UNIVERSITI SAINS MALAYSIA

Peperiksaan Semester Pertama

Sidang 1993/94

Oktober/November 1993

MSG466 - Analisis Multivariat

[Masa: 3 jam]

Jawab **SEMUA** soalan; semua soalan mesti dijawab dalam Bahasa Malaysia. Terdapat EMPAT soalan.

1. (a) Katakan $\underset{\sim}{X}' = (X_1, X_2, X_3)$ tertabur $N_3\left(\underline{\mu}, \underset{\sim}{\Sigma}\right)$, dimana

$$\underset{\sim}{\Sigma} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 4 & 3 \\ 1 & 3 & 2 \end{pmatrix} \text{ dan } \underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$$

Cari taburan bagi $4X_1 - 3X_2 + 2X_3$.

(b) Cari anggaran kebolehjadian maksimum bagi vektor min $\underset{\sim}{\mu}$ dan matriks kovarians $\underset{\sim}{\Sigma}$ berdasarkan pada sampel rawak

$$\underset{\sim}{X} = \begin{pmatrix} 3 & 7 & 4 & 6 \\ 5 & 3 & 7 & 9 \end{pmatrix}$$

daripada suatu populasi normal bivariat.

(c) Kelaskan matriks-matriks berikut sebagai tentu positif atau semi-tentu positif:

(i) $\begin{pmatrix} 4 & 1 & 2 \\ 1 & 4 & -1 \\ 2 & -1 & 4 \end{pmatrix}$ (ii) $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$ (iii) $\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}$

...2/-

(d) Diberikan matriks data

$$\underset{\sim}{X} = \begin{pmatrix} 2 & 3 & 4 \\ 1 & 4 & 1 \\ 2 & -2 & 3 \end{pmatrix}$$

dan gabungan-gabungan linear

$$\underset{\sim}{b}'\underset{\sim}{X} = \begin{pmatrix} 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

dan $$\underset{\sim}{c}'\underset{\sim}{X} = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix},$$

nilaikan min, varians dan kovarians sampel bagi $\underset{\sim}{b}'\underset{\sim}{X}$ dan $\underset{\sim}{c}'\underset{\sim}{X}$.

[100/100]

2. (a) Sampel-sampel peluh daripada 20 orang perempuan yang sihat dikaji. Tiga komponen, X_1 = kadar peluh, X_2 = kandungan kalium dan X_3 = kandungan sodium, disukatkan. Ringkasan data diberikan di bawah:

$$\underset{\sim}{\overline{X}} = \begin{pmatrix} 4.640 \\ 9.965 \\ 45.400 \end{pmatrix} \qquad \underset{\sim}{S} = \begin{pmatrix} 2.879 & -1.810 & 10.002 \\ -1.810 & 3.628 & -5.627 \\ 10.002 & -5.627 & 199.798 \end{pmatrix}$$

Dengan **mengandaikan**

$$\underset{\sim}{S}^{-1} = \begin{pmatrix} .586 & .258 & -.022 \\ .258 & .402 & -.002 \\ -.022 & -.002 & .006 \end{pmatrix},$$

ujikan hipotesis bahawa min populasi ialah $H_0 : \underset{\sim}{\mu}' = (4, 10, 48)$ lawan $H_1 : \underset{\sim}{\mu}' \neq (4, 10, 48)$ pada $\alpha = .10$

Nyatakan anggapan-anggapan yang telah anda menggunakan.

(b) Cerapan-cerapan pada dua balasan, X_1 dan X_2 , dikutip bagi tiga rawatan. Vektor-vektor cerapan $\underset{\sim}{x}' = (x_1\ x_2)$ ialah:

Rawatan 1 : $\begin{pmatrix}3\\9\end{pmatrix} \begin{pmatrix}2\\6\end{pmatrix} \begin{pmatrix}7\\9\end{pmatrix}$

Rawatan 2 : $\begin{pmatrix}4\\0\end{pmatrix} \begin{pmatrix}0\\2\end{pmatrix}$

Rawatan 3 : $\begin{pmatrix}8\\3\end{pmatrix} \begin{pmatrix}9\\1\end{pmatrix} \begin{pmatrix}7\\2\end{pmatrix}$

(i) Bina jadual MANOVA satu hala bagi data ini.
(ii) Nilaikan lambda Wilks, Λ^* , dan uji bagi kesan rawatan. Gunakan $\alpha = .01$.
(iii) Ulangkan ujian dengan menggunakan penghampiran khi-kuasa dua dengan pembetulan Bartlett.
(iv) Nyatakan anggapan-anggapan yang telah anda menggunakan.

[100/100]

3. (a) Pertimbangkan matriks korelasi

$$\underset{\sim}{\rho} = \begin{pmatrix}1 & \rho & \rho\\ \rho & 1 & \rho\\ \rho & \rho & 1\end{pmatrix} ,$$

di mana $\rho > 0$.

(i) Dapatkan nilai-nilai eigen dan vektor-vektor eigen bagi $\underset{\sim}{\rho}$.

