	Conform	Leader
27	13	14	20	11
12	13	24	25	6
⋮	⋮	⋮	⋮	⋮
27	19	22	7	9
10	17	22	22	8

Factor analysis was performed on this data set using the principal component method and the maximum likelihood method. The output of the analysis is displayed in Appendix B. Interpret and discuss the results.

[25 marks]

<u>Soalan 6</u>

(a) Jawab soalan-soalan berkenaan Analisis Faktor berikut:

(i) Apakah tujuan analisis tersebut?

(ii) Bagaimana ia berbeza daripada Analisis Komponen Utama?

(iii) Bagaimana pembolehubah-pembolehubah dikelaskan?

(iv) Apakah tujuan putaran faktor?

(b) Suatu set data mengandungi 130 cerapan yang dijana daripada skor suatu ujian psikologi terhadap remaja-remaja Peru (lihat jadual di bawah). Skor-skor dikumpul dalam lima skor subskala berlabel berdikari (indep), sokongan (supp), kebajikan (benev), pematuhan (conform) dan kepimpinan (leader).

Indep	*Supp*	*Benev*	*Conform*	*Leader*
27	*13*	*14*	*20*	*11*
12	*13*	*24*	*25*	*6*
⋮	⋮	⋮	⋮	⋮
27	*19*	*22*	*7*	*9*
10	*17*	*22*	*22*	*8*

Analisis faktor telah dijalankan pada set data ini menggunakan kaedah komponen utama dan kaedah kebolehjadian maksimum. Output analisis dipamerkan dalam Lampiran B. Tafsir dan bincang keputusannya.

[25 markah]

Question 7

The percentages of the labour force in three different types of industry (agriculture, mining and manufacturing) for five European countries are shown in the following table

Country	AGR	MIN	MAN
Belgium	2.6	0.2	20.8
Romania	22.0	2.6	37.9
Switzerland	5.6	0.0	24.7
France	5.1	0.3	20.2
Bulgaria	19.0	0.0	35.0

(a) Determine the distance matrix for the data using the city-block metric.

(b) Cluster the five countries using the complete linkage hierarchical procedure. Draw a dendogram and discuss your results.

[30 marks]

Soalan 7

Peratusan tenaga kerja dalam tiga jenis industri berbeza (pertanian, perlombongan dan pembuatan) bagi lima negara Eropah ditunjukkan dalam jadual di bawah :

Negara	*AGR*	*MIN*	*MAN*
Belgium	*2.6*	*0.2*	*20.8*
Romania	*22.0*	*2.6*	*37.9*
Switzerland	*5.6*	*0.0*	*24.7*
Perancis	*5.1*	*0.3*	*20.2*
Bulgaria	*19.0*	*0.0*	*35.0*

(a) Tentukan matriks jarak bagi data menggunakan metrik blok-bandar.

(b) Klusterkan kelima-lima negara tersebut menggunakan tatacara berhierarki pautan lengkap. Lukis dendogram dan bincang keputusan anda.

[30 markah]

Question 8

Observations on two responses are collected for four treatments. The summary statistics of the two variables for Treatment 1, Treatment 2, Treatment 3 and Treatment 4 are as follows:

Treatment 1: $\bar{\mathbf{x}}_1 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$, $\mathbf{S}_1 = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$, $n_1 = 18$

Treatment 2: $\bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, $\mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $n_2 = 13$

Treatment 3: $\bar{\mathbf{x}}_3 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, $\mathbf{S}_3 = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}$, $n_3 = 20$

Treatment 4: $\bar{\mathbf{x}}_4 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $\mathbf{S}_4 = \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}$, $n_4 = 14$

Construct a one-way MANOVA table and test for treatment effects using $\alpha = 0.05$. Give your conclusion and state any assumptions that you have made.

