MATH 4424 Multivariate Statistical Analysis
Course Outline --- Fall 2023

1. **Instructor(s)**

Name: Dr. XIA, Dong (L1)
Contact Details: Room 3477, tel: ext-2244, email: madxia@ust.hk
If you want to make an appointment with Dr. XIA, send him an email.

2. **Teaching Assistant(s)**

Name: SHEN, Yinan (T1A), email: yshenay@connect.ust.hk
If you want to make an appointment with your TA, send him an email.

3. **Meeting Time and Venue**

Lectures:

Date/Time: L1: TuTh. 9:00AM - 10:20AM
Venue: Lecture Theatre H

Tutorials:
Date/Time: T1A: Th. 7:00PM - 7:50PM;
Rm 2463 (lift 25-26)
(R programming is important. Our TA will give several tutorial sessions in the first few weeks for the basics of R programming.)

4. **Course Description**

Credit Points: 3
Pre-requisite: MATH 3423 and MATH 3424
Exclusion: ISOM 5530

Brief Information/synopsis:
Inferences of means and covariance matrices, canonical correlation, discriminant analysis, multivariate ANOVA, principal components analysis, factor analysis

5. **Intended Learning Outcomes**

Upon successful completion of this course, students should be able to:

No.	ILOs
1	Get familiar with multivariate normal distribution, random sample, the distributions of sample mean and covariance, Linear combination of normal random variables
2	Understand the testing of mean vector, Hotelling's test, Likelihood ratio test, Confidence regions and Large sample inference, Comparing mean vectors from two populations, and one-way MANOVA
3	Understand the population principal components, Sample variation summarisation
4	Understand Canonical variates, correlation, interpretation of canonical variables, Sample canonical correlation
5	Understand classification, Linear and quadratic classification rules, Evaluating classification performance

6. **Assessment Scheme**

a. Percentage of coursework, examination, etc.:

Assessment 1	Assessing Course ILOs
15% Assignments	1,2,3,4,5
30% Mid-term Test	1,2,3
55% Final Examination	1,2,3,4,5

Assessment 2	Assessing Course ILOs
15% Assignments	1,2,3,4,5
0% Mid-term Test (in case the student cannot attend the midterm)	1,2,3
85% Final Examination (in case the student cannot attend the midterm)	1,2,3,4,5

The final grade=max{ Assessment 1, Assessment 2}
(You can bring a computer to final exam, but your screen must ONLY shows the Rstudio interface.)

7. Student Learning Resources

Lecture notes and slides (on canvas)
Textbook: Johnson and Wichern, *Applied Multivariate Statistical Analysis*, Sixth Edition (The book is not required.)

8. Teaching and Learning Activities

Scheduled activities: 4 hrs (lecture + tutorial)

9. Tentative Course Schedule

Keyword Syllabus:

- **Chaps 1, 2 and 3 Introduction, Basic Prerequisite and Sample Geometry** (~6 hours)
  (i) Why multivariate analysis, data display and organisation and distance; (ii) Prerequisite on matrix algebra; (iii) Positive definite matrix and covariance matrix (iv) Geometry of sample; (v) Random sample; (vi) Sample mean, covariance and correlation; (vii) Linear combination of normal variables;

- **Chap 4 Multivariate Normal Distribution** (~5 hours)
  (i) Multivariate normal density and property; (ii) Maximum Likelihood Estimator; (iii) Sampling distribution of sample mean and covariance; (iv) Assess the assumption of normal distribution; (v) Detect outliers and data cleaning; (vi) Data transformation to near normality; (vii) Examples with R codes

- **Chap 5 Inference about a Mean Vector** (~4 hours)
  (i) Test the mean vector; (ii) Hotelling's test; (iii) Likelihood ratio test; (iv) Confidence region and simultaneous confidence interval; (v) Large sample inference; (vi) Examples with R codes

- **Chap 6 Comparisons of Several Multivariate Means** (~4 hours)
  (i) Paired comparison; (ii) Compare mean vectors from two populations; (iii) Compare several multivariate population means (MNOVA); (iv) Examples with R codes

- **Chap 7 Multiple Linear Regression** (~2 hours)
  (i) Least squares estimation; (ii) Inferences about regression model; (iii) Inferences from estimated regression function;

- **Chap 8 Principal Component Analysis** (~4 hours)
  (i) Population principal components; (ii) Summarizing sample variation by principal components; (iii) Cautions and scalings; (iv) Large sample inference; (v) Number of principal components; (vi) Examples with R codes

- **Chap 9 Canonical Correlation Analysis** (~3 hours)
  (i) Canonical variates and correlation; (ii) Interpret the population canonical variables; (iii) Sample canonical variates and sample canonical correlation; (iv) Large Sample Inference; (v) Examples with R codes

- **Chap 10 Discrimination and Classification** (~5 hours)
  (i) Separation and classification for two populations; (ii) Classification with two multivariate normal population; (iii) Evaluate classification functions; (iv) Classification with several populations; (v) Examples with R codes