FACULTY OF SCIENCE

SCHOOL OF MATHEMATICS
AND STATISTICS

MATH5855

MULTIVARIATE ANALYSIS

Semester 2 2018

MATH5855 – Course Outline

Information about the course

Course Authority: Associate Professor Spiridon Penev

e-mail `S.Penev@unsw.edu.au`

Room: RC-1038

Consultation: Thursday 11:00-12:30, Friday 11:00-12:30.

Individual times can also be arranged-please use email to arrange an appointment.

Credit, Prerequisites, Exclusions:

This course counts for 6 Units of Credit (6UOC). It is an optional component of Master of Statistics, Master of Biostatistics and Master of Financial Mathematics. MATH2801, MATH2901 or MATH5846 & MATH5856 is assumed minimal knowledge for this course. However once you have been admitted to the postgraduate program of the Department of Statistics, there are no further prerequisites.

Lectures: There will be three hours of lectures per week.

Wednesday	17:00-20:00	OMB149

Lectures will start in week 1 and continue until week 12. (The first lecture is on 25th July).

Tutorials: Tutorial and labs (using `SAS` and `R`) for this course are flexible and will be held in the time slot for the lectures. More precise information will be given during lectures.

Online materials: Further information, skeleton lecture notes, and other materials will be provided via Moodle.

Course aims

This course will give you a solid methodological background in Multivariate Analysis as a backbone of Applied Statistics. You will learn the theoretical foundations of the most commonly applied multivariate techniques such as mean vector and covariance matrix estimation and testing, estimation and testing of correlations, discriminant analysis, classification and Support Vector Machines, principal components, canonical correlations analysis, cluster analysis, factor analysis and structural equations. You will study the properties and the importance of the multivariate normality assumption in the context of each of these methods. `SAS`-based computing will feature prominently in the course. At the end of the course you should be able to use all of

the above techniques in your work as applied statistician, for practical analysis of real datasets.

Relation to other statistics courses

Being a backbone of Applied Statistics, this course can be very useful and is recommended for all the remaining postgraduate coursework programs of the Department of Statistics. It is particularly relevant in combination with courses such as MATH5806, MATH5836, MATH5845, MATH5945.

Student Learning Outcomes

- Able to use the general terminology, notation and concepts in the theory, methods and applications of Multivariate Analysis.

- Able to use creatively the properties of the multivariate normal distribution to justify optimality properties of Statistical Inference procedures based on multivariate normality.

- Able to formulate and solve inference problems that use one or a combination of the following multivariate statistical procedures: inferences about mean vector and the covariance matrix, inferences about correlations and partial correlations, canonical correlation analysis, discrimination and/or classification problems, clustering, principal components, factor analysis and covariance structure models.

- Able to perform the above multivariate inference procedures by using the available powerful procedures in the `SAS` system.

- Able to write simple `SAS` instructions about data input and output and to do a coding of your own simple multivariate analyses by using `SAS/IML`.

- Apply the multivariate techniques and models to the analysis of datasets, interpret the results and draw conclusions.

These outcomes are closely related to the graduate attributes "Research, inquiry and analytical thinking abilities" and "Information literacy" (through the computing component of the course).

Relation to graduate attributes

These outcomes are closely related to the graduate attributes "Research, inquiry and analytical thinking abilities", "Communication" and "Information literacy" (through the computing component of the course).

Teaching strategies underpinning the course

Lecture notes provide a brief reference source for this course. New ideas and skills are first introduced and demonstrated in lectures, then students develop these skills by applying them to specific tasks in tutorials and assessments. Computing skills are developed and practised in computer practical sessions.

Rationale for learning and teaching strategies

We believe that effective learning is best supported by a climate of inquiry, in which students are actively engaged in the learning process. Hence this course is structured with a strong emphasis on problem-solving tasks in tutorials and in assessment tasks, and students are expected to devote the majority of their study time to the solving of such tasks.

Effective learning is achieved when students attend all classes, have prepared effectively for classes by reading through previous lecture notes and by having made a serious attempt at doing for themselves the tutorial problems.

Furthermore, lectures should be viewed by students as an opportunity to learn, rather than just copy down or skim over lecture notes.

Assessment

Knowledge and abilities assessed: All assessment tasks will assess the learning outcomes outlined above. Assessment in this course will use problem-solving tasks of a similar form to those practised in lectures, tutorials and labs, to encourage the development of the core analytical and computing skills underpinning this course and the development of analytical thinking.

Assessment

Assessment task	%	Available	Due	Notes
Assignment 1	10	Week 3	24th August (Week 5)	No late assignments
Assignment 2	10	Week 6	21st September (Week 9)	No late assignments
Assignment 3	10	Week 9	12th October (Week 11)	No late assignments
Final exam	70	N/A	TBA	

In all assessments, marks will be awarded for correct working and appropriate explanations and not just the final answer.

