UNIVERSITI SAINS MALAYSIA

First Semester Examination
Academic Session 2007/2008

October/November 2007

MSG 366 – Multivariate Analysis
[Analisis Multivariat]

Duration : 3 hours
[Masa : 3 jam]

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

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

Instructions: Answer **all four** [4] questions.

*[**Arahan:** Jawab **semua empat** [4] soalan.]*

...2/-

1. (a) What is univariate analysis? Is multivariate analysis an extension of univariate analysis? Discuss.

[10 marks]

(b) How would you examine multivariate data and assess the assumption of normality?

[20 marks]

(c) The following are five measurements on the variables X_1, X_2 and X_3:

X_1	9	2	6	5	8
X_2	12	8	6	4	10
X_3	3	4	0	2	1

Find the arrays $\bar{\mathbf{X}}$, $\mathbf{S}_n$ and $\mathbf{R}$. Explain what the entries in these vector and matrices mean in your own words.

[30 marks]

(d) Derive expressions for the mean and variances of the following linear combinations in terms of the means and covariances of the random variables X_1, X_2, and X_3.

(i) $X_1 - 2X_2$
(ii) $-X_1 + 3X_2$
(iii) $X_1 + X_2 + X_3$
(iv) $X_1 + 2X_2 - X_3$

[40 marks]

1. (a) *Apakah analisis univariat? Adakah analisis multivariat suatu lanjutan analisis univariat? Bincangkan.*

[10 markah]

(b) *Bagaimanakah anda akan memeriksa data multivariat dan menilai andaian kenormalan?*

[20 markah]

(c) *Berikut adalah lima ukuran untuk pembolehubah* X_1, X_2 *and* X_3*:*

X_1	9	2	6	5	8
X_2	12	8	6	4	10
X_3	3	4	0	2	1

Cari tata susunan $\bar{\mathbf{X}}$, $\mathbf{S}_n$ *dan* $\mathbf{R}$. *Jelaskan maksud unsur-unsur di dalam vektor dan matriks tersebut.*

[30 markah]

(d) *Terbitkan min dan varians bagi gabungan-gabungan linear yang berikut dalam sebutan min dan kovarians bagi pembolehubah* X_1, X_2, *dan* X_3.

(i) $X_1 - 2X_2$
(ii) $-X_1 + 3X_2$
(iii) $X_1 + X_2 + X_3$
(iv) $X_1 + 2X_2 - X_3$

[40 markah]

...4/-

2. (a) (i) What is the meaning of inference about a mean vector? How does it play a role in multivariate analysis?

[10 marks]

(ii) Suggest an alternative method for multiple comparisons in the multivariate case. Is there any advantage in using this method?

[10 marks]

(iii) Evaluate T^2, for testing $H_0 : \boldsymbol{\mu}' = [7,11]$, using the data

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

Specify the distribution of T^2 and test H_0 at the $\alpha = 0.05$ level of significance. What is your conclusion?

[30 marks]

(b) (i) When do the comparisons of several multivariate means arise? Use your own made-up example to illustrate the comparisons.

[20 marks]

(ii) Peanuts are an important crop in parts of the southern United States. In an effort to develop improved plants, crop scientists routinely compare varieties with respect to several variables. The data for one two-factor experiment are given in Table 1.

2. (a) (i) *Apakah makna pentaabiran bagi min vektor? Bagaimanakah ia memainkan peranan dalam analisis multivariat?*

[10 markah]

(ii) *Cadangkan kaedah alternatif bagi perbandingan berganda dalam kes multivariat. Apakah kebaikan jika menggunakan kaedah ini?*

[10 markah]

(iii) *Kira* T^2, *bagi menguji* $H_0 : \boldsymbol{\mu}' = [7, 11]$, *dengan menggunakan data*

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

Nyatakan taburan T^2 *dan uji* H_0 *pada paras keertian* $\alpha = 0.05$. *Apakah kesimpulan anda?*

[30 markah]

(b) (i) *Bilakah perbandingan bagi beberapa min multivariat berlaku? Gunakan contoh anda sendiri untuk menunjukkan perbandingan.*

[20 markah]

(ii) *Kacang adalah tanaman yang penting di bahagian selatan Amerika Syarikat. Dalam usaha untuk mendapat tanaman yang lebih baik, saintis membandingkan secara rutin jenis tanaman terhadap beberapa pembolehubah. Data untuk eksperimen dua-faktor diberikan dalam Jadual 1.*

...6/-

Table 1. Peanut Data

Factor 1 Location	Factor 2 Variety	X_1 Yield	X_2 SdMatKer	X_3 SeedSize
1	5	195.3	153.1	51.4
1	5	194.3	167.7	53.7
2	5	189.7	139.5	55.5
2	5	180.4	121.1	44.4
1	6	203.0	156.8	49.8
1	6	195.9	166.0	45.8
2	6	202.7	166.1	60.4
2	6	197.6	161.8	54.1
1	8	193.5	164.5	57.8
1	8	187.0	165.1	58.6
2	8	201.5	166.8	65.0
2	8	200.0	173.8	67.2

Three varieties (5, 6, and 8) were grown at with two geographical locations (1, 2) and, in this case three variables representing yield and the two important grade-grain characteristics were measured. The three variables are

$X_1 =$ Yield (plot weight)

$X_2 =$ Sound mature kernels (weight in grams – maximum of 250 grams)

$X_3 =$ Seed size (weight, in grams, of 100 seeds)

There were two replications of the experiment.

The following SAS output displayed gives us some idea of the sort of analyses which can be performed on the peanut data. State the hypotheses, explain the type of tests carried out and interpret the results obtained. What can we conclude from these results?

[30 marks]

...7/-

Jadual 1. Data Kacang

Faktor 1 Lokasi	*Faktor 2 Jenis*	X_1 *Hasil*	X_2 *SdMatKer*	X_3 *Saiz benih*
1	*5*	*195.3*	*153.1*	*51.4*
1	*5*	*194.3*	*167.7*	*53.7*
2	*5*	*189.7*	*139.5*	*55.5*
2	*5*	*180.4*	*121.1*	*44.4*
1	*6*	*203.0*	*156.8*	*49.8*
1	*6*	*195.9*	*166.0*	*45.8*
2	*6*	*202.7*	*166.1*	*60.4*
2	*6*	*197.6*	*161.8*	*54.1*
1	*8*	*193.5*	*164.5*	*57.8*
1	*8*	*187.0*	*165.1*	*58.6*
2	*8*	*201.5*	*166.8*	*65.0*
2	*8*	*200.0*	*173.8*	*67.2*

Tiga jenis (5, 6, and 8) pokok kacang ditanam di dua lokasi geografik (1, 2) dan, dalam kes ini tiga pembolehubah yang mewakili hasil dan dua ciri biji-gred diukur. Tiga pembolehubah tersebut ialah

$X_1 =$ *Hasil (berat plot)*

$X_2 =$ *'Sound mature kernels' (berat dalam gram – maksimum 250 gram)*

$X_3 =$ *Saiz benih (berat dalam gram bagi 100 benih)*

Terdapat dua replikasi bagi eksperimen tersebut.

Output SAS yang berikut memberi sedikit ide tentang analisis yang dapat dilaksanakan pada data kacang. Nyatakan hipotesis-hipotesis, terangkan jenis ujian yang telah dijalankan dan tafsir keputusan yany diperoleh. Apakah yang kita dapat simpulkan dari keputusan ini?

