



STAT302

Graphics, Multivariate Methods and Data Mining

S2 Day 2014

Statistics

Contents

<u>General Information</u>	2
<u>Learning Outcomes</u>	2
<u>Assessment Tasks</u>	3
<u>Delivery and Resources</u>	5
<u>Unit Schedule</u>	6
<u>Policies and Procedures</u>	7
<u>Graduate Capabilities</u>	8
<u>Changes from Previous Offering</u>	13
<u>Teaching and Learning Strategy</u>	13

Disclaimer

Macquarie University has taken all reasonable measures to ensure the information in this publication is accurate and up-to-date. However, the information may change or become out-dated as a result of change in University policies, procedures or rules. The University reserves the right to make changes to any information in this publication without notice. Users of this publication are advised to check the website version of this publication [or the relevant faculty or department] before acting on any information in this publication.

General Information

Unit convenor and teaching staff

Ayse Bilgin

ayse.bilgin@mq.edu.au

Contact via E-mail

E4A 528

Credit points

3

Prerequisites

39cp including (STAT270(P) or STAT271(P) or BIOL235(P) or PSY222(P) or PSY248(P))

Corequisites

Co-badged status

Unit description

This unit is concerned with the structure of multivariate data which is explored graphically, analysed statistically and investigated using data mining methods. Multivariate methods covered include cluster analysis, principal components, and discriminant analysis. Knowledge of simple matrix algebra, although not essential, will be very helpful in understanding and working through these topics. Statistical packages are used extensively to illustrate the concepts in lectures and tutorials.

Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are available at <https://www.mq.edu.au/study/calendar-of-dates>

Learning Outcomes

On successful completion of this unit, you will be able to:

- Understand the principles underlying graphics, multivariate methods and data mining;
- Choose the appropriate statistical analysis, for a given data set, from a wide range of methods based on multivariate methods and data mining;
- Choose appropriate graphical techniques for displaying data;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Assessment Tasks

Name	Weighting	Due
<u>Participation</u>	5%	Weekly
<u>Presentation</u>	5%	Weekly
<u>Assignments (x2)</u>	30%	Week 5 & Week 11
<u>Final Examination</u>	60%	Exam Timetable

Participation

Due: **Weekly**

Weighting: **5%**

*Weeks 2 – 12 inclusive.

Every week lecture and tutorial participation will be monitored and most weeks there will be set homework to submit to the lecturer at the start of the following lecture.

On successful completion you will be able to:

- Understand the principles underlying graphics, multivariate methods and data mining;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Presentation

Due: **Weekly**

Weighting: **5%**

Each student will be given an opportunity to present to class either in the Lecture or in the Tutorial. The presentations will start on Week 2 until each student presents at least once.

Each week there will be three student presentations. Students are encouraged to volunteer for these presentations so that a timetable could be created, in the case of no one volunteering a random selection will be used. The three presentations will in the form of no more than two power point slides. The specifics of the presentations are:

1. Summary of the previous week's lecture (at the beginning of the lecture)
2. The most important aspects of the previous week's lecture and/or tutorial (at the beginning of the tutorial)
3. The most important aspects of the current week's lecture and/or tutorial (at the end of the

tutorial)

On successful completion you will be able to:

- Understand the principles underlying graphics, multivariate methods and data mining;
- Apply statistical techniques to problems arising from diverse fields of research.

Assignments (x2)

Due: **Week 5 & Week 11**

Weighting: **30%**

There will be two individual assignments due in weeks 5 and 11. The assignment questions will be made available through iLearn.

There is no “group work” assessment in this unit, however students are encouraged to work together and help each other to learn.

On successful completion you will be able to:

- Understand the principles underlying graphics, multivariate methods and data mining;
- Choose the appropriate statistical analysis, for a given data set, from a wide range of methods based on multivariate methods and data mining;
- Choose appropriate graphical techniques for displaying data;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Final Examination

Due: **Exam Timetable**

Weighting: **60%**

The examination will examine any material covered throughout the course. Students may bring one A4 sized sheet of hand written notes, formulae, etc., which may be written on both sides and is easily readable. This summary must be submitted with your exam paper and is marked. No other materials such as lecture notes and textbooks are permitted.