[25 marks]

Soalan 8

Cerapan daripada dua respon dikumpul bagi empat rawatan. Statistik ringkas daripada dua pembolehubah bagi Rawatan 1, Rawatan 2, Rawatan 3 dan Rawatan 4 adalah seperti berikut:

Rawatan 1: $\bar{\mathbf{x}}_1 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$, $\mathbf{S}_1 = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$, $n_1 = 18$

Rawatan 2: $\bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, $\mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $n_2 = 13$

Rawatan 3: $\bar{\mathbf{x}}_3 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, $\mathbf{S}_3 = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}$, $n_3 = 20$

Rawatan 4: $\bar{\mathbf{x}}_4 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $\mathbf{S}_4 = \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}$, $n_4 = 14$

Bina jadual MANOVA satu-hala dan uji kesan rawatan menggunakan $\alpha = 0.05$. *Berikan kesimpulan anda dan nyatakan sebarang andaian yang anda telah buat.*

[25 markah]

Appendix/Lampiran A

***Period: Early Predynastic (1), Roman (2)**

Descriptive Statistics: X1, X2, X3, X4

Variable	Period	Total Count	Mean	Variance	Minimum	Maximum
X1	1	20	132.00	29.68	119.00	141.00
	2	20	135.90	28.73	126.00	145.00
X2	1	20	134.10	21.88	121.00	143.00
	2	20	130.90	28.52	120.00	138.00
X3	1	20	97.80	29.43	89.00	109.00
	2	20	94.350	15.292	86.000	101.000
X4	1	20	50.500	7.105	44.000	56.000
	2	20	51.700	12.642	45.000	58.000

Correlations: X1, X2, X3, X4

	X1	X2	X3
X2	-0.029		
	0.861		
X3	-0.068	0.207	
	0.678	0.201	
X4	0.221	0.240	0.060
	0.170	0.136	0.714

```
Cell Contents: Pearson correlation
               P-Value
```

ANOVA: X1, X2, X3, X4 versus Period

Factor	Type	Levels	Values
Period	fixed	2	1, 2

Analysis of Variance for X1

Source	DF	SS	MS	F	P
Period	1	152.10	152.10	5.21	0.028
Error	38	1109.80	29.21		
Total	39	1261.90			

S = 5.40419   R-Sq = 12.05%   R-Sq(adj) = 9.74%

Analysis of Variance for X2

Source	DF	SS	MS	F	P
Period	1	102.40	102.40	4.06	0.051
Error	38	957.60	25.20		
Total	39	1060.00			

S = 5.01996   R-Sq = 9.66%   R-Sq(adj) = 7.28%

Analysis of Variance for X3

Source	DF	SS	MS	F	P
Period	1	119.03	119.03	5.32	0.027
Error	38	849.75	22.36		
Total	39	968.78			

S = 4.72883   R-Sq = 12.29%   R-Sq(adj) = 9.98%

Analysis of Variance for X4

Source	DF	SS	MS	F	P
Period	1	14.400	14.400	1.46	0.235
Error	38	375.200	9.874		
Total	39	389.600			

S = 3.14224   R-Sq = 3.70%   R-Sq(adj) = 1.16%

MANOVA for Period
s = 1   m = 1.0   n = 16.5

Criterion	Test Statistic	F	DF Num	DF Denom	P
Wilks'	0.67173	4.276	4	35	0.006
Lawley-Hotelling	0.48869	4.276	4	35	0.006
Pillai's	0.32827	4.276	4	35	0.006
Roy's	0.48869				

SSCP Matrix for Period

	X1	X2	X3	X4
X1	152.1	-124.8	-134.6	46.80
X2	-124.8	102.4	110.4	-38.40
X3	-134.6	110.4	119.0	-41.40
X4	46.8	-38.4	-41.4	14.40

SSCP Matrix for Error

	X1	X2	X3	X4
X1	1109.80	91.80	59.70	108.40
X2	91.80	957.60	99.10	192.40
X3	59.70	99.10	849.75	78.10
X4	108.40	192.40	78.10	375.20