(ii) Dapatkan komponen-komponen prinsipal bagi $\underset{\sim}{\rho}$.

(b) Pertimbangkan dua set data:

$$\underset{\sim}{X}_1 = \begin{pmatrix}3 & 2 & 4\\ 7 & 4 & 7\end{pmatrix} \text{ dan } \underset{\sim}{X}_2 = \begin{pmatrix}6 & 5 & 4\\ 9 & 7 & 8\end{pmatrix}$$

di mana

$$\bar{\underset{\sim}{x}}_1 = \begin{pmatrix}3\\6\end{pmatrix} , \quad \bar{\underset{\sim}{x}}_2 = \begin{pmatrix}5\\8\end{pmatrix} ,$$

dan $\underset{\sim}{S}_p = \begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}$

(i) Hitungkan fungsi pembezalayan linear

$$y = \left(\bar{\underset{\sim}{x}}_1 - \bar{\underset{\sim}{x}}_2 \right)' \underset{\sim}{S}_p^{-1} \underset{\sim}{x} .$$

(ii) Kelaskan cerapan $\underset{\sim}{x}_0' = (2, 7)$ sebagai dari populasi π_1 atau populasi π_2 dengan menggunakan petua berikut:

Umpukkan $\underset{\sim}{x}_0$ kepada populasi π_1 jika

$$y_0 \geq \hat{m} ,$$

dan umpukkan $\underset{\sim}{x}_0$ kepada populasi π_2 jika

$$y_0 < \hat{m}$$

di mana $y_0 = \left(\bar{\underset{\sim}{x}}_1 - \bar{\underset{\sim}{x}}_2 \right)' \underset{\sim}{S}_p^{-1} \underset{\sim}{x}_0$

dan $\hat{m} = \frac{1}{2}\left(\bar{\underset{\sim}{x}}_1 - \bar{\underset{\sim}{x}}_2 \right)' \underset{\sim}{S}_p^{-1} \left(\bar{\underset{\sim}{x}}_1 + \bar{\underset{\sim}{x}}_2 \right) .$

(c) Korelasi-korelasi sampel bagi saham dari 5 syarikat diberikan di bawah. (Korelasi-korelasi ini dibulatkan kepada dua titik perpuluhan):

	A	B	C	D	E
A	1				
B	.58	1			
C	.51	.60	1		
D	.39	.39	.44	1	
E	.46	.32	.43	.52	1

Dengan menggunakan korelasi-korelasi sampel sebagai sukatan keserupaan, kelompokkan saham-saham tersebut dengan menggunakan prosedur-prosedur hirarki pautan tunggal dan pautan lengkap. Lukiskan dendogram-dendogram dan bandingan hasil-hasil.

[100/100]

...5/-

Bagi setiap bahagian yang berikut, tulis suatu perenggan yang menghuraikan kesimpulan-kesimpulan anda. Output-output bagi setiap bahagian dilampirkan pada akhir soalan ini.

4. (a) Sukatan-sukatan dibuat pada tengkorak orang lelaki negara Mesir daripada kawasan bandar Thebes. Terdapat lima sampel yang terdiri daripada 30 tengkorak daripada setiap zaman, iaitu, dari zaman "early predynastic (circa 4000 BC)", zaman "late predynastic (circa 3300 BC)", zaman "12th. and 13th. dynasties (circa 1800 BC)", zaman "Ptolemaic (circa 200 BC)", dan zaman "Roman (circa AD 150)".

Empat sukatan adalah tersedia bagi setiap tengkorak, iaitu,

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

Semua sukatan adalah dalam mm.

Data tersebut dianalisiskan melalui prosedur, PROC GLM dengan opsyen, MANOVA, daripada pakej SAS.

(b) H. Bumpus (1898) mengkaji morfologi burung-burung pipit yang dikutip selepas suatu angin ribut yang kencang. Beliau mengambil 8 sukatan morfologi pada setiap burung dan juga menimbang burung tersebut. Di sini, kita hanya mempertimbangkan 5 pembolehubah bagi burung betina sahaja. Pembolehubah-pembolehubah yang berkaitan ialah:

X_1 = jumlah panjang (total length).
X_2 = "alar extent"
X_3 = "length of beak and head"
X_4 = "length of humerous"
X_5 = "length of keel of sternum".

Semua sukatan ialah di dalam mm. Terdapat 21 burung yang terus hidup (Kumpulan 1) dan 28 yang mati (Kumpulan 2).

Analisis faktor telah dijalankan dengan menggunakan prosedur PROC FACTOR daripada pakej SAS bagi Kumpulan 1 dan bagi Kumpulan 2.

(c) Industri perikanan Salmon ialah suatu sumber yang berharga bagi kedua-dua negara Amerika Syarikat dan Kanada. Kerana ia ialah suatu sumber yang terhad, ia mesti diuruskan secara cekapnya. Tambahan pula, kerana lebih daripada satu negara dibabitkan, masalah-masalah mesti diselesaikan secara adilnya. Iaitu, penangkap ikan komersial Alaska tidak boleh menangkap terlalu banyak ikan salmon Kanada dan sebaliknya.