[30 markah]

...8/-

SAS OUTPUT for Peanut Data

MANOVA
The GLM Procedure

Class Level Information

Class	Levels	Values
factor1	2	1 2
factor2	3	5 6 8

Number of observations 12

Dependent Variable: x1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	401.9175000	80.3835000	4.63	0.0446
Error	6	104.2050000	17.3675000		
Corrected Total	11	506.1225000			

R-Square	Coeff Var	Root MSE	x1 Mean
0.794111	2.136324	4.167433	195.0750

Source	DF	Type III SS	Mean Square	F Value	Pr > F
factor1	1	0.7008333	0.7008333	0.04	0.8474
factor2	2	196.1150000	98.0575000	5.65	0.0418
factor1*factor2	2	205.1016667	102.5508333	5.90	0.0382

...9/-

SAS OUTPUT for Peanut Data

MANOVA
The GLM Procedure

Class Level Information

Class	Levels	Values
factor1	2	1 2
factor2	3	5 6 8

Number of observations 12

Dependent Variable: x1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	401.9175000	80.3835000	4.63	0.0446
Error	6	104.2050000	17.3675000		
Corrected Total	11	506.1225000			

R-Square	Coeff Var	Root MSE	x1 Mean
0.794111	2.136324	4.167433	195.0750

Source	DF	Type III SS	Mean Square	F Value	Pr > F
factor1	1	0.7008333	0.7008333	0.04	0.8474
factor2	2	196.1150000	98.0575000	5.65	0.0418
factor1*factor2	2	205.1016667	102.5508333	5.90	0.0382

...10/-

Dependent Variable: x2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2031.777500	406.355500	6.92	0.0177 Error
6 352.105000	58.684167				
Corrected Total	11	2383.882500			

R-Square	Coeff Var	Root MSE	x2 Mean
0.852298	4.832398	7.660559	158.5250

Source	DF	Type III SS	Mean Square	F Value	Pr > F
factor1	1	162.067500	162.067500	2.76	0.1476
factor2	2	1089.015000	544.507500	9.28	0.0146
factor1*factor2	2	780.695000	390.347500	6.65	0.0300

Dependent Variable: x3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	442.5741667	88.5148333	5.60	0.0292
Error	6	94.8350000	15.8058333		
Corrected Total	11	537.4091667			

R-Square	Coeff Var	Root MSE	x3 Mean
0.823533	7.188166	3.975655	55.30833

Source	DF	Type III SS	Mean Square	F Value	Pr > F
factor1	1	72.5208333	72.5208333	4.59	0.0759
factor2	2	284.1016667	142.0508333	8.99	0.0157
factor1*factor2	2	85.9516667	42.9758333	2.72	0.1443

...11/-

Dependent Variable: x2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2031.777500	406.355500	6.92	0.0177 Error
6 352.105000	58.684167				
Corrected Total	11	2383.882500			

R-Square	Coeff Var	Root MSE	x2 Mean
0.852298	4.832398	7.660559	158.5250

Source	DF	Type III SS	Mean Square	F Value	Pr > F
factor1	1	162.067500	162.067500	2.76	0.1476
factor2	2	1089.015000	544.507500	9.28	0.0146
factor1*factor2	2	780.695000	390.347500	6.65	0.0300

Dependent Variable: x3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	442.5741667	88.5148333	5.60	0.0292
Error	6	94.8350000	15.8058333		
Corrected Total	11	537.4091667			

R-Square	Coeff Var	Root MSE	x3 Mean
0.823533	7.188166	3.975655	55.30833

Source	DF	Type III SS	Mean Square	F Value	Pr > F
factor1	1	72.5208333	72.5208333	4.59	0.0759
factor2	2	284.1016667	142.0508333	8.99	0.0157
factor1*factor2	2	85.9516667	42.9758333	2.72	0.1443

...12/-

The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	x1	x2	x3
x1	104.205	49.365	76.48
x2	49.365	352.105	121.995
x3	76.48	121.995	94.835

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 6	x1	x2	x3
x1	1.000000	0.257714 0.5769	0.769342 0.0432
x2	0.257714 0.5769	1.000000	0.667608 0.1013
x3	0.769342 0.0432	0.667608 0.1013	1.000000

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for factor1
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 x1	x2	x3
8.38824348	100.00	-0.13688388	-0.07628041	0.23952166
0.00000000	0.00	0.10187838	0.00216080	-0.00678495
0.00000000	0.00	-0.06307410	0.03725453	0.06189287

...13/-

The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	x1	x2	x3
x1	104.205	49.365	76.48
x2	49.365	352.105	121.995
x3	76.48	121.995	94.835

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 6	x1	x2	x3
x1	1.000000	0.257714 0.5769	0.769342 0.0432
x2	0.257714 0.5769	1.000000	0.667608 0.1013
x3	0.769342 0.0432	0.667608 0.1013	1.000000

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for factor1
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 x1	x2	x3
8.38824348	100.00	-0.13688388	-0.07628041	0.23952166
0.00000000	0.00	0.10187838	0.00216080	-0.00678495
0.00000000	0.00	-0.06307410	0.03725453	0.06189287

...14/-

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall factor1 Effect
H = Type III SSCP Matrix for factor1
E = Error SSCP Matrix

S=1 M=0.5 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.10651620	11.18	3	4	0.0205
Pillai's Trace	0.89348380	11.18	3	4	0.0205
Hotelling-Lawley Trace	8.38824348	11.18	3	4	0.0205
Roy's Greatest Root	8.38824348	11.18	3	4	0.0205

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for factor2
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 x1	x2	x3
18.1876113	85.09	-0.16986539	-0.06425268	0.23943636
3.1880638	14.91	0.00137509	0.03769309	0.03800092
0.0000000	0.00	0.06510456	-0.04076880	0.04973481

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall factor2 Effect
H = Type III SSCP Matrix for factor2
E = Error SSCP Matrix

S=2 M=0 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.01244417	10.62	6	8	0.0019
Pillai's Trace	1.70910921	9.79	6	10	0.0011
Hotelling-Lawley Trace	21.37567504	10.69	6	6	0.0055
Roy's Greatest Root	18.18761127	30.31	3	5	0.0012

...15/-

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall factor1 Effect
H = Type III SSCP Matrix for factor1
E = Error SSCP Matrix

S=1 M=0.5 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.10651620	11.18	3	4	0.0205
Pillai's Trace	0.89348380	11.18	3	4	0.0205
Hotelling-Lawley Trace	8.38824348	11.18	3	4	0.0205
Roy's Greatest Root	8.38824348	11.18	3	4	0.0205

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for factor2
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 x1	x2	x3
18.1876113	85.09	-0.16986539	-0.06425268	0.23943636
3.1880638	14.91	0.00137509	0.03769309	0.03800092
0.0000000	0.00	0.06510456	-0.04076880	0.04973481

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall factor2 Effect
H = Type III SSCP Matrix for factor2
E = Error SSCP Matrix

S=2 M=0 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.01244417	10.62	6	8	0.0019
Pillai's Trace	1.70910921	9.79	6	10	0.0011
Hotelling-Lawley Trace	21.37567504	10.69	6	6	0.0055
Roy's Greatest Root	18.18761127	30.31	3	5	0.0012

...16/-

The GLM Procedure
Multivariate Analysis of Variance

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

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for factor1*factor2
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 x1	x2	x3
6.82409388	90.45	0.15723347	0.06948572	-0.18762316
0.72019649	9.55	-0.08644203	0.01400396	0.12612011
0.00000000	0.00	0.03000259	-0.04676424	0.10069089

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall factor1*factor2 Effect
H = Type III SSCP Matrix for factor1*factor2
E = Error SSCP Matrix

S=2 M=0 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.07429984	3.56	6	8	0.0508
Pillai's Trace	1.29086073	3.03	6	10	0.0587
Hotelling-Lawley Trace	7.54429038	3.77	6	6	0.0655
Roy's Greatest Root	6.82409388	11.37	3	5	0.0113

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

...17/-

The GLM Procedure
Multivariate Analysis of Variance

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

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for factor1*factor2
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 x1	x2	x3
6.82409388	90.45	0.15723347	0.06948572	-0.18762316
0.72019649	9.55	-0.08644203	0.01400396	0.12612011
0.00000000	0.00	0.03000259	-0.04676424	0.10069089

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall factor1*factor2 Effect
H = Type III SSCP Matrix for factor1*factor2
E = Error SSCP Matrix

S=2 M=0 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.07429984	3.56	6	8	0.0508
Pillai's Trace	1.29086073	3.03	6	10	0.0587
Hotelling-Lawley Trace	7.54429038	3.77	6	6	0.0655
Roy's Greatest Root	6.82409388	11.37	3	5	0.0113

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

...18/-

3. (i) What are the objectives of analysis of principal components (PCA) and factor analysis (FA)?

[10 marks]

(ii) Describe the salient features of both multivariate techniques, PCA and FA.

[20 marks]

(iii) Is there any difference between PCA and FA? Elaborate.

