Second Semester Examination
2017/2018 Academic Session

May / June 2018

MSG466 - Multivariate Analysis
(Analisis Multivariat)

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of **FIFTEEN (15)** pages of printed material before you begin the examination.

*[Sila pastikan bahawa kertas peperiksaan ini mengandungi **LIMA BELAS (15)** muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]*

Instructions : Answer **all six (6)** questions.

*[**Arahan :** Jawab **semua enam (6)** 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].

...2/-

Question 1

(a) Write down three properties that must be satisfied by a distance measure, $d(P,Q)$ between two points *P* and *Q*.

(b) Are the following distance function valid for distance from the origin? Explain. If it is a distance function, sketch the ellipse.

$$4x_1^2 + 3x_2^2 - 2\sqrt{2}x_1x_2 = 1$$

[8 marks]

Soalan 1

(a) Tulis tiga sifat yang mesti dipenuhi oleh suatu ukuran jarak, $d(P,Q)$ *antara dua titik P dan Q.*

(b) Adakah fungsi jarak berikut suatu fungsi jarak dari asal? Terangkan. Jika ia adalah fungsi jarak, lakarkan elipsnya.

$$4x_1^2 + 3x_2^2 - 2\sqrt{2}x_1x_2 = 1$$

[8 markah]

Question 2

Suppose the random vector **X** has multivariate normal distribution $\mathrm{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & r & r^2 \\ r & 1 & r \\ r^2 & r & 1 \end{pmatrix}.$$

(a) Find the distribution of $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$.

(b) Find the conditional distribution of X_1 and X_3 given $X_2 = x_2$.

(c) Are $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ and X_2 independent?

[10 marks]

Soalan 2

Andaikan vektor rawak **X** *mempunyai taburan normal multivariat* $\mathrm{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *dengan*

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & r & r^2 \\ r & 1 & r \\ r^2 & r & 1 \end{pmatrix}.$$

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

(b) *Cari taburan bersyarat bagi* X_1 *dan* X_3 *diberi* $X_2 = x_2$.

(c) *Adakah* $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ *dan* X_2 *tidak bersandar?*

[10 markah]

Question 3

The data in Table 1 (Q3tbl1.mtw) shows judges' scores on fish prepared by three methods. Twelve fish were cooked by each method, and several judges tasted fish samples and rated each on four variables: y_1 = aroma, y_2 = flavor, y_3 = texture, and y_4 = moisture. Each entry in Table 1 is an average score for the judges on that fish.

Table 1 Judges' Score on Fish Prepared by Three Methods

Method 1				Method 2				Method 3			
y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
5.4	6.0	6.3	6.7	5.0	5.3	5.3	6.5	4.8	5.0	6.5	7.0
5.2	6.2	6.0	5.8	4.8	4.9	4.2	5.6	5.4	5.0	6.0	6.4
6.1	5.9	6.0	7.0	3.9	4.0	4.4	5.0	4.9	5.1	5.9	6.5
4.8	5.0	4.9	5.0	4.0	5.1	4.8	5.8	5.7	5.2	6.4	6.4
5.0	5.7	5.0	6.5	5.6	5.4	5.1	6.2	4.2	4.6	5.3	6.3
5.7	6.1	6.0	6.6	6.0	5.5	5.7	6.0	6.0	5.3	5.8	6.4
6.0	6.0	5.8	6.0	5.2	4.8	5.4	6.0	5.1	5.2	6.2	6.5
4.0	5.0	4.0	5.0	5.3	5.1	5.8	6.4	4.8	4.6	5.7	5.7
5.7	5.4	4.9	5.0	5.9	6.1	5.7	6.0	5.3	5.4	6.8	6.6
5.6	5.2	5.4	5.8	6.1	6.0	6.1	6.2	4.6	4.4	5.7	5.6
5.8	6.1	5.2	6.4	6.2	5.7	5.9	6.0	4.5	4.0	5.0	5.9
5.3	5.9	5.8	6.0	5.1	4.9	5.3	4.8	4.4	4.2	5.6	5.5

(a) Construct the one-way MANOVA table and test the hypothesis that there is no difference in average score among the three methods. Give your conclusion and state any assumptions you make.