Ikan-ikan ini mempunyai suatu kitaran hidup (life cycle) yang istimewa. Ikan-ikan tersebut dilahirkan dalam sungai air tawar dan selepas satu atau dua tahun berenang ke dalam lautan. Selepas dua tahun di dalam air masin, ikan-ikan itu kembali ke tempat lahir mereka untuk beranak dan mati. Pada waktu ikan-ikan itu akan kembali sebagai ikan matang, mereka ditangkap semasa di lautan.

Untuk menolong kawalan tangkapan ikan, sampel-sampel ikan diambil semasa tangkapan mesti dicamkan sebagai daripada air Alaska atau air Kanada. Ikan-ikan itu mempunyai sedikit maklumat mengenai tempat kelahirannya dalam lingkaran pertumbuhan (growth rings) pada sisiknya. Biasanya, lingkaran-lingkaran disekutu dengan pertumbuhan air tawar lebih kecil bagi ikan-ikan dilahirkan di Alaska apabila dibandingkan dengan ikan-ikan dilahirkan di Kanada.

Diameter-diameter rantau lingkaran pertumbuhan (diperbesarkan 100 kali) disukatkan. Pembolehubah-pembolehubah yang sebenarnya dikutip ialah

X_1 = diameter lingkaran bagi pertumbuhan air tawar tahun pertama (dalam 0.01 inci)

X_2 = diameter lingkaran bagi pertumbuhan air laut tahun pertama (dalam 0.01 inci)

Tambahan pula, ikan betina di kod sebagai 1 dan ikan jantan di kod sebagai 2.

Sampel-sampel latihan dengan 50 ikan dilahirkan di Alaska dan 50 ikan dilahirkan di Kanada digunakan.

Data-data dikutip dianalisiskan melalui prosedur, PROC GLM dengan opsyen MANOVA, daripada pakej SAS.

(d) Pertimbangkan bahagian (c). Data tersebut juga dianalisiskan dengan menggunakan prosedur, PROC DISCRIM, daripada pakej SAS.

[100/100]

OUTPUT FOR 4(a):

General Linear Models Procedure
Class Level Information

Class	Levels	Values
PERIOD	5	12th and 13th dy Early predynasti Late predynastic Ptolemic period Roman period (ci

Number of observations in data set = 150

Dependent Variable: X1 maximum breadth (mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	502.8266667	125.7066667	5.95	0.0002
Error	145	3061.0666667	21.1108046		
Corrected Total	149	3563.8933333			

R-Square	C.V.	Root MSE	X1 Mean
0.141089	3.429525	4.594650	133.973333

Dependent Variable: X1 maximum breadth (mm)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
PERIOD	4	502.8266667	125.7066667	5.95	0.0002

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PERIOD	4	502.8266667	125.7066667	5.95	0.0002

Dependent Variable: X2 basibregmatic height (mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	229.9066667	57.4766667	2.45	0.0490
Error	145	3405.2666667	23.4845977		
Corrected Total	149	3635.1733333			

R-Square	C.V.	Root MSE	X2 Mean
0.063245	3.656139	4.846091	132.546667

Dependent Variable: X2 basibregmatic height (mm)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
PERIOD	4	229.9066667	57.4766667	2.45	0.0490

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PERIOD	4	229.9066667	57.4766667	2.45	0.0490

Dependent Variable: X3 basialveolar length (mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	803.2933333	200.8233333	8.31	0.0001
Error	145	3505.9666667	24.1790805		
Corrected Total	149	4309.2600000			

R-Square	C.V.	Root MSE	X3 Mean
0.186411	5.097681	4.917223	96.4600000

Dependent Variable: X3 basialveolar length (mm)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
PERIOD	4	803.2933333	200.8233333	8.31	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PERIOD	4	803.2933333	200.8233333	8.31	0.0001

Dependent Variable: X4 nasal height (mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	61.20000000	15.30000000	1.51	0.2032
Error	145	1472.13333333	10.15264368		
Corrected Total	149	1533.33333333			

R-Square	C.V.	Root MSE	X4 Mean
0.039913	6.255867	3.186321	50.9333333

Dependent Variable: X4 nasal height (mm)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
PERIOD	4	61.20000000	15.30000000	1.51	0.2032

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PERIOD	4	61.20000000	15.30000000	1.51	0.2032

General Linear Models Procedure
Multivariate Analysis of Variance

E = Error SS&CP Matrix

	X1	X2	X3	X4
X1	3061.0666667	5.333333333	11.466666667	291.3
X2	5.333333333	3405.2666667	754	412.53333333
X3	11.466666667	754	3505.9666667	164.33333333
X4	291.3	412.53333333	164.33333333	1472.1333333

Partial Correlation Coefficients from the Error SS&CP Matrix / Prob > |r|

DF = 144	X1	X2	X3	X4
X1	1.000000 0.0	0.001652 0.9842	0.003500 0.9666	0.137224 0.0986
X2	0.001652 0.9842	1.000000 0.0	0.218219 0.0081	0.184251 0.0260
X3	0.003500 0.9666	0.218219 0.0081	1.000000 0.0	0.072335 0.3856