Calculators will be needed but must not be of the text/programmable type.

On successful completion you will be able to:

- Understand the principles underlying graphics, multivariate methods and data mining;
- Choose the appropriate statistical analysis, for a given data set, from a wide range of methods based on multivariate methods and data mining;

- Choose appropriate graphical techniques for displaying data;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Delivery and Resources

Lectures begin in Week 1.

Tutorials begin in Week 2.

The timetable for classes can be found on the University web site at: <http://www.timetables.mq.edu.au/>

Recommended Texts and/or Materials

There are no **required** texts for this unit, but the following list provides useful references, which are available in the Library.

Weeks 1-6 Material

Chambers J M et al (1983) Graphical Methods for Data Analysis. (QA276.3 .G73/1983)

Cleveland W S (1994) Elements of Graphing Data. (QA90 .C54/1994)

Cleveland W S & McGill M E (1988) Dynamic Graphics for Statistics. (QA276.3.D96/1988)

Du Toit S H C et al (1986) Graphical Exploratory Data Analysis. (QA276.3 .D778/1986)

Ehrenberg A S C (1982) Primer in Data Reduction. (QA276.12 .E37/1982)

Tufte E R (2001) The Visual Display of Quantitative Information. (QA276.3 T8 2001)

Weeks 7-13 Material

Everitt B S et al (2001) Applied multivariate data analysis. (QA278 .E914/2001)

Johnson, D.E (1998) Applied Multivariate Methods for Data Analysts. (QA278.J615/1998)

Johnson, R.A. & Wichern, D.W. (2002) Applied Multivariate Statistical Analysis. (QA278 .J63/2002)

Manly, B F J (2004) Multivariate Statistical Methods - A Primer. (QA278 .M35 2004)

iLearn

There is an iLearn site for this subject that will contain all the required course materials and allows communication between students. You can access the unit iLearn site from the address <http://learn.mq.edu.au> using your Student ID number and myMQ Portal password. You can only access the material if you are enrolled in the unit. If you have any problems accessing this website, go to the Online Teaching Facility support web page at <http://online.mq.edu.au/docs/teconf.html>

The lecturer will make announcements via the iLearn. You should regularly log in and read the posts at least twice a week.

The Discussion Board on iLearn can be used to communicate with other students.

Software

The main software packages will be used are IBM SPSS Analytics, IBM SPSS Modeler, R (a dialect of S-Plus which is open source, it is available from <http://cran.r-project.org/>), Mondrian (open source, available from <http://rosuda.org/mondrian/Mondrian.html>), Excel & Microsoft Word.

Students are expected to

- attend all the lectures and the tutorials;
- **submit** solutions to weekly homework exercises through iLearn by due date;
- contact the lecturer in advance if for any reason, students cannot submit their assessment tasks on time;
- for the assignments students expected to submit a soft copy through iLearn and hand in a hard copy to the lecturer at the beginning of the lecture;
- Collect their marked assessments and have a discussion with their peer and/or lecturer to improve their learning based on the feedback provided on the assessment.

Unit Schedule

Week	Topic
1	Introduction & presenting data numerically
2	Good and bad graphical displays
3	Choosing different graphic displays
4	Displaying multivariate data
5	Similarities and distances
6	Hierarchical cluster analysis
7	K-means clustering
	Mid-semester break - two weeks
8	Eigenvalues and eigenvectors
9	Principal component analysis

10	Discriminant analysis
11	Multiple discriminant analysis
12	Classification and regression trees
13	Review

The order of the lectures might change.