(b) Perform a discriminate analysis on the data in Table 1.

(c) What can you conclude from your results obtained in part (b)?

[25 marks]

Soalan 3

Data dalam Jadual 1 (Q3tbl1.mtw) menunjukkan skor pengadil pada ikan yang disediakan dengan tiga kaedah. Dua belas ikan dimasak oleh setiap kaedah, dan beberapa pengadil merasakan sampel ikan dan memberi nilai masing-masing pada empat pembolehubah: y_1 *= aroma,* y_2 *= rasa,* y_3 *= tekstur, dan* y_4 *= kelembapan. Setiap catatan dalam Jadual 1 adalah skor purata oleh para pengadil pada ikan tersebut.*

Jadual 1 *Skor Pengadil terhadap Ikan yang Disediakan dengan Tiga Kaedah*

Kaedah 1				*Kaedah 2*				*Kaedah 3*			
y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
5.4	6.0	6.3	6.7	5.0	5.3	5.3	6.5	4.8	5.0	6.5	7.0
5.2	6.2	6.0	5.8	4.8	4.9	4.2	5.6	5.4	5.0	6.0	6.4
6.1	5.9	6.0	7.0	3.9	4.0	4.4	5.0	4.9	5.1	5.9	6.5
4.8	5.0	4.9	5.0	4.0	5.1	4.8	5.8	5.7	5.2	6.4	6.4
5.0	5.7	5.0	6.5	5.6	5.4	5.1	6.2	4.2	4.6	5.3	6.3
5.7	6.1	6.0	6.6	6.0	5.5	5.7	6.0	6.0	5.3	5.8	6.4
6.0	6.0	5.8	6.0	5.2	4.8	5.4	6.0	5.1	5.2	6.2	6.5
4.0	5.0	4.0	5.0	5.3	5.1	5.8	6.4	4.8	4.6	5.7	5.7
5.7	5.4	4.9	5.0	5.9	6.1	5.7	6.0	5.3	5.4	6.8	6.6
5.6	5.2	5.4	5.8	6.1	6.0	6.1	6.2	4.6	4.4	5.7	5.6
5.8	6.1	5.2	6.4	6.2	5.7	5.9	6.0	4.5	4.0	5.0	5.9
5.3	5.9	5.8	6.0	5.1	4.9	5.3	4.8	4.4	4.2	5.6	5.5

(a) Bina jadual MANOVA sehala dan uji hipotesis bahawa tidak ada perbezaan skor purata antara tiga kaedah tersebut. Berikan kesimpulan anda dan nyatakan sebarang andaian yang telah dibuat.

(b) Lakukan suatu analisis diskriminasi bagi data dalam Jadual 1.

(c) Apakah yang boleh anda simpulkan daripada keputusan yang diperolehi dalam bahagian (b)?

[25 markah]

Question 4

The data in Table 2 consist of measurement on three variables y_1, variable soil calcium, y_2, exchangeable soil calcium, and y_3, turnip green calcium at 10 different locations in Malaysia.

Table 2 Calcium in Soil and Turnip green

Location Number	y_1	y_2	y_3
1	35	3.5	2.80
2	35	4.9	2.70
3	40	30.0	4.38
4	10	2.8	3.21
5	6	2.7	2.73
6	20	2.8	2.81
7	35	4.6	2.88
8	35	10.9	2.90
9	35	8.0	3.28
10	30	1.6	3.20

(a) The following linear combinations are defined for the variables in Table 2:

$$z_1 = y_1 + y_2 + y_3,$$
$$z_2 = 2y_1 - 3y_2 + 2y_3,$$
$$z_3 = -y_1 - 2y_2 - 3y_3.$$

(i) Find the sample mean vector of the **z**, $\bar{\mathbf{z}}$.

(ii) Find the sample covariance matrix of the **z**, $\mathbf{s_z}$.

(iii) Find the sample correlation matrix of the **z**, $\mathbf{R_z}$.

(b) Suppose $\mathbf{Y} \sim \mathrm{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Based on the data in Table 2, find the maximum likelihood estimates mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$.