Partial Correlations for the Error SSCP Matrix

	X1	X2	X3	X4
X1	1.00000	0.08905	0.06148	0.16799
X2	0.08905	1.00000	0.10986	0.32098
X3	0.06148	0.10986	1.00000	0.13832
X4	0.16799	0.32098	0.13832	1.00000

Discriminant Analysis: Period versus X1, X2, X3, X4

Linear Method for Response: Period

Predictors: X1, X2, X3, X4

Group	1	2
Count	20	20

Summary of classification

	True Group	
Put into Group	1	2
1	17	3
2	3	17
Total N	20	20
N correct	17	17
Proportion	0.850	0.850

N = 40 N Correct = 34 Proportion Correct = 0.850

Squared Distance Between Groups

	1	2
1	0.00000	1.85703
2	1.85703	0.00000

Linear Discriminant Function for Groups

	1	2
Constant	-746.05	-737.21
X1	3.87	4.01
X2	4.38	4.22
X3	3.50	3.33
X4	1.02	1.22

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
6**	1	2	1	8.422	0.124
			2	4.520	0.876
12**	1	2	1	10.172	0.080
			2	5.291	0.920
18**	1	2	1	2.108	0.370
			2	1.040	0.630
26**	2	1	1	2.564	0.931
			2	7.775	0.069
28**	2	1	1	5.709	0.640
			2	6.861	0.360
36**	2	1	1	3.866	0.550
			2	4.267	0.450

Appendix/Lampiran B

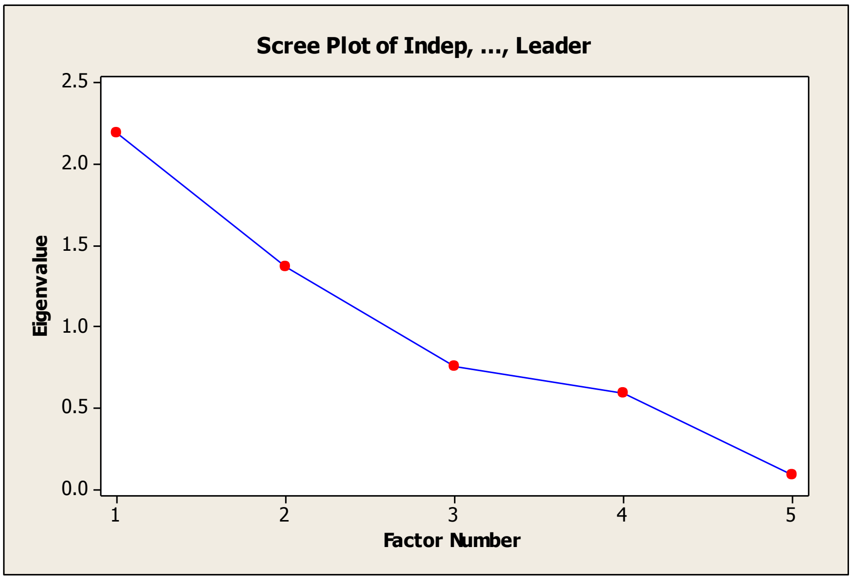
Descriptive Statistics: Indep, Supp, Benev, Conform, Leader

Variable	Total Count	Mean	StDev	Variance
Indep	130	15.669	5.895	34.750
Supp	130	17.077	4.185	17.513
Benev	130	18.785	5.463	29.845
Conform	130	15.500	5.748	33.043
Leader	130	11.731	5.192	26.958

Correlations: Indep, Supp, Benev, Conform, Leader

	Indep	Supp	Benev	Conform
Supp	-0.173 0.049			
Benev	-0.561 0.000	0.018 0.836		
Conform	-0.471 0.000	-0.327 0.000	0.298 0.001	
Leader	0.187 0.033	-0.401 0.000	-0.492 0.000	-0.333 0.000

Cell Contents: Pearson correlation
P-Value



Factor Analysis: Indep, Supp, Benev, Conform, Leader

Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
Indep	-0.772	0.101	0.606
Supp	0.180	0.922	0.882
Benev	0.813	-0.009	0.660
Conform	0.651	-0.574	0.753
Leader	-0.696	-0.422	0.662
Variance	2.1966	1.3682	3.5649
% Var	0.439	0.274	0.713