Policies and Procedures

Macquarie University policies and procedures are accessible from [Policy Central](#). Students should be aware of the following policies in particular with regard to Learning and Teaching:

Academic Honesty Policy http://mq.edu.au/policy/docs/academic_honesty/policy.html

Assessment Policy <http://mq.edu.au/policy/docs/assessment/policy.html>

Grading Policy <http://mq.edu.au/policy/docs/grading/policy.html>

Grade Appeal Policy <http://mq.edu.au/policy/docs/gradeappeal/policy.html>

Grievance Management Policy http://mq.edu.au/policy/docs/grievance_management/policy.html

Disruption to Studies Policy http://www.mq.edu.au/policy/docs/disruption_studies/policy.html *The Disruption to Studies Policy is effective from March 3 2014 and replaces the Special Consideration Policy.*

In addition, a number of other policies can be found in the [Learning and Teaching Category](#) of Policy Central.

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: https://students.mq.edu.au/support/student_conduct/

Student Support

Macquarie University provides a range of support services for students. For details, visit <http://students.mq.edu.au/support/>

Learning Skills

Learning Skills (mq.edu.au/learningskills) provides academic writing resources and study strategies to improve your marks and take control of your study.

- [Workshops](#)
- [StudyWise](#)

- [Academic Integrity Module for Students](#)
- [Ask a Learning Adviser](#)

Student Services and Support

Students with a disability are encouraged to contact the [Disability Service](#) who can provide appropriate help with any issues that arise during their studies.

Student Enquiries

For all student enquiries, visit Student Connect at ask.mq.edu.au

IT Help

For help with University computer systems and technology, visit <http://informatics.mq.edu.au/help/>.

When using the University's IT, you must adhere to the [Acceptable Use Policy](#). The policy applies to all who connect to the MQ network including students.

Graduate Capabilities

Capable of Professional and Personal Judgement and Initiative

We want our graduates to have emotional intelligence and sound interpersonal skills and to demonstrate discernment and common sense in their professional and personal judgement. They will exercise initiative as needed. They will be capable of risk assessment, and be able to handle ambiguity and complexity, enabling them to be adaptable in diverse and changing environments.

This graduate capability is supported by:

Learning outcomes

- Understand the principles underlying graphics, multivariate methods and data mining;
- Choose the appropriate statistical analysis, for a given data set, from a wide range of methods based on multivariate methods and data mining;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Assessment tasks

- Participation
- Presentation
- Assignments (x2)
- Final Examination

Commitment to Continuous Learning

Our graduates will have enquiring minds and a literate curiosity which will lead them to pursue knowledge for its own sake. They will continue to pursue learning in their careers and as they participate in the world. They will be capable of reflecting on their experiences and relationships with others and the environment, learning from them, and growing - personally, professionally and socially.

This graduate capability is supported by:

Learning outcomes

- Choose the appropriate statistical analysis, for a given data set, from a wide range of methods based on multivariate methods and data mining;
- Choose appropriate graphical techniques for displaying data;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Assessment tasks

- Participation
- Assignments (x2)
- Final Examination

Discipline Specific Knowledge and Skills

Our graduates will take with them the intellectual development, depth and breadth of knowledge, scholarly understanding, and specific subject content in their chosen fields to make them competent and confident in their subject or profession. They will be able to demonstrate, where relevant, professional technical competence and meet professional standards. They will be able to articulate the structure of knowledge of their discipline, be able to adapt discipline-specific knowledge to novel situations, and be able to contribute from their discipline to inter-disciplinary solutions to problems.

This graduate capability is supported by:

Learning outcomes

- Understand the principles underlying graphics, multivariate methods and data mining;
- Choose the appropriate statistical analysis, for a given data set, from a wide range of methods based on multivariate methods and data mining;
- Choose appropriate graphical techniques for displaying data;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for

publication;

- Apply statistical techniques to problems arising from diverse fields of research.

Assessment tasks

- Participation
- Presentation
- Assignments (x2)
- Final Examination

Critical, Analytical and Integrative Thinking

We want our graduates to be capable of reasoning, questioning and analysing, and to integrate and synthesise learning and knowledge from a range of sources and environments; to be able to critique constraints, assumptions and limitations; to be able to think independently and systemically in relation to scholarly activity, in the workplace, and in the world. We want them to have a level of scientific and information technology literacy.