(c) Test at the 5% level the null hypothesis that $\boldsymbol{\mu}' = (30, 7, 3)$. State any assumptions you make before performing the test.

(d) Obtain the 95% simultaneous confidence intervals for the three population mean calcium measurements.

(e) Obtain the 95% Bonferroni confidence intervals for the three mean in part (d).

(f) Compare the confidence intervals obtain in parts (d) and (e).

[17 marks]

Soalan 4

Data dalam Jadual 2 terdiri daripada pengukuran bagi tiga pembolehubah, y_1*, pembolehubah kalsium tanah,* y_2*, kalsium tanah yang boleh ditukar, dan* y_3*, turnip kalsium hijau di 10 lokasi di Malaysia.*

***Jadual 2** Kalsium dalam Tanah dan Turnip hijau*

nombor lokasi	y_1	y_2	y_3
1	35	3.5	2.80
2	35	4.9	2.70
3	40	30.0	4.38
4	10	2.8	3.21
5	6	2.7	2.73
6	20	2.8	2.81
7	35	4.6	2.88
8	35	10.9	2.90
9	35	8.0	3.28
10	30	1.6	3.20

(a) *Gabungan linear berikut didefinisikan untuk pembolehubah dalam Jadual 2:*

$$z_1 = y_1 + y_2 + y_3,$$
$$z_2 = 2y_1 - 3y_2 + 2y_3,$$
$$z_3 = -y_1 - 2y_2 - 3y_3.$$

(i) *Cari vektor min sampel bagi z,* $\bar{\mathbf{z}}$.

(ii) *Cari matriks kovarians sampel bagi z,* $\mathbf{s_z}$.

(iii) *Cari matriks korelasi sampel bagi z,* $\mathbf{R_z}$.

(b) *Andaikan* $\mathbf{Y} \sim \mathrm{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Berdasarkan data dalam Jadual 2, dapatkan anggaran kebolehjadian maksimum bagi vektor min* $\boldsymbol{\mu}$ *dan matriks kovarian* $\boldsymbol{\Sigma}$.

(c) *Uji pada aras 5% hipotesis nol bahawa* $\boldsymbol{\mu}' = (30, 7, 3)$. *Nyatakan sebarang andaian yang telah anda buat sebelum melakukan ujian tersebut.*

(d) *Dapatkan selang keyakinan serentak 95% bagi tiga min pengukuran kalsium populasi.*

(e) *Dapatkan selang keyakinan Bonferroni 95% untuk tiga min dalam bahagian (d).*

(f) *Bandingkan selang-selang keyakinan yang diperolehi dalam bahagian (d) and (e).*

[17 markah]

Question 5

The data in Table 3 (Q5tbl3.mtw) consist of 130 observations generated by scores on a psychological test administered to Peruvian teenagers (ages 15, 16, and 17). For each of these teenagers the scores were accumulated into five subscale labeled independence (indep), support (supp), benevolence (benev), conformity (conform), and leadership (leader). Perform a principal component analysis using the covariance matrix **S** and the correlation matrix **R**. Interpret your results.

Table 3 Psychological Profile Data

Indep	Supp	benev	conform	leader
27	13	14	20	11
12	13	24	25	6
14	20	15	16	7
⋮	⋮	⋮	⋮	⋮
19	11	23	18	13
27	19	22	7	9
10	17	22	22	8

[15 marks]

Soalan 5

Data dalam Jadual 3 (Q5tbl3.mtw) mengandungi 130 cerapan yang dihasilkan daripada skor ujian psikologi yang diberikan kepada remaja Peru (umur 15, 16, dan 17). Bagi setiap remaja ini, markah telah dikumpulkan ke dalam lima sub skala yang dilabelkan kebebasan (indep), sokongan (supp), kebajikan (benev), kesesuaian (conform), dan kepimpinan (leader). Lakukan analisis komponen utama dengan menggunakan matriks kovarians S dan matriks korelasi R. Tafsirkan keputusan anda.