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
Indep	0.775	0.076	0.606
Supp	0.033	-0.939	0.882
Benev	-0.794	-0.174	0.660
Conform	-0.764	0.413	0.753
Leader	0.583	0.568	0.662
Variance	2.1544	1.4105	3.5649
% Var	0.431	0.282	0.713

Factor Score Coefficients

Variable	Factor1	Factor2
Indep	0.359	0.007
Supp	0.072	-0.675
Benev	-0.362	-0.077
Conform	-0.383	0.342
Leader	0.239	0.372

Factor Analysis: Indep, Supp, Benev, Conform, Leader

Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Communality
Indep	-0.772	0.101	-0.580	0.943
Supp	0.180	0.922	0.163	0.909
Benev	0.813	-0.009	0.100	0.670
Conform	0.651	-0.574	-0.256	0.819
Leader	-0.696	-0.422	0.563	0.979
Variance	2.1966	1.3682	0.7559	4.3207
% Var	0.439	0.274	0.151	0.864

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Factor3	Communality
Indep	-0.971	0.018	-0.003	0.943
Supp	0.136	-0.312	0.890	0.909
Benev	0.700	-0.418	-0.081	0.670
Conform	0.419	-0.379	-0.707	0.819
Leader	-0.155	0.971	-0.111	0.979

Variance	1.6506	1.3587	1.3114	4.3207
% Var	0.330	0.272	0.262	0.864

Factor Score Coefficients

Variable	Factor1	Factor2	Factor3
Indep	-0.752	-0.362	-0.147
Supp	0.119	-0.129	0.690
Benev	0.372	-0.127	-0.010
Conform	0.073	-0.277	-0.545
Leader	0.240	0.832	0.008

Factor Analysis: Indep, Supp, Benev, Conform, Leader

Maximum Likelihood Factor Analysis of the Correlation Matrix

* NOTE * Heywood case

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
Indep	0.679	0.173	0.492
Supp	0.000	-1.000	1.000
Benev	-0.689	-0.018	0.475
Conform	-0.668	0.327	0.553
Leader	0.529	0.401	0.441
Variance	1.6620	1.2985	2.9605
% Var	0.332	0.260	0.592

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
Indep	0.679	0.173	0.492
Supp	0.000	-1.000	1.000
Benev	-0.689	-0.018	0.475
Conform	-0.668	0.327	0.553
Leader	0.529	0.401	0.441
Variance	1.6620	1.2985	2.9605
% Var	0.332	0.260	0.592

Factor Score Coefficients

Variable	Factor1	Factor2
Indep	0.310	0.000
Supp	0.034	-1.000
Benev	-0.305	0.000
Conform	-0.347	-0.000
Leader	0.219	0.000

* WARNING * Too many factors, solution is not unique

Factor Analysis: Indep, Supp, Benev, Conform, Leader

Maximum Likelihood Factor Analysis of the Correlation Matrix

* NOTE * Heywood case

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Communality
Indep	-0.788	0.187	0.587	1.000
Supp	-0.464	-0.401	-0.790	1.000
Benev	0.532	-0.492	-0.086	0.532
Conform	0.664	-0.333	0.194	0.589
Leader	0.000	1.000	0.000	1.000
Variance	1.5591	1.5486	1.0133	4.1211
% Var	0.312	0.310	0.203	0.824

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Factor3	Communality
Indep	-0.992	0.034	0.122	1.000
Supp	0.048	-0.192	-0.980	1.000
Benev	0.562	-0.454	0.098	0.532
Conform	0.515	-0.371	0.432	0.589
Leader	-0.129	0.968	0.213	1.000
Variance	1.5842	1.3199	1.2170	4.1211
% Var	0.317	0.264	0.243	0.824

Factor Score Coefficients

Variable	Factor1	Factor2	Factor3
Indep	-1.016	-0.130	-0.024
Supp	-0.123	0.219	-1.069
Benev	-0.000	0.000	0.000
Conform	-0.000	0.000	-0.000
Leader	0.011	1.081	-0.211

Appendix/Lampiran C

FORMULAE SHEET

1. Suppose $\mathbf{X}$ has $E(\mathbf{X})=\boldsymbol{\mu}$ and $\mathrm{Cov}(\mathbf{X})=\boldsymbol{\Sigma}$. Thus $\mathbf{c}'\mathbf{X}$ has mean $\mathbf{c}'\boldsymbol{\mu}$ and variance $\mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}$.