[20 marks]

(iv) A naturalist for the Alaska Fish and Game Department studies grizzly bears with the goal of maintaining a healthy population. Measurements on $n = 61$ bears provided the following summary statistics:

Variable	Weight (kg)	Body Length (cm)	Neck (cm)	Girth (cm)	Head Length (cm)	Head Width (cm)
Sample mean	95.52	164.39	55.69	93.39	17.98	31.13

Covariance matrix

$$\mathbf{S} = \begin{bmatrix} 3266.46 & 1343.97 & 731.54 & 1175.50 & 162.68 & 238.37 \\ 1343.97 & 721.91 & 324.25 & 537.35 & 80.17 & 117.73 \\ 731.54 & 324.25 & 179.28 & 281.17 & 39.15 & 56.80 \\ 1175.50 & 537.35 & 281.17 & 474.98 & 63.73 & 94.85 \\ 162.68 & 80.17 & 39.15 & 63.73 & 9.95 & 13.88 \\ 238.37 & 117.73 & 56.80 & 94.85 & 13.88 & 21.26 \end{bmatrix}$$

Can the data be effectively summarized in fewer than six dimensions? Discuss. What is your conclusion based on the following SAS output?

[50 marks]

...19/-

3. (i) *Apakah objektif-objektif analisis komponen prinsipal (PCA) dan analisis faktor (FA)?*

[10 markah]

(ii) *Huraikan ciri-ciri penting bagi kedua-dua teknik multivariat, PCA dan FA.*

[20 markah]

(iii) *Adakah terdapat perbezaan di antara PCA dan FA? Jelaskan.*

[20 markah]

(iv) *Seorang pengkaji haiwan bagi Jabatan Haiwan negeri Alaska sedang mengkaji beruang yang berbulu kelabu dengan tujuan untuk menjamin populasi yang sihat. Ukuran pada* $n = 61$ *beruang berbulu kelabu memberi statistik ringkasan yang berikut:*

Pemboleh ubah	*Berat (kg)*	*Panjang badan (sm)*	*Leher (sm)*	*Lilitan (sm)*	*Panjang kepala (sm)*	*Lebar kepala (sm)*
Min Sampel	*95.52*	*164.39*	*55.69*	*93.39*	*17.98*	*31.13*

Matriks kovarians

$$\mathbf{S} = \begin{bmatrix} 3266.46 & 1343.97 & 731.54 & 1175.50 & 162.68 & 238.37 \\ 1343.97 & 721.91 & 324.25 & 537.35 & 80.17 & 117.73 \\ 731.54 & 324.25 & 179.28 & 281.17 & 39.15 & 56.80 \\ 1175.50 & 537.35 & 281.17 & 474.98 & 63.73 & 94.85 \\ 162.68 & 80.17 & 39.15 & 63.73 & 9.95 & 13.88 \\ 238.37 & 117.73 & 56.80 & 94.85 & 13.88 & 21.26 \end{bmatrix}$$

Bolehkah data diringkaskan secara berkesan dalam dimensi yang kurang daripada enam? Bincangkan. Apakah kesimpulan anda berdasarkan output SAS yang berikut?

[50 markah]

...20/-

SAS OUTPUT for the Bear Data

Factor Analysis of Bear Data
The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 6 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	5.64457153	5.46879703	0.9408	0.9408
2	0.17577450	0.11925590	0.0293	0.9701
3	0.05651860	0.00727782	0.0094	0.9795
4	0.04924078	0.00190153	0.0082	0.9877
5	0.04733925	0.02078393	0.0079	0.9956
6	0.02655533		0.0044	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern

	Factor1	Factor2
weight	0.95908	-0.23406
bodylgth	0.96057	0.22316
neck	0.97398	-0.16320
girth	0.97885	-0.09347
headlgth	0.97204	0.13375
headwidt	0.97490	0.13391

Variance Explained by Each Factor

Factor1	Factor2
5.6445715	0.1757745

Final Communality Estimates: Total = 5.820346

weight	bodylgth	neck	girth	headlgth	headwidt
0.97461318	0.97248445	0.97526168	0.96687840	0.96274190	0.96836643

...21/-

SAS OUTPUT for the Bear Data

Factor Analysis of Bear Data
The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 6 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	5.64457153	5.46879703	0.9408	0.9408
2	0.17577450	0.11925590	0.0293	0.9701
3	0.05651860	0.00727782	0.0094	0.9795
4	0.04924078	0.00190153	0.0082	0.9877
5	0.04733925	0.02078393	0.0079	0.9956
6	0.02655533		0.0044	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern

	Factor1	Factor2
weight	0.95908	-0.23406
bodylgth	0.96057	0.22316
neck	0.97398	-0.16320
girth	0.97885	-0.09347
headlgth	0.97204	0.13375
headwidt	0.97490	0.13391

Variance Explained by Each Factor

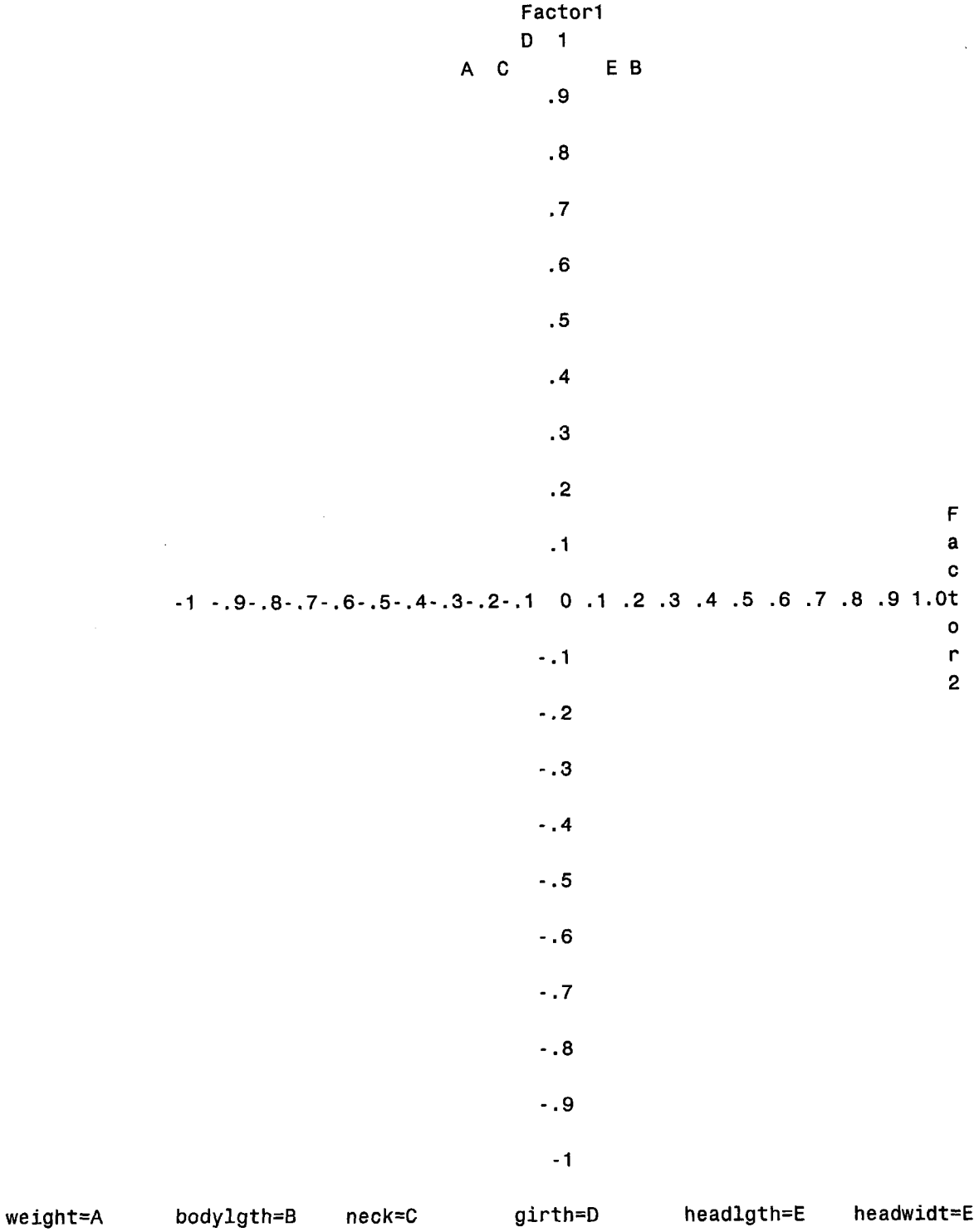
Factor1	Factor2
5.6445715	0.1757745

Final Communality Estimates: Total = 5.820346

weight	bodylgth	neck	girth	headlgth	headwidt
0.97461318	0.97248445	0.97526168	0.96687840	0.96274190	0.96836643