This graduate capability is supported by:

Learning outcomes

- Understand the principles underlying graphics, multivariate methods and data mining;
- Choose the appropriate statistical analysis, for a given data set, from a wide range of methods based on multivariate methods and data mining;
- Choose appropriate graphical techniques for displaying data;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Assessment tasks

- Participation
- Presentation
- Assignments (x2)
- Final Examination

Problem Solving and Research Capability

Our graduates should be capable of researching; of analysing, and interpreting and assessing data and information in various forms; of drawing connections across fields of knowledge; and they should be able to relate their knowledge to complex situations at work or in the world, in order to diagnose and solve problems. We want them to have the confidence to take the initiative in doing so, within an awareness of their own limitations.

This graduate capability is supported by:

Learning outcomes

- Choose the appropriate statistical analysis, for a given data set, from a wide range of methods based on multivariate methods and data mining;
- Choose appropriate graphical techniques for displaying data;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Assessment tasks

- Participation
- Presentation
- Assignments (x2)
- Final Examination

Creative and Innovative

Our graduates will also be capable of creative thinking and of creating knowledge. They will be imaginative and open to experience and capable of innovation at work and in the community. We want them to be engaged in applying their critical, creative thinking.

This graduate capability is supported by:

Learning outcomes

- Choose appropriate graphical techniques for displaying data;
- Apply statistical techniques to problems arising from diverse fields of research.

Assessment tasks

- Participation
- Presentation
- Assignments (x2)

Effective Communication

We want to develop in our students the ability to communicate and convey their views in forms effective with different audiences. We want our graduates to take with them the capability to read, listen, question, gather and evaluate information resources in a variety of formats, assess, write clearly, speak effectively, and to use visual communication and communication technologies as appropriate.

This graduate capability is supported by:

Learning outcomes

- Choose appropriate graphical techniques for displaying data;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Assessment tasks

- Participation
- Presentation
- Assignments (x2)
- Final Examination

Engaged and Ethical Local and Global citizens

As local citizens our graduates will be aware of indigenous perspectives and of the nation's historical context. They will be engaged with the challenges of contemporary society and with knowledge and ideas. We want our graduates to have respect for diversity, to be open-minded, sensitive to others and inclusive, and to be open to other cultures and perspectives: they should have a level of cultural literacy. Our graduates should be aware of disadvantage and social justice, and be willing to participate to help create a wiser and better society.

This graduate capability is supported by:

Learning outcomes

- Choose the appropriate statistical analysis, for a given data set, from a wide range of methods based on multivariate methods and data mining;
- Use a statistical computer package to carry out chosen analyses and interpret the results with understanding; present the results of analyses in a form which is suitable for publication;
- Apply statistical techniques to problems arising from diverse fields of research.

Assessment tasks

- Participation
- Presentation
- Assignments (x2)

Socially and Environmentally Active and Responsible

We want our graduates to be aware of and have respect for self and others; to be able to work with others as a leader and a team player; to have a sense of connectedness with others and country; and to have a sense of mutual obligation. Our graduates should be informed and active

participants in moving society towards sustainability.

This graduate capability is supported by:

Learning outcomes

- Understand the principles underlying graphics, multivariate methods and data mining;
- Apply statistical techniques to problems arising from diverse fields of research.

Assessment tasks

- Participation
- Assignments (x2)
- Final Examination

Changes from Previous Offering

Random class tests are replaced by student presentations. Practical test is cancelled. Assignment weights are increased from %5 to 15% each.

Teaching and Learning Strategy

All unit related queries should be directed to the unit convenor Dr Nino Kordzakhia using the Macquarie University e-mail system.

Lectures begin in Week 1.

Tutorials (1 x 2 hour tutorial) will start in the second week. In weeks 2 to 13 you will be required to submit homework and this work will also count towards your assessment.

The timetable for classes can be found on the University web site at <https://timetables.mq.edu.au/2013/>

Students are expected to:

- attend all the lectures (beginning in Week 1) and tutorials (beginning in Week 2);
- submit assignment solutions by due dates;
- submit homeworks by due dates;
- contact the unit convenor in advance if, for any reason, they cannot hand in their assessment tasks on time.