2. Bivariate normal p.d.f.:

$$f(x_1,x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}\left(1-\rho_{12}{}^2\right)}} \exp\left\{-\frac{1}{2\left(1-\rho_{12}{}^2\right)} \left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

3. Multivariate normal p.d.f.:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\,|\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})/2}$$

4. If $\mathbf{X} : \mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, then
   (a) $\mathbf{a}'\mathbf{X} : N\left(\mathbf{a}'\boldsymbol{\mu},\ \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}\right)$
   (b) $\mathbf{AX} : \mathrm{N}_q\left(\mathbf{A}\boldsymbol{\mu},\ \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)$
   (c) $\mathbf{X}+\mathbf{d} : \mathrm{N}_p\left(\boldsymbol{\mu}+\mathbf{d},\ \boldsymbol{\Sigma}\right)$
   (d) $\mathbf{AX}+\mathbf{d} : \mathrm{N}_q\left(\mathbf{A}\boldsymbol{\mu}+\mathbf{d},\ \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)$
   (e) $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) : \chi^2_p$

5. Let $\mathbf{X}_j : \mathrm{N}_p(\boldsymbol{\mu}_j,\boldsymbol{\Sigma}),\ j=1,...,n$ be mutually independent. Then $\mathbf{V}_1=\sum_{j=1}^{n} c_j\mathbf{X}_j : \mathrm{N}_p\left(\sum_{j=1}^{n} c_j\boldsymbol{\mu}_j,\ \left(\sum_{j=1}^{n} c_j{}^2\right)\boldsymbol{\Sigma}\right)$. Moreover, $\mathbf{V}_1$ and $\mathbf{V}_2=\sum_{j=1}^{n} b_j\mathbf{X}_j$ are jointly multivariate normal with covariance matrix $\begin{pmatrix} \left(\sum_{j=1}^{n} c_j{}^2\right)\boldsymbol{\Sigma} & (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} \\ (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} & \left(\sum_{j=1}^{n} b_j{}^2\right)\boldsymbol{\Sigma} \end{pmatrix}$.

6. Let $\mathbf{X}=\begin{bmatrix}\mathbf{X}_1\\ \hdashline \mathbf{X}_2\end{bmatrix}$ be distributed as $\mathbf{X} : \mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ with $\boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}_1\\ \hdashline \boldsymbol{\mu}_2\end{bmatrix}$, $\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$ and $|\boldsymbol{\Sigma}_{22}|>0$. Then the conditional distribution of $\mathbf{X}_1$, given that $\mathbf{X}_2=\mathbf{x}_2$, is normal and has mean = $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}{}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)$ and covariance = $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}{}^{-1}\boldsymbol{\Sigma}_{21}$.

7. One-sample results:

(a) $T^2 = n\left(\bar{\mathbf{X}}-\boldsymbol{\mu}\right)'\mathbf{S}^{-1}\left(\bar{\mathbf{X}}-\boldsymbol{\mu}\right)$

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j, \quad \mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}\left(\mathbf{X}_j-\bar{\mathbf{X}}\right)\left(\mathbf{X}_j-\bar{\mathbf{X}}\right)'$$

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

(b) $100(1-\alpha)\%$ simultaneous confidence intervals for $\mathbf{a}'\boldsymbol{\mu}$:

$$\mathbf{a}'\bar{\mathbf{X}} \quad \pm \quad \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)\,\mathbf{a}'\mathbf{S}\mathbf{a}}$$