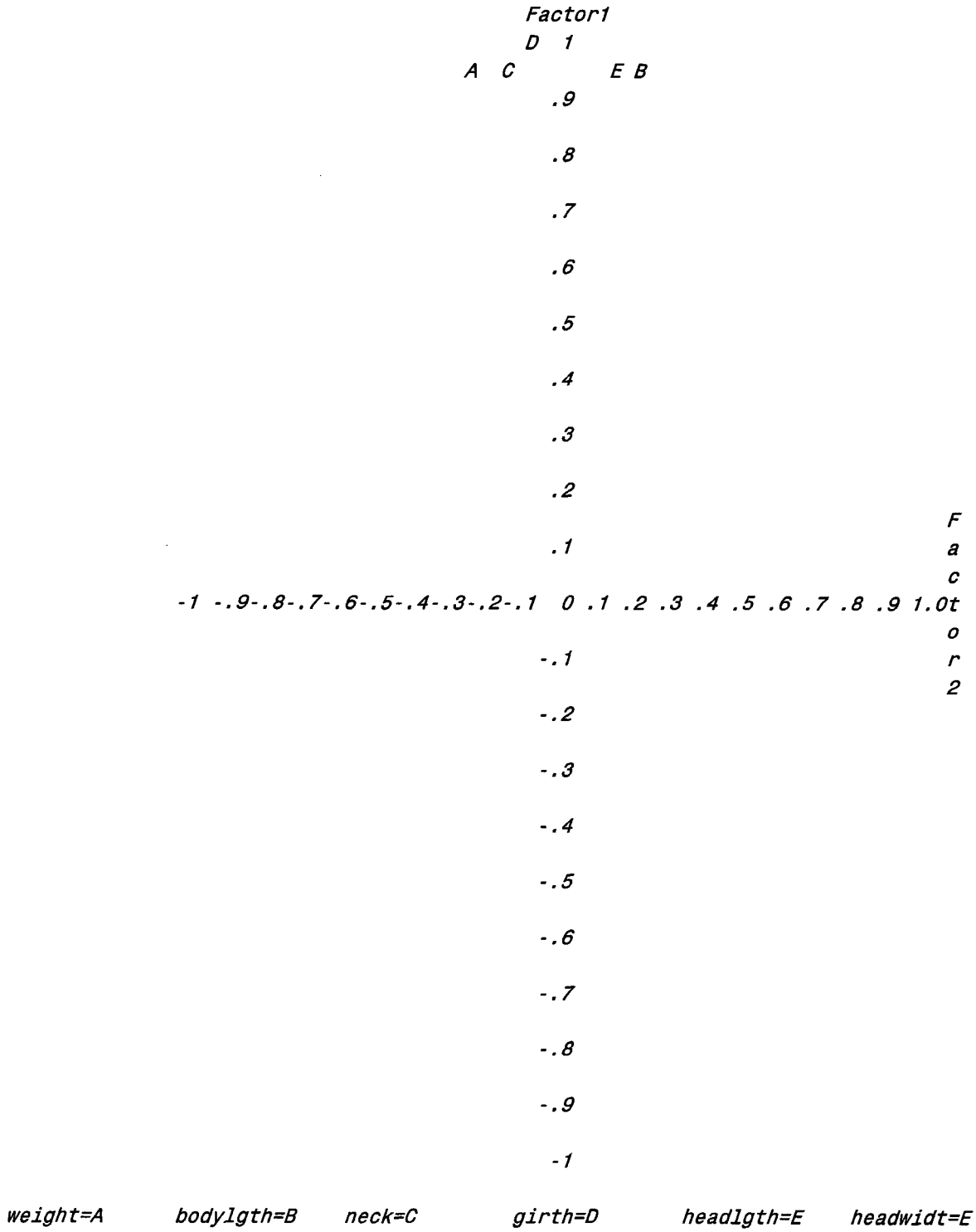
...22/-

Plot of Factor Pattern for Factor1 and Factor2



...23/-

Plot of Factor Pattern for Factor1 and Factor2



...24/-

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2
1	0.71127	0.70292
2	0.70292	-0.71127

Rotated Factor Pattern

	Factor1	Factor2
weight	0.51764	0.84063
bodylgth	0.84009	0.51647
neck	0.57805	0.80070
girth	0.63053	0.75453
headlgth	0.78540	0.58812
headwidt	0.78755	0.59002

Variance Explained by Each Factor

Factor1	Factor2
2.9424953	2.8778507

Final Communality Estimates: Total = 5.820346

weight	bodylgth	neck	girth	headlgth	headwidt
0.97461318	0.97248445	0.97526168	0.96687840	0.96274190	0.96836643

...25/-

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2
1	0.71127	0.70292
2	0.70292	-0.71127

Rotated Factor Pattern

	Factor1	Factor2
weight	0.51764	0.84063
bodylgth	0.84009	0.51647
neck	0.57805	0.80070
girth	0.63053	0.75453
headlgth	0.78540	0.58812
headwidt	0.78755	0.59002

Variance Explained by Each Factor

Factor1	Factor2
2.9424953	2.8778507

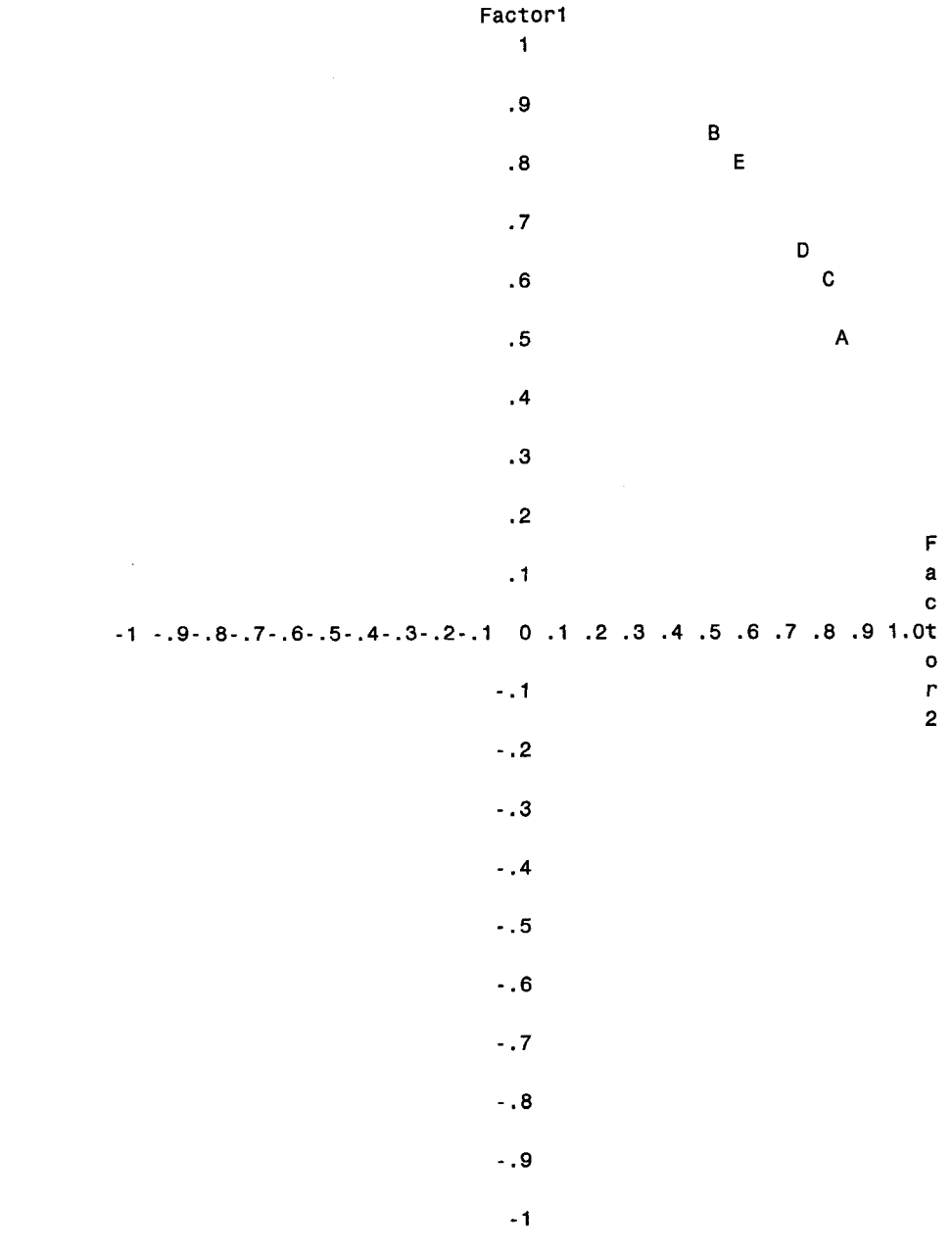
Final Communality Estimates: Total = 5.820346

weight	bodylgth	neck	girth	headlgth	headwidt
0.97461318	0.97248445	0.97526168	0.96687840	0.96274190	0.96836643