Partial Correlation Coefficients from the Error SS&CP Matrix / Prob > |r|

DF = 144	X1	X2	X3	X4
X4	0.137224 0.0986	0.184251 0.0260	0.072335 0.3856	1.000000 0.0

H = Type III SS&CP Matrix for PERIOD

	X1	X2	X3	X4
X1	502.82666667	-228.1466667	-626.6266667	135.43333333
X2	-228.1466667	229.90666667	292.28	-66.06666667
X3	-626.6266667	292.28	803.29333333	-180.7333333
X4	135.43333333	-66.06666667	-180.7333333	61.2

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SS&CP Matrix for PERIOD   E = Error SS&CP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 X1	X2	X3	X4
0.4250953752	88.23	-0.01051989	0.00307535	0.01205199	-0.00688042
0.0389988979	8.09	0.00321700	0.01744767	-0.00565660	-0.00641882
0.0157044595	3.26	-0.00770399	0.00204030	-0.00122480	0.02446431

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SS&CP Matrix for PERIOD   E = Error SS&CP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 X1	X2	X3	X4
0.0020203462	0.42	0.01236048	-0.00003489	0.01100359	0.00555233

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

S=4   M=-0.5   N=70

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.66358580	3.9009	16	434.4548	0.0001
Pillai's Trace	0.35330557	3.5120	16	580	0.0001
Hotelling-Lawley Trace	0.48181908	4.2310	16	562	0.0001
Roy's Greatest Root	0.42509538	15.4097	4	145	0.0001

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

OUTPUT FOR 4(b):

(i) Output for Birds that Survived:

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 5 Average = 1

	1	2	3	4	5
Eigenvalue	3.253981	0.775736	0.402743	0.348548	0.218993
Difference	2.478245	0.372993	0.054196	0.129555	
Proportion	0.6508	0.1551	0.0805	0.0697	0.0438
Cumulative	0.6508	0.8059	0.8865	0.9562	1.0000

2 factors will be retained by the NFACTOR criterion.

Initial Factor Method: Principal Components

Factor Pattern

	FACTOR1	FACTOR2	
X1	0.85862	0.07172	total length
X2	0.84077	-0.39336	alar extent
X3	0.83239	-0.03840	length of beak and head
X4	0.85831	-0.19468	length of humerus
X5	0.61666	0.75926	length of keel of sternum

Variance explained by each factor

FACTOR1	FACTOR2
3.253981	0.775736

Initial Factor Method: Principal Components

Final Communality Estimates: Total = 4.029716

X1	X2	X3	X4	X5
0.742372	0.861637	0.694348	0.774604	0.956755

Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2
1	0.88288	0.46960
2	-0.46960	0.88288

Rotated Factor Pattern

	FACTOR1	FACTOR2	
X1	0.72438	0.46653	total length
X2	0.92703	0.04754	alar extent
X3	0.75293	0.35699	length of beak and head
X4	0.84921	0.23119	length of humerus
X5	0.18789	0.95992	length of keel of sternum

Rotation Method: Varimax

Variance explained by each factor

FACTOR1	FACTOR2
2.707465	1.322251

Final Communality Estimates: Total = 4.029716

X1	X2	X3	X4	X5
0.742372	0.861637	0.694348	0.774604	0.956755

(ii) Output for Birds that Died:

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 5 Average = 1

	1	2	3	4	5
Eigenvalue	3.801064	0.470988	0.373218	0.242405	0.112325
Difference	3.330076	0.097771	0.130812	0.130080	
Proportion	0.7602	0.0942	0.0746	0.0485	0.0225
Cumulative	0.7602	0.8544	0.9291	0.9775	1.0000

2 factors will be retained by the NFACTOR criterion.

Initial Factor Method: Principal Components

Factor Pattern

	FACTOR1	FACTOR2	
X1	0.87237	0.14779	total length
X2	0.89414	-0.05271	alar extent
X3	0.87310	-0.35831	length of beak and head
X4	0.91472	-0.20678	length of humerus
X5	0.80095	0.52461	length of keel of sternum

Variance explained by each factor

FACTOR1	FACTOR2
3.801064	0.470988

Initial Factor Method: Principal Components

Final Communality Estimates: Total = 4.272052

X1	X2	X3	X4	X5
0.782876	0.802270	0.890694	0.879470	0.916741

Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2
1	0.77932	0.62662
2	-0.62662	0.77932

Rotated Factor Pattern

	FACTOR1	FACTOR2	
X1	0.58725	0.66183	total length
X2	0.72986	0.51921	alar extent
X3	0.90496	0.26786	length of beak and head
X4	0.84244	0.41203	length of humerus
X5	0.29547	0.91074	length of keel of sternum