Jadual 3 *Data Profil Psikologi*

Indep	*Supp*	*benev*	*conform*	*leader*
27	13	14	20	11
12	13	24	25	6
14	20	15	16	7
⋮	⋮	⋮	⋮	⋮
19	11	23	18	13
27	19	22	7	9
10	17	22	22	8

[15 markah]

Question 6

Table 4 (Q6tbl4.mtw) shows air pollution data from 41 US cities. The variables are as follows:

y_1 = SO_2 content of air in micrograms per cubie meter,
y_2 = Average annual temperature in ^{o}F,
y_3 = Number of manufacturing enterprise employing 20 or more workers,
y_4 = Population size (1970 census) in thousands,
y_5 = Average annual wind speed in miles per hour,
y_6 = Average annual precipitation in inches,
y_7 = Average number of days with precipitation per year.

Table 4 Air pollution Level in US Cities

Cities	y_1	y_2	y_3	y_4	y_5	y_6	y_7
Phoenix	10	70.3	213	582	6.0	7.05	36
Little Rock	13	61.0	91	132	8.2	48.52	100
San Francisco	12	56.7	453	716	8.7	20.66	67
Denver	17	51.9	454	515	9.0	12.95	86
Hartford	56	49.1	412	158	9.0	43.37	127
Wilmington	36	54.0	80	80	9.0	40.25	114
Washington	29	57.3	434	757	9.3	38.89	111
Jacksonville	14	68.4	136	529	8.8	54.47	116
Miami	10	75.5	207	335	9.0	59.8	128
Atlanta	24	61.5	368	497	9.1	48.34	115
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Nashville	18	59.4	275	448	7.9	46.00	119
Dallas	9	66.2	641	844	10.9	35.94	78
Houston	10	68.9	721	1233	108	48.19	103
Salt Lake City	28	51.0	137	176	8.7	15.17	89
Norfolk	31	59.3	96	308	10.6	44.68	116
Richmond	26	57.8	197	299	7.6	42.59	115
Seattle	29	51.1	379	531	9.4	38.79	164
Charleston	31	55.2	35	71	6.5	40.75	148
Milwaukee	16	45.7	569	717	11.8	29.07	123

(a) Perform a factor analysis of the data in Table 4.

(b) Perform a suitable cluster analysis of the data in Table 4.

(c) Interpret the results obtained in parts (a) and (b).

[25 marks]

Soalan 6

Jadual 4 (Q6tbl4.mtw) *menunjukkan data pencemaran udara daripada 41 bandar di US. Pembolehubah-pembolehubah adalah seperti berikut:*

y_1 = *kandungan udara dalam mikrogram setiap meter cubie,*

y_2 = *Purata suhu tahunan dalam* $^{\circ}\mathrm{F}$,

y_3 = *Bilangan perusahaan perkilangan yang mempunyai 20 atau lebih pekerja,*

y_4 = *Saiz penduduk (penapisan 1970) dalam ribu,*

y_5 = *Purata kelajuan angin tahunan dalam batu sejam,*

y_6 = *Purata hujan tahunan dalam inci,*

y_7 = *Purata bilangan hari hujan setahun.*

Jadual 4 Tahap Pencemaran Udara di bandar US

Bandar	y_1	y_2	y_3	y_4	y_5	y_6	y_7
Phoenix	10	70.3	213	582	6.0	7.05	36
Little Rock	13	61.0	91	132	8.2	48.52	100
San Francisco	12	56.7	453	716	8.7	20.66	67
Denver	17	51.9	454	515	9.0	12.95	86
Hartford	56	49.1	412	158	9.0	43.37	127
Wilmington	36	54.0	80	80	9.0	40.25	114
Washington	29	57.3	434	757	9.3	38.89	111
Jacksonville	14	68.4	136	529	8.8	54.47	116
Miami	10	75.5	207	335	9.0	59.8	128
Atlanta	24	61.5	368	497	9.1	48.34	115
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Nashville	18	59.4	275	448	7.9	46.00	119
Dallas	9	66.2	641	844	10.9	35.94	78
Houston	10	68.9	721	1233	108	48.19	103
Salt Lake City	28	51.0	137	176	8.7	15.17	89
Norfolk	31	59.3	96	308	10.6	44.68	116
Richmond	26	57.8	197	299	7.6	42.59	115
Seattle	29	51.1	379	531	9.4	38.79	164
Charleston	31	55.2	35	71	6.5	40.75	148
Milwaukee	16	45.7	569	717	11.8	29.07	123