Assignments

Rationale: Assignments will give an opportunity for students to try their hand at more difficult problems requiring more than one line of argument and also introduce them to aspects of the subject which are not explicitly covered in lectures.

Assignments must be YOUR OWN WORK, or severe penalties will be incurred.

You should consult the University web page on plagiarism

Late assignments will not be accepted.

Examination

Duration: Three hours.

Rationale: The final examination will assess student mastery of the material covered in the lectures and tutorials.

The examination will be held in the computer lab and will involve both computing and theoretical component.

Further details about the final examination will be available in class closer to the time.

Additional resources and support

Tutorial Exercises

A set of tutorial exercises will be available on Moodle. These problems are for YOU to do to enhance mastery of the course.

SOME of the problems will be done in tutorials, but you will learn a lot more if you try to do them before the tutorial.

Lectures and lecture notes

A set of skeleton notes will be provided on Moodle. Formally, each week, there will be three hours of lectures. However, a small part of the lectures will be used for solving exercises/tutorial type of problems. Sets of tutorial questions will be given out from time to time for your individual work at home (these will not be collected and marked and partial solutions or hints to them will be given later). Approximately one hour every other week will be spent in the computer lab RC-G012C. More precise information will be given as we go, during lecture time.

Textbooks

The textbooks are listed below according to their relative importance for this course.

JW Johnson, R. A. and Wichern, D.W. *Applied Multivariate Statistical Analysis*, Sixth Edition, Prentice Hall 2007. (This is the recommended text).

HS Härdle, W. and Simar, L. *Applied Multivariate Statistical Analysis*, Third Edition, Springer 2012.

A Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, Third Edition, Wiley 2003.

HFT Hastie, T., Friedman, J. and Tibshirani, R. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Second Edition, Springer 2009.

ED Der, G. and Everitt, B. *A Handbook of Statistical Analyses using SAS*. Third Edition, CRC Press 2009.

LN Lecture notes : Lecture notes will be given out in lectures.

Most of the material will be in the lecture notes. From the textbooks, the two recommended text **JW** and **HS** would be most useful. The reference **HFT** contains some recent developments in Multivariate Analysis. It is available as an e-book from the UNSW library.

Moodle

Skeleton lecture notes and other useful materails will be available on Moodle. You should check regularly for updates. Some notes and tutorial solutions may be handed out as a hard copy only.

Computer laboratories

Computer laboratories (RC-G012 and RC-M020) are open 9-5 Monday-Friday on teaching days. RC-M020 has extended teaching hours (usually 8:30-9pm Monday-Friday, and 9-5 Monday-Friday on non-teaching weeks).

Course Evaluation and Development

The School of Mathematics and Statistics evaluates each course each time it is run. We carefully consider the student responses and their implications for course development. It is common practice to discuss informally with students how the course and their mastery of it are progressing.

Administrative matters

Additional Assessment

See attached handout.

School Rules and Regulations

Fuller details of the general rules regarding attendance, release of marks, special consideration etc are available via the School of Mathematics and Statistics Web page at
`http://www.maths.unsw.edu.au/currentstudents/assessment-policies`.

Plagiarism and academic honesty

Plagiarism is the presentation of the thoughts or work of another as one's own. Issues you must be aware of regarding plagiarism and the university's policies on academic honesty and plagiarism can be found at
`http://www.maths.unsw.edu.au/currentstudents/assessment-policies`.

Occupational Health and safety

Please refer to the UNSW Occupational Health and Safety policies and expectations:
`http://www.gs.unsw.edu.au/policy/documents/ohspolicy.pdf`.

Equity and diversity

Any equity and diversity issues should be directed to the Student Equity Officers (Disability) in the Student Equity and Diversity Unit (9385-4734). Further information for students with disabilities is available at

`http://www.studentequity.unsw.edu.au/`.

Detailed course schedule

It is intended that the following topics will be covered in the given order. Any variation from this will be indicated by the lecturer. Some material may not be possible to cover in details and in such cases, at the discretion of the lecturer, part of some lectures will be used as a Problem Class where instead of new theoretical material, just some examples are shown.

	Topic	Week (approx)
1.	Introduction and Linear Algebra Background	1
2.	Multivariate Distributions. The Multivariate Normal	2
3.	Estimation of the Mean Vector and the Covariance Matrix	3
5.	Generalized T^2 Statistics and Applications	4
5.	Distributions and uses of Sample Correlation Coefficients. Copulae	4
6.	Discriminant Analysis and linear Classification	5
7.	Support Vector Machines and non-linear classification	6
8.	Distribution of Sample Covariance Matrix. Tests on covariance matrices	7
9.	Principal Component Analysis and Applications	7-8
10.	Canonical Correlation Analysis	9
11.	Factor Analysis	10
12.	Covariance Structure Analysis	11
13.	Cluster Analysis	12