Rotation Method: Varimax

Variance explained by each factor

FACTOR1	FACTOR2
2.493498	1.778554

Final Communality Estimates: Total = 4.272052

X1	X2	X3	X4	X5
0.782876	0.802270	0.890694	0.879470	0.916741

OUTPUT FOR 4(c):

General Linear Models Procedure
Class Level Information

Class	Levels	Values
SALMON	2	Alaskan Canadian
SEX	2	female male

Number of observations in data set = 100

Dependent Variable: X1 diameter of rings (freshwater)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	38591.26064	12863.75355	43.58	0.0001
Error	96	28338.09936	295.18853		
Corrected Total	99	66929.36000			

R-Square	C.V.	Root MSE	X1 Mean
0.576597	14.57009	17.18105	117.920000

Dependent Variable: X1 diameter of rings (freshwater)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SALMON	1	37805.20673	37805.20673	128.07	0.0001
SEX	1	2.05391	2.05391	0.01	0.9337
SALMON*SEX	1	408.04673	408.04673	1.38	0.2426

Dependent Variable: X2 diameter of rings (marine)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	101611.8637	33870.6212	29.54	0.0001
Error	96	110064.1763	1146.5018		
Corrected Total	99	211676.0400			

R-Square	C.V.	Root MSE	X2 Mean
0.480035	8.504554	33.86003	398.140000

Dependent Variable: X2 diameter of rings (marine)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SALMON	1	100294.7452	100294.7452	87.48	0.0001
SEX	1	356.1185	356.1185	0.31	0.5786
SALMON*SEX	1	1904.7052	1904.7052	1.66	0.2005

Multivariate Analysis of Variance

E = Error SS&CP Matrix

	X1	X2
X1	28338.099359	-3529.349359
X2	-3529.349359	110064.17628

Partial Correlation Coefficients from the Error SS&CP Matrix / Prob > |r|

DF = 95	X1	X2
X1	1.000000	-0.063196
	0.0	0.5386
X2	-0.063196	1.000000
	0.5386	0.0

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SS&CP Matrix for SALMON E = Error SS&CP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 X1	X2
2.1144049338	100.00	0.00449006	-0.00183481
0.0000000000	0.00	0.00390756	0.00239907

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

S=1   M=0   N=46.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.32108863	100.4342	2	95	0.0001
Pillai's Trace	0.67891137	100.4342	2	95	0.0001
Hotelling-Lawley Trace	2.11440493	100.4342	2	95	0.0001
Roy's Greatest Root	2.11440493	100.4342	2	95	0.0001

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SS&CP Matrix for SEX   E = Error SS&CP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 X1	X2
0.0032598447	100.00	-0.00051382	0.00298651
0.0000000000	0.00	0.00593006	0.00045035

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

S=1   M=0   N=46.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.99675075	0.1548	2	95	0.8568
Pillai's Trace	0.00324925	0.1548	2	95	0.8568
Hotelling-Lawley Trace	0.00325984	0.1548	2	95	0.8568
Roy's Greatest Root	0.00325984	0.1548	2	95	0.8568

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SS&CP Matrix for SALMON*SEX   E = Error SS&CP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 X1	X2
0.0338349138	100.00	0.00416036	0.00228909
0.0000000000	0.00	0.00425689	-0.00197030

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

S=1 M=0 N=46.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.96727242	1.6072	2	95	0.2059
Pillai's Trace	0.03272758	1.6072	2	95	0.2059
Hotelling-Lawley Trace	0.03383491	1.6072	2	95	0.2059
Roy's Greatest Root	0.03383491	1.6072	2	95	0.2059

Level of		X1		X2	
SALMON	N	Mean	SD	Mean	SD
Alaskan	50	98.380000	16.1433502	429.660000	37.4043597
Canadian	50	137.460000	18.0579679	366.620000	29.8874692

Level of		X1		X2	
SEX	N	Mean	SD	Mean	SD
female	52	118.057692	27.8767691	396.326923	42.5296764
male	48	117.770833	24.0984149	400.104167	50.3334581

OUTPUT FOR 4(d):

Discriminant Analysis

100 Observations	99 DF Total
2 Variables	98 DF Within Classes
2 Classes	1 DF Between Classes

Class Level Information

SALMON	Frequency	Weight	Proportion	Prior Probability
Alaskan	50	50.0000	0.500000	0.500000
Canadian	50	50.0000	0.500000	0.500000

Discriminant Analysis Within-Class Covariance Matrices

SALMON = Alaskan DF = 49

Variable	X1	X2	
X1	260.607755	-188.092653	diameter of rings (freshwater)
X2	-188.092653	1399.086122	diameter of rings (marine)

SALMON = Canadian DF = 49

Variable	X1	X2	
X1	326.0902041	133.5048980	diameter of rings (freshwater)
X2	133.5048980	893.2608163	diameter of rings (marine)

Pooled Within-Class Covariance Matrix DF = 98

Variable	X1	X2	
X1	293.348980	-27.293878	diameter of rings (freshwater)
X2	-27.293878	1146.173469	diameter of rings (marine)

Discriminant Analysis Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
2	12.7233291