(a) *Lakukan analisis faktor bagi data dalam Jadual 4.*

(b) *Lakukan analisis kluster yang sesuai bagi data dalam Jadual 4.*

(c) *Tafsirkan keputusan yang diperoleh dalam bahagian (a) dan (b).*

[25 markah]

APPENDIX / LAMPIRAN

Summary of Multivariate Formulae / *Ringkasan Rumus-Rumus Multivariat*

1. Sample means: $\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}$, where $\bar{x}_k = \dfrac{1}{n}\sum_{j=1}^{n} x_{jk},\ k = 1, 2, \ldots, p$.

2. Sample variances and covariances:

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

where $s_{ik} = \dfrac{1}{n-1}\sum_{j=1}^{n}\left(x_{ji} - \bar{x}_i\right)\left(x_{jk} - \bar{x}_k\right) \qquad i = 1, 2, \ldots, p \quad k = 1, 2, \ldots, p$.

3. The sample correlation coefficient:

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = \frac{\sum_{j=1}^{n}\left(x_{ji} - \bar{x}_i\right)\left(x_{jk} - \bar{x}_k\right)}{\sqrt{\sum_{j=1}^{n}\left(x_{ji} - \bar{x}_i\right)^2}\sqrt{\sum_{j=1}^{n}\left(x_{jk} - \bar{x}_k\right)^2}},$$

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

5. Multivariate normal p.d.f.:

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

6. If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

   (a) $\mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$

   (b) $\mathbf{AX} \sim N_q(\mathbf{A}\boldsymbol{\mu},\ \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

(c) $\mathbf{X}+\mathbf{d} \sim N_p(\boldsymbol{\mu}+\mathbf{d}, \boldsymbol{\Sigma})$

(d) $\mathbf{AX}+\mathbf{d} \sim N_q(\mathbf{A}\boldsymbol{\mu}+\mathbf{d},\ \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

(e) $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \sim \chi_p^2$

7. If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_p\left(\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$ and $|\boldsymbol{\Sigma}_{22}| > \mathbf{0}$. 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 = $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

8. $\mathbf{X}_j \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$, $j = 1, 2, \ldots, n$ be mutually independent. Then $\mathbf{V}_1 = \sum_{j=1}^{n} c_j \mathbf{X}_j \sim N_p\left(\sum_{j=1}^{n} c_j \boldsymbol{\mu}_j, \ \left(\sum_{j=1}^{n} c_j^2 \right) \boldsymbol{\Sigma} \right)$. Moreover, $\mathbf{V}_1$ and $\mathbf{V}_2 = \sum_{j=1}^{n} b_j \mathbf{X}_j$ are jointly multivariate normal with covariance matrix $\begin{bmatrix} \left(\sum_{j=1}^{n} c_j^2 \right) \boldsymbol{\Sigma} & (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} \\ (\mathbf{b}'\mathbf{c})\boldsymbol{\Sigma} & \left(\sum_{j=1}^{n} b_j^2 \right) \boldsymbol{\Sigma} \end{bmatrix}$.

9. One-sample results:

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

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

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

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

$$\bar{x}_i \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha) \frac{s_{ii}}{n}}$$

(c) The 100(1 $-\alpha$)% Bonferroni intervals for the component mean μ_i :

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

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

$$\bar{x}_i \pm \sqrt{\chi^2{}_p(\alpha)}\sqrt{\frac{s_{ii}}{n}}$$

10. Two-sample results (paired comparison):

(a) $T^2 = n\,\bar{\mathbf{d}}'\,\mathbf{S}_d^{-1}\,\bar{\mathbf{d}}$ and $T^2 \sim \dfrac{(n-1)p}{(n-p)}F_{p,n-p}$

(b) The $100(1-\alpha)\%$ simultaneous confidence intervals for the individual mean differences δ_i:

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

(c) The Bonferroni $100(1-\alpha)\%$ simultaneous confidence intervals for the individual mean differences δ_i :

