FACULTY OF SCIENCE

SCHOOL OF MATHEMATICS
AND STATISTICS

MATH5855

MULTIVARIATE ANALYSIS

Trimester 3, 2019
Information about the course

Course Authority

Dr Pavel Krivitsky

E-mail: p.krivitsky@unsw.edu.au

Room: RC-1032

Consultation:

- Tuesday 14:00–15:30
- Friday 13:00–14:30
- 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 four hours of lectures per week:

- Tuesdays 17:00-19:00 in K-H6-LG05
- Friday 15:00-17:00 in K-H13-M032

Lectures will start in week 1 and continue until week 10. (The first lecture is on 17th September).

Tutorials

Tutorial and labs (using R and SAS) 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. R-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 R and/or SAS system.

• Able to write simple R instructions about data input and output and to do a coding of your own simple multivariate analyses by using R.

• Familiar with 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

<table>
<thead>
<tr>
<th>Assessment task</th>
<th>%</th>
<th>Available</th>
<th>Due</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>10</td>
<td>Week 1</td>
<td>4 October (Week 3, Friday)</td>
<td>No late assignments</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>10</td>
<td>Week 6</td>
<td>22 November (Week 10, Friday)</td>
<td>No late assignments</td>
</tr>
<tr>
<td>Mid-session test</td>
<td>20</td>
<td>Week 6</td>
<td></td>
<td>No late assignments</td>
</tr>
<tr>
<td>Final exam</td>
<td>60</td>
<td>N/A</td>
<td>TBA</td>
<td></td>
</tr>
</tbody>
</table>

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.

Note that while the instructor will endeavour to provide detailed feedback on all assessment items, due to time constraints, it may not be practical, and only a selection of the items may count towards the assessment mark.

Mid-session Test

Duration: Two hours, in class.

Rationale: The Mid-session Test will give students feedback on their progress and mastery of the material.
Both short answer questions and some longer questions requiring clear and logical presentation of correct solutions will be given.

If you are absent from the test, you must provide a medical certificate. No further test will be offered.

**Examination**

**Duration:** Two 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 four 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 two hours every other week will be spent in the computer lab RC-M032. 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.


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 materials 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


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. Due to the transition to UNSW3+ schedule, the exact weeks are subject to change.

Some material may not be possible to cover in detail 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.
<table>
<thead>
<tr>
<th>Topic</th>
<th>Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction and Linear Algebra Background</td>
<td>1</td>
</tr>
<tr>
<td>2. Multivariate Distributions. The Multivariate Normal</td>
<td>1–2</td>
</tr>
<tr>
<td>3. Estimation of the Mean Vector and the Covariance Matrix</td>
<td>2–3</td>
</tr>
<tr>
<td>4. Generalized $T^2$ Statistics and Applications</td>
<td>3</td>
</tr>
<tr>
<td>5. Distributions and uses of Sample Correlation Coefficients. Copulae</td>
<td>4</td>
</tr>
<tr>
<td>6. Discriminant Analysis and linear Classification</td>
<td>4–5</td>
</tr>
<tr>
<td>7. Support Vector Machines and non-linear classification</td>
<td>5–6</td>
</tr>
<tr>
<td>8. Distribution of Sample Covariance Matrix. Tests on covariance matrices</td>
<td>7</td>
</tr>
<tr>
<td>9. Principal Component Analysis and Applications</td>
<td>7–8</td>
</tr>
<tr>
<td>10. Canonical Correlation Analysis</td>
<td>8–9</td>
</tr>
<tr>
<td>11. Factor Analysis</td>
<td>9</td>
</tr>
<tr>
<td>12. Covariance Structure Analysis</td>
<td>9–10</td>
</tr>
<tr>
<td>13. Cluster Analysis</td>
<td>10</td>
</tr>
</tbody>
</table>