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV^{-1} \, (\bar{X}_i - \bar{X}_j)$$

Generalized Squared Distance to SALMON

From SALMON	Alaskan	Canadian
Alaskan	0	8.29187
Canadian	8.29187	0

Multivariate Statistics and Exact F Statistics

S=1 M=0 N=47.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.32099922	102.5907	2	97	0.0001
Pillai's Trace	0.67900078	102.5907	2	97	0.0001
Hotelling-Lawley Trace	2.11527236	102.5907	2	97	0.0001
Roy's Greatest Root	2.11527236	102.5907	2	97	0.0001

Discriminant Analysis Linear Discriminant Function

$$\text{Constant} = -.5 \, \bar{X}_j' \, COV^{-1} \, \bar{X}_j \qquad \text{Coefficient Vector} = COV^{-1} \, \bar{X}_j$$

SALMON

	Alaskan	Canadian	Label
CONSTANT	-100.68337	-95.14216	
X1	0.37107	0.49946	diameter of rings (freshwater)
X2	0.38370	0.33176	diameter of rings (marine)

Classification Summary for Calibration Data: WORK.SALMON

Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function:

$$D_j^2(X) = (X-\bar{X}_j)' \, COV^{-1} \, (X-\bar{X}_j)$$

Posterior Probability of Membership in each SALMON:

$$Pr(j|X) = \exp(-.5\ D_j^2(X)) \ /\ \text{SUM}_k \exp(-.5\ D_k^2(X))$$

Classification Summary for Calibration Data: WORK.SALMON

Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into SALMON:

From SALMON	Alaskan	Canadian	Total
Alaskan	44 88.00	6 12.00	50 100.00
Canadian	1 2.00	49 98.00	50 100.00
Total Percent	45 45.00	55 55.00	100 100.00

Classification Summary for Calibration Data: WORK.SALMON

Resubstitution Summary using Linear Discriminant Function

Error Count Estimates for SALMON:

	Alaskan	Canadian	Total
Rate	0.1200	0.0200	0.0700
Priors	0.5000	0.5000	

LAMPIRAN

Tatatanda adalah seperti di dalam kuliah.

1. Penguraian spektrum bagi suatu matriks simetrik k x k, $\underset{\sim}{A}$ diberikan oleh

$$\underset{\sim}{A} = \lambda_1 \underset{\sim}{e}_1 \underset{\sim}{e}_1' + \lambda_2 \underset{\sim}{e}_2 \underset{\sim}{e}_2' + \ldots + \lambda_k \underset{\sim}{e}_k \underset{\sim}{e}_k'$$

di mana $\lambda_1, \lambda_2, \ldots, \lambda_k$ adalah nilai-nilai eigen $\underset{\sim}{A}$ dan $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_k$ adalah vektor-vektor eigen terpiawai yang berkaitan.

2. Katakan $\underset{\sim}{X}$ mempunyai $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$ dan Kov $(\underset{\sim}{X}) = \underset{\sim}{\Sigma}$. Maka $\underset{\sim}{c}'\underset{\sim}{X}$ mempunyai min, $\underset{\sim}{c}'\underset{\sim}{\mu}$ dan varians, $\underset{\sim}{c}'\underset{\sim}{\Sigma}\,\underset{\sim}{c}$.

3. f.k.k. normal bivariat:

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

4. f.k.k. normal multivariat:

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

5. Jika $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$, maka $\underset{\sim}{A}\underset{\sim}{X} \sim N_q(\underset{\sim}{A}\underset{\sim}{\mu}, \underset{\sim}{A}\,\underset{\sim}{\Sigma}\,\underset{\sim}{A}')$.

...2/-

6. Satu sampel:

(a) $T^2 = n(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu})' \underset{\sim}{S}^{-1}(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu})$

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

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

(b) Lambda Wilks $\Lambda^{2/n} = |\hat{\underset{\sim}{\Sigma}}| \Big/ |\hat{\underset{\sim}{\Sigma}}_0| = \left[1 + \frac{T^2}{(n-1)}\right]^{-1}$

(c) Selang keyakinan serentak 100(1-α)% bagi $\underset{\sim}{\ell}'\underset{\sim}{\mu}$:

$$\underset{\sim}{\ell}' \bar{\underset{\sim}{X}} \pm \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)\ \underset{\sim}{\ell}'\underset{\sim}{S}\,\underset{\sim}{\ell}}$$

(d) Selang keyakinan serentak Bonferroni 100(1-α)% bagi

μ_i, $i = 1, \ldots, p$:

$$\bar{X}_i + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{S_{ii}}{n}}$$

7. Dua sampel tak bersandar:

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

$$\left[\bar{\underset{\sim}{X}}_1 - \bar{\underset{\sim}{X}}_2 - \left(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2\right)\right]$$