(c) $100(1-\alpha)\%$ Bonferroni confidence interval for μ_i:

$$\bar{x}_i \quad \pm \quad t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{ii}}{n}}$$

(d) $100(1-\alpha)\%$ large sample confidence interval for μ_i:

$$\bar{x}_i \quad \pm \quad \sqrt{\chi_p^2(\alpha)}\,\sqrt{\frac{s_{ii}}{n}}$$

8. Two-sample results (Paired comparisons):

(a) $T^2 = n\left(\bar{\mathbf{D}}-\boldsymbol{\delta}\right)'\boldsymbol{\delta}_d^{-1}\left(\bar{\mathbf{D}}-\boldsymbol{\delta}\right)$

$$\bar{\mathbf{D}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{D}_j, \quad \mathbf{S}_d = \frac{1}{n-1}\sum_{j=1}^{n}\left(\mathbf{D}_j-\bar{\mathbf{D}}\right)\left(\mathbf{D}_j-\bar{\mathbf{D}}\right)'$$

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

(b) $100(1-\alpha)\%$ simultaneous confidence intervals for δ_i:

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

9. Two-sample results (Independent samples):

(a) $T^2 = \left[\left(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2\right)-\left(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\right)\right]'\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{S}_p\right]^{-1}\left[\left(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2\right)-\left(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\right)\right]$

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

$$\mathbf{S}_p = \frac{(n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2}{n_1+n_2-2}$$

$$\mathbf{S}_i = \frac{\sum_{j=1}^{n_i}(\mathbf{x}_{ij}-\bar{\mathbf{x}}_i)(\mathbf{x}_{ij}-\bar{\mathbf{x}}_i)'}{n_i-1}$$

(b) $100(1-\alpha)\%$ simultaneous confidence interval for $\mathbf{a}'(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$:

$$\mathbf{a}'(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2) \pm c\sqrt{\mathbf{a}'\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{S}_p\mathbf{a}}$$

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

(c) For large n_1-p and n_2-p , an approximate $100(1-\alpha)\%$ simultaneous confidence interval for $\mathbf{a}'(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$:

$$\mathbf{a}'(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2) \pm c\sqrt{\mathbf{a}'\left(\frac{1}{n_1}\mathbf{S}_1+\frac{1}{n_2}\mathbf{S}_2\right)\mathbf{a}}$$

$$c^2 = \chi^2_p(\alpha)$$

10. One-way MANOVA

$$\mathbf{B} = \sum_{l=1}^{g} n_l\,(\bar{\mathbf{x}}_l-\bar{\mathbf{x}})(\bar{\mathbf{x}}_l-\bar{\mathbf{x}})'$$

$$\mathbf{W} = \sum_{l=1}^{g}\sum_{j=1}^{n_l}(\mathbf{x}_{lj}-\bar{\mathbf{x}}_l)(\mathbf{x}_{lj}-\bar{\mathbf{x}}_l)' = (n_1-1)\mathbf{S}_1+(n_2-1)\mathbf{S}_2+\ldots+(n_g-1)\mathbf{S}_g$$

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

Distribution of Λ^* :

For $p=1,\ g\ge 2$: $\left(\frac{n-g}{g-1}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{g-1,n-g}$

For $p=2,\ g\ge 2$: $\left(\frac{n-g-1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1),2(n-g-1)}$

For $p \geq 1,\ g = 2:\ \left(\frac{n-p-1}{p}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right):\ F_{p,n-p-1}$

For $p \geq 1,\ g = 3:\ \left(\frac{n-p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right):\ F_{2p,\,2(n-p-2)}$

$n = \sum n_l$

11. The Estimated Minimum ECM Rule for two normal populations:

Allocate $\mathbf{x}_0$ to population 1 if

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{p}}^{-1}\,\mathbf{x}_0 \;-\; \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{p}}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \;\geq\; \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

Allocate $\mathbf{x}_0$ to population 2 otherwise.

- ooo O ooo -