...26/-

The FACTOR Procedure
Rotation Method: Varimax

Plot of Factor Pattern for Factor1 and Factor2



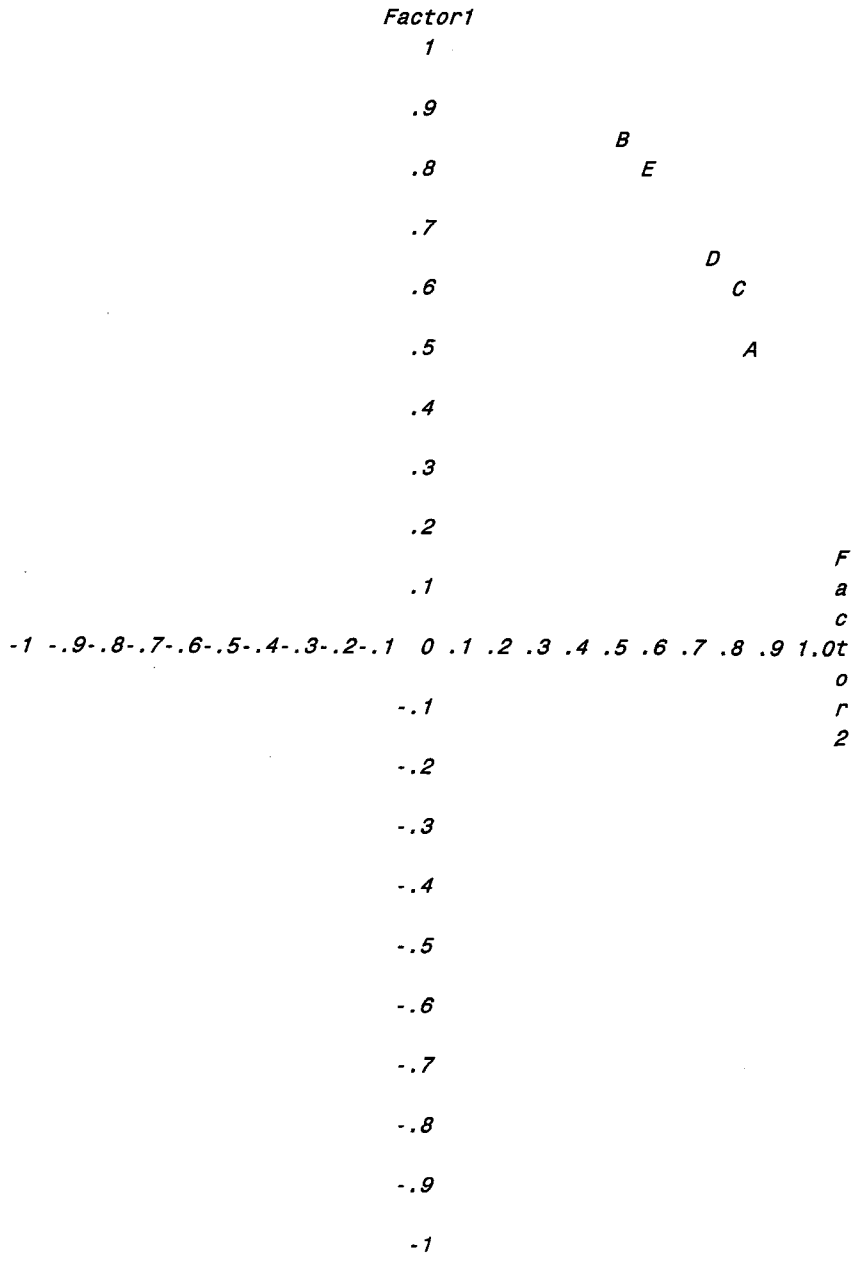
weight=A bodylgth=B neck=C girth=D headlgth=E headwidt=E

...27/-

The FACTOR Procedure

Rotation Method: Varimax

Plot of Factor Pattern for Factor1 and Factor2



...28/-

4. (i) What is an appropriate technique for analyzing multivariate data in categorical form? Explain the ideas behind this technique. You can illustrate the ideas with an example.

[25 marks]

(ii) Consider the two data sets

$$\mathbf{X}_1 = \begin{bmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2 = \begin{bmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{bmatrix}$$

for which

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

and

$$\mathbf{S}_{pooled} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

(a) Calculate the linear discriminant function given by

$$\hat{y} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} \mathbf{x}$$

(b) Classify the observation $\mathbf{x}_0' = [2 \quad 7]$ as population π_1 or population π_2, using the classification statistic

$$\hat{w} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} \mathbf{x}_0 - \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)$$

with equal priors and equal costs.

[25 marks]

...29/-

4. (i) *Apakah teknik yang sesuai untuk menganalisis data multivariat dalam bentuk berkategori? Terangkan ide-ide dalam teknik ini. Anda boleh mengilustrasi ide-ide dengan menggunakan satu contoh.*

[25 markah]

(ii) *Pertimbangkan dua set data*

$$\mathbf{X}_1 = \begin{bmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{bmatrix} \quad \text{dan} \quad \mathbf{X}_2 = \begin{bmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{bmatrix}$$

di mana

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

dan

$$\mathbf{S}_{pooled} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

(a) *Kira fungsi pembezalayan linear yang diberi oleh*

$$\hat{y} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} \mathbf{x}$$

(b) *Kelaskan cerapan* $\mathbf{x}_0' = [2 \quad 7]$ *sebagai populasi* π_1 *atau populasi* π_2, *dengan menggunakan statistik pengkelasan*

$$\hat{w} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} \mathbf{x}_0 - \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)$$

dengan prior sama dan kos sama.

[25 markah]

(iii) Data were collected by Bryce and Barker (2006) as part of a preliminary study of a possible link between football helmet design and neck injuries. Six head measurements were made on each subject. There were 30 subjects in each of three groups: high school football players (group 1), college football players (group 2), and non-football players (group 3). The six variables are

WDIM = head width at widest dimension,
CIRCUM = head circumference,
FBEYE = front-to-back measurement at eye level,
EYEHD = eye-to-top-of-head measurement,
EARHD = ear-to-top-of-head measurement,
JAW = jaw width.

The results of a multivariate data analysis using SPSS are displayed as follows:

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.921[a]	94.3	94.3	.811
2	.116[a]	5.7	100.0	.322

a. First 2 canonical discriminant functions were used in the analysis.

Classification Function Coefficients

	Group		
	1	2	3
Head Width	7.655	10.140	10.895
Head Circumference	13.278	13.256	13.244
Front-to-back	3.935	3.938	3.935
Eye-to-top-of-head	-1.244	-3.403	-2.688
Ear-to-top-of-head	14.752	13.325	13.224
Jaw width	8.250	6.190	5.297
(Constant)	-641.325	-608.083	-614.614

Fisher's linear discriminant functions

...31/-

(iii) *Data telah dikutip oleh Bryce dan Barker (2006) sebagai sebahagian kajian awal bagi kemungkinan terdapat kaitan di antara rekabentuk topi bola sepak dan kecederaan leher. Enam ukuran kepala dibuat pada setiap pemain. Terdapat 30 pemain dalam setiap tiga kumpulan: pemain bola sepak sekolah tinggi (kumpulan 1), pemain bola sepak kolej (kumpulan 2), dan bukan pemain bola sepak (kumpulan 3). Enam pembolehubah ialah*

WDIM = lebar kepala pada dimensi yang terlebar,
CIRCUM = lilitan kepala,
FBEYE = ukuran depan-ke-belakang di paras mata,
EYEHD = ukuran mata-ke-atas-kepala,
EARHD = ukuran telinga-ke-atas-kepala,
JAW = lebar rahang.