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

...3/-

(b) Selang keyakinan serentak 100(1-α)% bagi $\underset{\sim}{\ell}'(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)$:

$$\underset{\sim}{\ell}'\left[\underset{\sim}{\bar{X}}_1 - \underset{\sim}{\bar{X}}_2\right] \pm c \sqrt{\underset{\sim}{\ell}'\left[\frac{1}{n_1} + \frac{1}{n_2}\right] \underset{\sim}{S}_p \underset{\sim}{\ell}}$$

$$\text{di mana } c^2 = \frac{\left[n_1 + n_2 - 2\right]p}{n_1 + n_2 - p - 1} F_{p,\, n_1 + n_2 - p - 1}$$

8. MANOVA satu-hala:

(a) $$\underset{\sim}{B} = \sum_{\ell=1}^{g} n_\ell (\underset{\sim}{\bar{x}}_\ell - \underset{\sim}{\bar{x}})(\underset{\sim}{\bar{x}}_\ell - \underset{\sim}{\bar{x}})'$$

$$\underset{\sim}{W} = \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} \left[\underset{\sim}{x}_{\ell j} - \underset{\sim}{\bar{x}}_\ell\right]\left[\underset{\sim}{x}_{\ell j} - \underset{\sim}{\bar{x}}_\ell\right]'$$

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

(b) Selang keyakinan serentak 100(1-α)% bagi $\tau_{ki} - \tau_{\ell i}$:

$$\bar{X}_{ki} - \bar{X}_{\ell i} \pm t_{n-g}\left[\frac{\alpha}{pg(g-1)}\right] \sqrt{\frac{W_{ii}}{n-g}\left[\frac{1}{n_k} + \frac{1}{n_\ell}\right]} ,$$

$i = 1, 2, \ldots, p \;,\; \ell < k = 1, 2, \ldots, g$

9. Andaikan $\underset{\sim}{E}$ mempunyai d.k. m_E dan H mempunyai d.k. m_H.

Katakan $\Lambda = \dfrac{|\underset{\sim}{E}|}{|\underset{\sim}{E} + \underset{\sim}{H}|}$.

...4/-

Maka (1) Untuk p = 1,

$$\left(\frac{1 - \Lambda}{\Lambda}\right) \frac{m_E}{m_H} \sim F_{m_H\,,\, m_E} \quad \text{bagi sebarang } m_H .$$

(2) Untuk $m_H = 1$,

$$\left(\frac{1 - \Lambda}{\Lambda}\right) \frac{m_E + 1 - p}{P} \sim F_{P,\; m_E + 1 - p} \quad \text{bagi sebarang P.}$$

(3) Untuk p = 2,

$$\left(\frac{1 - \Lambda^{1/2}}{\Lambda^{1/2}}\right) \left(\frac{m_E - 1}{m_H}\right) \sim F_{2m_H, 2(m_E - 1)}$$

untuk $m_H \geq 2$.

(4) Untuk $m_H = 2$,

$$\left(\frac{1 - \Lambda^{1/2}}{\Lambda^{1/2}}\right) \left(\frac{m_E + 1 - p}{p}\right) \sim F_{2p,\; 2(m_E + 1 - p)}$$

untuk $p \geq 2$.

Pembetulan Bartlett: Katakan $n_o = m_E + m_H$.

Bagi m_E besar,

$$-f \log \Lambda \sim X^2_{pm_H}$$

di mana $f = m_E - \frac{1}{2}\left(p - m_H + 1\right)$

$$= n_o - \frac{1}{2}\left(p + m_H + 1\right)$$

10. <u>MANOVA dua-hala</u>:

$$SSP_{faktor\ 1} = \sum_{\ell=1}^{g} bn\left(\bar{\underset{\sim}{x}}_{\ell .} - \bar{\underset{\sim}{x}}\right)\left(\bar{\underset{\sim}{x}}_{\ell .} - \bar{\underset{\sim}{x}}\right)'$$

...5/-

$$SSP_{\text{faktor 2}} = \sum_{k=1}^{b} gn\left(\bar{\underset{\sim}{x}}_{.k} - \bar{\underset{\sim}{x}}\right)\left(\bar{\underset{\sim}{x}}_{.k} - \bar{\underset{\sim}{x}}\right)'$$

$$SSP_{\substack{\text{tindakan} \\ \text{bersaling}}} = \sum_{\ell=1}^{g} \sum_{k=1}^{b} n \left(\bar{\underset{\sim}{x}}_{\ell k} - \bar{\underset{\sim}{x}}_{\ell .} - \bar{\underset{\sim}{x}}_{.k} + \bar{\underset{\sim}{x}}\right)\left(\bar{\underset{\sim}{x}}_{\ell k} - \bar{\underset{\sim}{x}}_{\ell .} - \bar{\underset{\sim}{x}}_{.k} + \bar{\underset{\sim}{x}}\right)'$$

$$SSP_{\text{residual}} = \sum_{\ell=1}^{g} \sum_{k=1}^{b} \sum_{r=1}^{n} \left(\underset{\sim}{x}_{\ell k r} - \bar{\underset{\sim}{x}}_{\ell k}\right)\left(\underset{\sim}{x}_{\ell k r} - \bar{\underset{\sim}{x}}_{\ell k}\right)'$$