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

11. Two-sample results (independent samples):

(a) $$T^2 = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - \boldsymbol{\delta}_0)'\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\right]^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - \boldsymbol{\delta}_0)$$

where $\mathbf{S}_{\text{pooled}} = \dfrac{n_1-1}{n_1+n_2-2}\mathbf{S}_1 + \dfrac{n_2-1}{n_1+n_2-2}\mathbf{S}_2$,

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

(b) A $100(1-\alpha)\%$ confidence region (ellipsoid centered at $\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$) for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is determined by all $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ such that

$$\left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right)'\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\right]^{-1}\left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right) \le \frac{(n_1+n_2-2)p}{(n_1+n_2-p-1)}F_{p,n_1+n_2-p-1}(\alpha)$$

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

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

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

(d) The Bonferroni 100(1 $-\alpha$)% simultaneous confidence intervals for $\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$:

$$\mathbf{a}'(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) \pm t_{n_1+n_2-2}(\alpha/2p)\sqrt{\mathbf{a}'\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\,\mathbf{a}}$$

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

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

where $c^2 = \chi_p^2(\alpha)$

12. One-way MANOVA:

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

$$\mathbf{W} = \sum_{l=1}^{g}\sum_{j=1}^{n_l}(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)'$$

$$= (n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2 + \ldots + (n_g-1)\mathbf{S}_g$$

Distribution of Λ^* :

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

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

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

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

$n = \sum n_l$

13. Two-way MANOVA

SSCP:

$$\sum_{l=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(\mathbf{x}_{lkr}-\bar{\mathbf{x}})(\mathbf{x}_{lkr}-\bar{\mathbf{x}})' = \sum_{l=1}^{g} bn(\bar{\mathbf{x}}_{l\cdot}-\bar{\mathbf{x}})(\bar{\mathbf{x}}_{l\cdot}-\bar{\mathbf{x}})' + \sum_{k=1}^{b} gn(\bar{\mathbf{x}}_{\cdot k}-\bar{\mathbf{x}})(\bar{\mathbf{x}}_{\cdot k}-\bar{\mathbf{x}})'$$
$$+ \sum_{l=1}^{g}\sum_{k=1}^{b} n(\bar{\mathbf{x}}_{lk}-\bar{\mathbf{x}}_{l\cdot}-\bar{\mathbf{x}}_{\cdot k}+\bar{\mathbf{x}})(\bar{\mathbf{x}}_{lk}-\bar{\mathbf{x}}_{l\cdot}-\bar{\mathbf{x}}_{\cdot k}+\bar{\mathbf{x}})'$$
$$+ \sum_{l=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n}(\mathbf{x}_{lkr}-\bar{\mathbf{x}}_{lk})(\mathbf{x}_{lkr}-\bar{\mathbf{x}}_{lk})'$$

$$\text{SSCP}_{\text{cor}} = \text{SSCP}_{\text{fac 1}} + \text{SSCP}_{\text{fac 2}} + \text{SSCP}_{\text{int}} + \text{SSCP}_{\text{res}}$$

(a) No Interaction effect:

$$\Lambda^* = \frac{|\text{SSCP}_{\text{res}}|}{|\text{SSCP}_{\text{int}} + \text{SSCP}_{\text{res}}|}$$

$$-\left[gb(n-1) - \frac{p+1-(g-1)(b-1)}{2}\right]\ln\Lambda^* > \chi^2_{(g-1)(b-1)}(\alpha)$$

(b) No factor 1 effect:

$$\Lambda^* = \frac{|\text{SSCP}_{\text{res}}|}{|\text{SSCP}_{\text{fac 1}} + \text{SSCP}_{\text{res}}|}$$

$$-\left[gb(n-1) - \frac{p+1-(g-1)}{2}\right]\ln\Lambda^* > \chi^2_{(g-1)p}(\alpha)$$

14. Euclidean (straight-line) distance:

$$d(\mathbf{x},\mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2}$$
$$= \sqrt{(\mathbf{x}-\mathbf{y})'(\mathbf{x}-\mathbf{y})}$$

15. Minkowski metric:

$$d(\mathbf{x},\mathbf{y}) = \left[\sum_{i=1}^{p} \left| x_i - y_i \right|^m \right]^{1/m}$$

- oooOOooo -