Keputusan analisis data multivariat dengan menggunakan SPSS dipamerkan seperti berikut:

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.921[a]	94.3	94.3	.811
2	.116[a]	5.7	100.0	.322

a. First 2 canonical discriminant functions were used in the analysis.

Classification Function Coefficients

	Group		
	1	2	3
Head Width	7.655	10.140	10.895
Head Circumference	13.278	13.256	13.244
Front-to-back	3.935	3.938	3.935
Eye-to-top-of-head	-1.244	-3.403	-2.688
Ear-to-top-of-head	14.752	13.325	13.224
Jaw width	8.250	6.190	5.297
(Constant)	-641.325	-608.083	-614.614

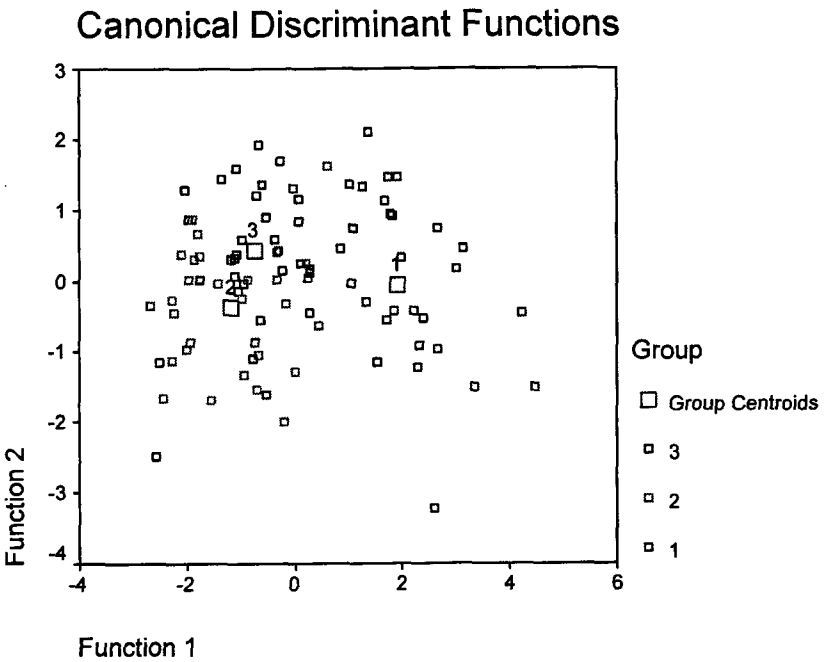
Fisher's linear discriminant functions

...32/-

Classification Results

		Predicted Group Membership			
	Group	1	2	3	Total
Original Count	1	26	1	3	30
	2	1	20	9	30
	3	2	8	20	30
%	1	86.7	3.3	10.0	100.0
	2	3.3	66.7	30.0	100.0
	3	6.7	26.7	66.7	100.0

a.73.3% of original grouped cases correctly classified.



Explain step-by-step what the displayed results mean. Then give your conclusion for this data analysis.

[30 marks]

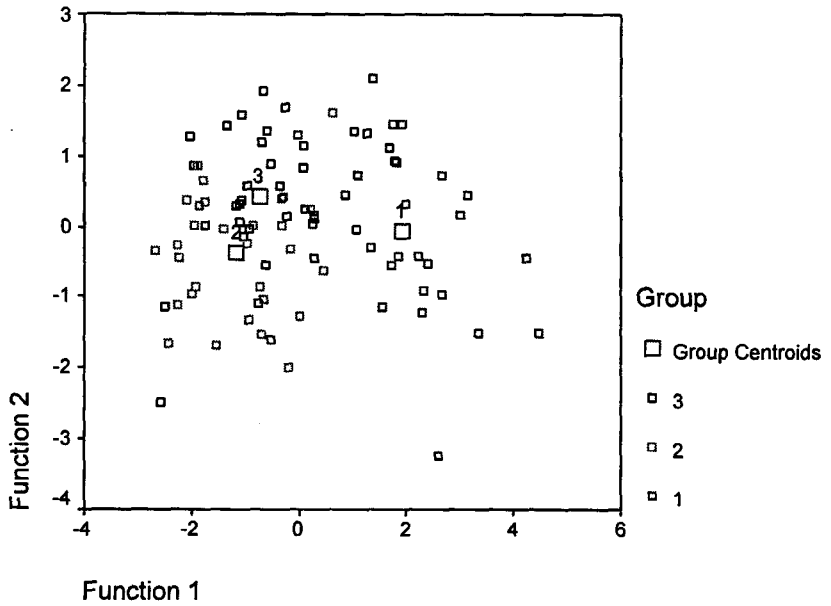
...33/-

Classification Results

		Predicted Group Membership			
	Group	1	2	3	Total
Original Count	1	26	1	3	30
	2	1	20	9	30
	3	2	8	20	30
%	1	86.7	3.3	10.0	100.0
	2	3.3	66.7	30.0	100.0
	3	6.7	26.7	66.7	100.0

a.73.3% of original grouped cases correctly classified.

Canonical Discriminant Functions



Jelaskan langkah-demi-langkah apa maksudnya keputusan yang dipamerkan. Kemudian beri kesimpulan anda untuk analisis data ini.

[30 markah]

...34/-

(iv) Elston and Grizzle (1962) collected data which consist of measurements y_1, y_2, y_3 and y_4 of the ramus bone at four different ages (8 yr, 8 yr 6 mths, 9 yr, 9 yr 6 mths) on each of 20 boys (see Table 2). Hierarchical Clustering was carried out using SPSS and the results are obtained for two methods, single linkage and average linkage. Compare the two methods. Which do you think is better? Explain.

[20 marks]

Table 2. Ramus Bone Length at Four Ages for 20 boys

Boy	8 yr (y_1)	8 yr 6 mths (y_2)	9 yr (y_3)	9 yr 6 mths (y_4)
1	47.8	48.8	49.0	49.7
2	46.4	47.3	47.7	48.4
3	46.3	46.8	47.8	48.5
4	45.1	45.3	46.1	47.2
5	47.6	48.5	48.9	49.3
6	52.5	53.2	53.3	53.7
7	51.2	53.0	54.3	54.5
8	49.8	50.0	50.3	52.7
9	48.1	50.8	52.3	54.4
10	45.0	47.0	47.3	48.3
11	51.2	51.4	51.6	51.9
12	48.5	49.2	53.0	55.5
13	52.1	52.8	53.7	55.0
14	48.2	48.9	49.3	49.8
15	49.6	50.4	51.2	51.8
16	50.7	51.7	52.7	53.3
17	47.2	47.7	48.4	49.5
18	53.3	54.6	55.1	55.3
19	46.2	47.5	48.1	48.4
20	46.3	47.6	51.3	51.8

...35/-

(iv) *Elston dan Grizzle (1962) mengutip data yang terdiri daripada ukuran-ukuran y_1, y_2, y_3 and y_4 bagi tulang ramus 20 budak lelaki yang berumur 8 tahun, 8 tahun 6 bulan, 9 tahun, dan 9 tahun 6 bulan masing-masing (lihat* ***Jadual 2****). Kelompok hirarkikal dijalankan dengan menggunakan SPSS dan keputusan diperoleh bagi dua kaedah, pautan tunggal dan pautan purata. Bandingkan dua kaedah tersebut. Yang manakah lebih baik? Jelaskan.*

[20 markah]