11. <u>Komponen Prinsipal</u>

(a) $\underset{\sim}{Y}_i = \underset{\sim}{e}_i' \underset{\sim}{X}$, $i = 1, 2, \ldots, p.$

$$\rho_{Y_i, X_k} = \frac{e_{ki}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, \qquad i, k = 1, 2, \ldots, p.$$

(b) $\underset{\sim}{Y}_i = \underset{\sim}{e}_i' \underset{\sim}{Z}$

$$\rho_{Y_i, Z_k} = e_{ki}\sqrt{\lambda_i}, \quad i,k = 1, 2, \ldots, p.$$

12. <u>Analisis Faktor</u>

(a) $\underset{\sim}{X} - \underset{\sim}{\mu} = \underset{\sim}{L}\,\underset{\sim}{F} + \underset{\sim}{\epsilon}$

(b) $\text{Kov}(\underset{\sim}{X}) = \underset{\sim}{L}\,\underset{\sim}{L}' + \underset{\sim}{\Psi}$

$\text{Kov}(\underset{\sim}{X}, \underset{\sim}{F}) = \underset{\sim}{L}$

...6/-

(c) $h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \ldots + \ell_{im}^2$, $i = 1, 2, \ldots, p$.

$\sigma_{ii} = h_i^2 + \psi_i$, $i = 1, 2, \ldots, p$.

(d) Kriterium varimax: Pilih transformasi ortogon T yang menjadikan

$$V = \frac{1}{p} \sum_{j=1}^{m} \left[\sum_{i=1}^{p} \tilde{\ell}_{ij}^{*4} - \left(\sum_{i=1}^{p} \tilde{\ell}_{ij}^{*2} \right)^2 \Big/ p \right]$$

sebesar yang mungkin.

13. <u>Analisis Pembezalayan</u>

(a) $Y = \underset{\sim}{\ell}' \underset{\sim}{X} = (\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{X}$

$m = \frac{1}{2}\left(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2\right)' \underset{\sim}{\Sigma}^{-1} \left(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2\right)$

(b) $y = \hat{\underset{\sim}{\ell}}' \underset{\sim}{x} = \left(\bar{\underset{\sim}{x}}_1 - \bar{\underset{\sim}{x}}_2\right)' \underset{\sim}{S}_p^{-1} \underset{\sim}{x}$

$\hat{m} = \frac{1}{2} \left(\bar{\underset{\sim}{x}}_1 - \bar{\underset{\sim}{x}}_2\right)' S_p^{-1} \left(\bar{\underset{\sim}{x}}_1 + \bar{\underset{\sim}{x}}_2\right)$

(c) Petua peruntukan:

$$\text{Untukkan } \underset{\sim}{x}_o \text{ kepada } \begin{cases} \pi_1 \text{ jika } y_o \geq \hat{m} \\ \pi_2 \text{ jika } y_o < \hat{m} \end{cases}$$

...7/-

(d) $\underset{\sim}{B}_o = \sum_{i=1}^{g} \left(\underset{\sim}{\mu}_i - \bar{\underset{\sim}{\mu}}\right)\left(\underset{\sim}{\mu}_i - \bar{\underset{\sim}{\mu}}\right)'$

$\lambda_1, \ldots, \lambda_s$ nilai eigen dan

$\underset{\sim}{e}_1, \ldots, \underset{\sim}{e}_s$ vektor eigen $\underset{\sim}{\Sigma}^{-1}\underset{\sim}{B}_o$.

$\underset{\sim}{\ell}_i\underset{\sim}{X} = \underset{\sim}{e}_i\underset{\sim}{X}$ pembezalayan ke-i , i = 1, 2, ..., s.

(e) $\hat{\underset{\sim}{B}}_o = \sum_{i=1}^{g} \left(\bar{\underset{\sim}{x}}_i - \bar{\underset{\sim}{x}}\right)\left(\bar{\underset{\sim}{x}}_i - \bar{\underset{\sim}{x}}\right)'$

$\underset{\sim}{W} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(\underset{\sim}{x}_{ij} - \bar{\underset{\sim}{x}}_i\right)\left(\underset{\sim}{x}_{ij} - \bar{\underset{\sim}{x}}_i\right)'$

$\hat{\underset{\sim}{\ell}}_i \underset{\sim}{x} = \hat{\underset{\sim}{e}}_i\underset{\sim}{x}$ pembezalayan sampel ke-i, i = 1, ..., s.

(f) Petua peruntukan:

Untukkan $\underset{\sim}{x}$ kepada π_k jika

$$\sum_{j=1}^{r} \left(\hat{y}_j - \bar{y}_{kj}\right)^2 = \sum_{j=1}^{r} \left[\hat{\ell}'_j \left(\underset{\sim}{x} - \bar{x}_k\right)\right]^2$$

$$\leq \sum_{j=1}^{r} \left[\hat{\underset{\sim}{\ell}}'_j \left(\underset{\sim}{x} - \bar{x}_i\right)\right]^2$$

bagi semua i ≠ k, r ≤ s.

- oooOooo -