Jadual 2. Panjang Tulang Ramus pada Empat Umur bagi 20 budak lelaki

Budak lelaki	***8 thn*** (y_1)	***8 thn 6 bulan*** (y_2)	***9 thn*** (y_3)	***9 thn 6 bulan*** (y_4)
1	47.8	48.8	49.0	49.7
2	46.4	47.3	47.7	48.4
3	46.3	46.8	47.8	48.5
4	45.1	45.3	46.1	47.2
5	47.6	48.5	48.9	49.3
6	52.5	53.2	53.3	53.7
7	51.2	53.0	54.3	54.5
8	49.8	50.0	50.3	52.7
9	48.1	50.8	52.3	54.4
10	45.0	47.0	47.3	48.3
11	51.2	51.4	51.6	51.9
12	48.5	49.2	53.0	55.5
13	52.1	52.8	53.7	55.0
14	48.2	48.9	49.3	49.8
15	49.6	50.4	51.2	51.8
16	50.7	51.7	52.7	53.3
17	47.2	47.7	48.4	49.5
18	53.3	54.6	55.1	55.3
19	46.2	47.5	48.1	48.4
20	46.3	47.6	51.3	51.8

...36/-

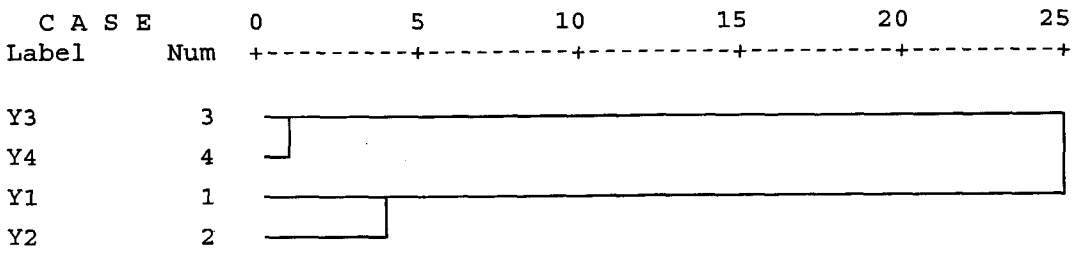
Single Linkage

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	4	24.800	0	0	3
2	1	2	26.440	0	0	3
3	1	3	38.010	2	1	0

Dendrogram using Single Linkage

Rescaled Distance Cluster Combine



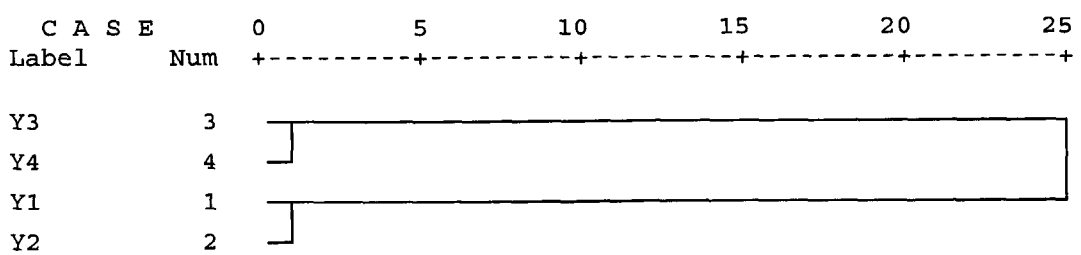
Average Linkage (Within Groups)

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	4	24.800	0	0	3
2	1	2	26.440	0	0	3
3	1	3	84.697	2	1	0

Dendrogram using Average Linkage (Within Group)

Rescaled Distance Cluster Combine



...37/-

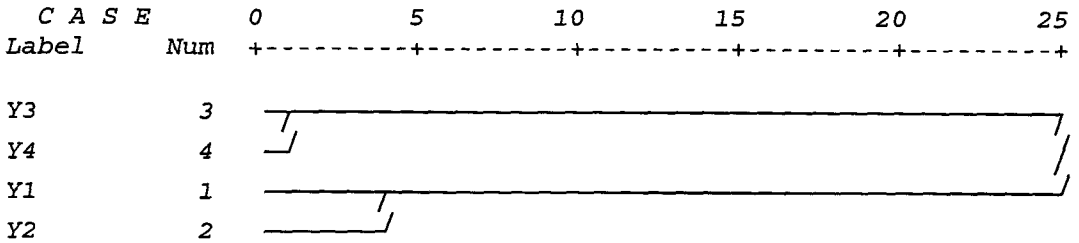
Single Linkage

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	4	24.800	0	0	3
2	1	2	26.440	0	0	3
3	1	3	38.010	2	1	0

Dendrogram using Single Linkage

Rescaled Distance Cluster Combine



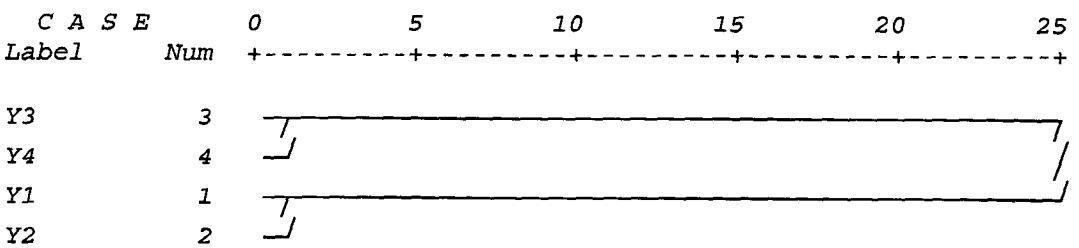
Average Linkage (Within Groups)

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	4	24.800	0	0	3
2	1	2	26.440	0	0	3
3	1	3	84.697	2	1	0

Dendrogram using Average Linkage (Within Group)

Rescaled Distance Cluster Combine



...38/-

APPENDIX

The notations are as given in the lectures.

1. Spectral decomposition for a $k \times k$ symmetric matrix, A, is given by

$$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \ldots + \lambda_k e_k e_k'$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are eigenvalues of A and $e_1, e_2, \ldots, e_k$ are the corresponding standardized eigenvectors.

2. Suppose X has $E(X) = \mu$ and $\text{Cov}(\mathbf{X}) = \Sigma$. Thus $c'X$ has mean, $c'\mu$, and variance, $c'\Sigma c$.

3. Bivariate normal pdf:

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

$$\left.\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right) - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

4. Multivariate normal pdf:

$$f(x_1, x_2) = \frac{1}{(2\pi)^{p/2}\left|\sum\right|^{1/2}} \; e^{-(1/2)(x-\mu)'\sum^{-1}(x-\mu)}$$

5. If $X \sim N_p(\mu, \Sigma)$, then $AX \sim N_q(A\mu, A\Sigma A')$.

6. One-sampel :

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

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

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

(b) Wilks' Lambda $\Lambda^{\frac{2}{n}} = \dfrac{\left|\hat{\Sigma}\right|}{\left|\hat{\Sigma}_0\right|} = \left(1 + \dfrac{T^2}{n-1}\right)^{-1}$

...39/-

LAMPIRAN

Tatatanda adalah seperti di dalam kuliah.

1. *Penguraian spektrum bagi suatu matriks simetrik* $k \times k$, A *diberikan oleh*

$$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \ldots + \lambda_k e_k e_k'$$

di mana $\lambda_1, \lambda_2, \ldots, \lambda_k$ *adalah nilai-nilai eigen* A *dan* $e_1, e_2, \ldots, e_k$ *adalah vektor-vektor eigen terpiawai yang berkaitan.*

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

3. *f.k.k. normal bivariat:*

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

$$\left.\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right) - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

4. *f.k.k. normal multivariat:*

$$f(x_1, x_2) = \frac{1}{(2\pi)^{p/2}\left|\sum\right|^{1/2}} \; e^{-(1/2)(x-\mu)'\sum^{-1}(x-\mu)}$$

5. *Jika* $X \sim N_p(\mu, \Sigma)$, *maka* $AX \sim N_q(A\mu, A\Sigma A')$.

6. *Satu sampel :*

(a)

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

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

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

(b) *Lambda Wilks* $A^{\frac{2}{n}} = \dfrac{\left|\hat{\Sigma}\right|}{\left|\hat{\Sigma}_0\right|} = \left(1+\dfrac{T^2}{n-1}\right)^{-1}$

...40/-

(c) $100(1-\alpha)\%$ simultaneous confidence intervals for $l'\mu$:

$$l'\bar{X} \pm \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)\ l'Sl}$$

(d) $100\ (1-\alpha)\%$ Bonferroni confidence interval for $\mu_i,\ i=1,2,...,p$:

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

7. Two independent samples:

(a) $$T^2 = \left[\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)\right]' \left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_p\right]^{-1}$$
$$\left[\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)\right]$$
$$T^2 \sim \frac{(n_1+n_2-2)p}{(n_1+n_2-p-1)}F_{p,n_1+n_2-p-1}$$

(b) $100\ (1-\alpha)\%$ simultaneous confidence intervals for $l'(\mu_1-\mu_2)$:

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

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

8. One-way MANOVA:

(a) $$B = \sum_{\ell=1}^{g} n_\ell (\bar{x}_\ell - \bar{x})(\bar{x}_\ell - \bar{x})'$$
$$W = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)(x_{\ell j} - \bar{x}_\ell)'$$
$$\Lambda^* = \frac{|W|}{|B+W|}$$

...41/-

(c) *Selang keyakinan serentak* $100(1-\alpha)\%$ *bagi* $l'\mu$ *:*

$$l'\bar{X} \pm \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)\ l'Sl}$$

(d) *Selang keyakinan serentak Bonferroni* $100\ (1-\alpha)\%$ *bagi*

$\mu_i,\ i=1,2,...,p$:

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

7. *Dua sampel tak bersandar:*

(a) $$T^2 = \left[\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)\right]' \left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_p\right]^{-1}$$
$$\left[\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)\right]$$
$$T^2 \sim \frac{(n_1+n_2-2)P}{(n_1+n_2-p-1)}F_{p,n_1+n_2-p-1}$$

(b) *Selang keyakinan serentak* $100\ (1-\alpha)\%$ *bagi* $l'(\mu_1-\mu_2)$:

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

di mana $c^2 = \dfrac{(n_1+n_2-2)p}{n_1+n_2-p-1}\ F_{p,n_1+n_2-p-1}$

8. *MANOVA satu-hala:*

(a) $$B = \sum_{\ell=1}^{g} n_\ell (\bar{x}_\ell - \bar{x})(\bar{x}_\ell - \bar{x})'$$
$$W = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)(x_{\ell j} - \bar{x}_\ell)'$$
$$\Lambda^* = \frac{|W|}{|B+W|}$$

(b) $100\,(1-\alpha)\%$ simultaneous confidence intervals for $\tau_{ki}-\tau_{\ell i}$:

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

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

9. Assume that $\boldsymbol{E}$ has m_E d.f. and $\boldsymbol{H}$ has m_H d.f.

Suppose $\wedge=\dfrac{|\boldsymbol{E}|}{|\boldsymbol{E}+\boldsymbol{H}|}$

Hence (1) For p=1,

$$\left(\frac{1-\wedge}{\wedge}\right)\frac{m_E}{m_H}\ \sim\ F_{m_H,\,m_E}\ \text{for any } m_H$$

(2) For $m_H=1$,

$$\left(\frac{1-\wedge}{\wedge}\right)\frac{m_E+1-p}{p}\ \sim\ F_{p,m_E+1-p}\ \text{for any } p$$

(3) For $p=2$,

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

(4) For $m_H=2$,

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

for $p\geq 2$.

Bartlett's correction: Say $n_0=m_E+m_H$.

For large m_E,

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

where $f=m_E-\dfrac{1}{2}(p-m_H+1)$

$=n_0-\dfrac{1}{2}(p-m_H+1)$

...43/-

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

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

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

9. *Andaikan* $\boldsymbol{E}$ *mempunyai d.k.* m_E *dan* $\boldsymbol{H}$ *mempunyai d.k.* m_H.

Katakan $\wedge=\dfrac{|\boldsymbol{E}|}{|\boldsymbol{E}+\boldsymbol{H}|}$

Maka (1) *Untuk* $p=1$,

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

(2) *Untuk* $m_H=1$,

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

(3) *Untuk* $p=2$,

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

(4) *Untuk* $m_H=2$,

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

untuk $p\geq 2$.

Pembetulan Bartlett: Katakan $n_0=m_E+m_H$.

Bagi m_E *besar,*

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

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

$=n_0-\dfrac{1}{2}(p-m_H+1)$

...44/-

10. Two-way MANOVA:

$$SSP_{\text{factor1}} = \sum_{\ell=1}^{g} bn(\bar{x}_{\ell.} - \bar{x})(\bar{x}_{\ell.} - \bar{x})'$$

$$SSP_{\text{factor 2}} = \sum_{k=1}^{b} gn(\bar{x}_{.k} - \bar{x})(\bar{x}_{.k} - \bar{x})'$$

$$SSP_{\text{interaction}} = \sum_{\ell=1}^{g} \sum_{k=1}^{b} n(\bar{x}_{\ell k} - \bar{x}_{\ell.} - \bar{x}_{.k} + \bar{x})$$
$$(\bar{x}_{\ell k} - \bar{x}_{\ell.} - \bar{x}_{.k} + \bar{x})'$$

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

11. Principal Components:

(a) $Y_i = e_i' X, \quad i = 1, 2, \ldots, p,$

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

(b) $Y_i = e_i' Z$

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

12. Factor Analysis:

(a) $X - \mu = L F + \epsilon$

(b) $\text{Cov}(X) = L L' + \psi$

$\text{Cov}(X, F) = L$

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

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

(d) Varimax criterion: Choose an orthogonal transformation such that

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

is as large as possible.

...45/-

10. *MANOVA dua –hala:*

$$SSP_{\text{faktor1}} = \sum_{\ell=1}^{g} bn(\bar{x}_{\ell.} - \bar{x})(\bar{x}_{\ell.} - \bar{x})'$$

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

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

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

11. *Komponen Prinsipal:*

(a) $Y_i = e_i' X, \quad i = 1, 2, \ldots, p,$

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

(b) $Y_i = e_i' Z$

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

12. *Analisis Faktor:*

(a) $X - \mu = L F + \epsilon$

(b) $\text{Kov}(X) = L L' + \psi$

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

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

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

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

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

sebesar yang mungkin.

...46/-

13. Discriminant Analysis:

(a) $Y = l'X = (\mu_1 - \mu_2)' \sum^{-1} X$

$$m = \frac{1}{2}(\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2)$$

(b) $y = \hat{l}'X = (\bar{x}_1 - \bar{x}_2)' S_p^{-1} x$

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

(c) Allocation rule:

$$\text{Allocate } x_0 \text{ to } \begin{cases} \pi_1 \text{ if } y_0 \geq \hat{m} \\ \pi_2 \text{ if } y_0 < \hat{m} \end{cases}$$

(d) $B_0 = \sum_{i=1}^{g} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})$

Eigenvalues $\lambda_1, \ \ldots, \ \lambda_s$ and

Eigenvectors $e_1, \ \ldots, \ e_s$ of $\sum^{-1} B_0$.

i th discriminant $\ell_i X = e_i X$, $i = 1, 2, \ldots, s$

(e) $B_0 = \sum_{i=1}^{g} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$

$$W = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)'$$

i th discriminant $\hat{\ell}_i x = \hat{e}_i x$, $i = 1, 2, \ldots, s$.

(f) Allocation rule:

Allocate x to π_k if

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

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

for all $i \neq k$, $r \leq s$.

...47/-

13. *Analisis Pembezalayan:*

(a) $Y = l'X = (\mu_1 - \mu_2)' \sum^{-1} X$

$$m = \frac{1}{2}(\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2)$$

(b) $y = \hat{l}'X = (\bar{x}_1 - \bar{x}_2)' S_p^{-1} x$

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

(c) *Petua peruntukan:*

$$\text{Untukkan } x_0 \text{ kepada } \begin{cases} \pi_1 \text{ jika } y_0 \geq \hat{m} \\ \pi_2 \text{ jika } y_0 < \hat{m} \end{cases}$$

(d) $B_0 = \sum_{i=1}^{g} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})$

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

$e_1, \ \ldots, \ e_s$ *vektor eigen* $\sum^{-1} B_0$.

$\ell_i X = e_i X$ *pembezalayan ke* $-i, i = 1, 2, \ldots, s$

(e) $B_0 = \sum_{i=1}^{g} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$

$$W = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)'$$

$\hat{\ell}_i x = \hat{e}_i x$ *pembezalayan ke* $-i, i = 1, 2, \ldots, s.$

(f) *Petua peruntukan :*

Untukkan x *kepada* π_k *jika*

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

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

bagi semua $i \neq k, \quad r \leq s.$

- 000 O 000 -