DIVISION OF ECONOMIC AND FINANCIAL STUDIES

STAT302: GRAPHICS, MULTIVARIATE METHODS AND DATA MINING

Second Semester, 2008 – D2

UNIT OUTLINE

Unit convenor: Dr Tania Prvan

Prerequisites: STAT270(P) or STAT271(P) or BIOL235(P) or PSY222(P)

Students in this unit should read this unit outline carefully at the start of semester. It contains important information about the unit. If anything in it is unclear, please consult one of the teaching staff in the unit.

ABOUT THIS UNIT

This unit consists of two modules concerned with the structure of multivariate data, one analytical, the other graphical.

The graphical module provides an introduction to a selection of topics related to new computer-based displays of multivariate data. Topics include: table presentation, principles of good graphical design, the scatterplot matrix, and dot charts for one and two-way classified data.

The multivariate analysis module provides an introduction to selected topics in multivariate analysis; namely, cluster analysis, principal components, multidimensional scaling and discriminant analysis. Knowledge of simple matrix algebra, although not essential, would be very helpful in understanding and working through these topics. Extensive use will be made of statistical packages to illustrate the concepts in lectures and tutorials.

Software
SPSS, Clementine, Minitab & Microsoft Word.

TEACHING STAFF

Convenor: Dr Tania Prvan, E4A 518, ph: 9850-8561, tprvan@efs.mq.edu.au
Co-Convenor: Ms Suzanne Curtis, E5A 549, ph: 9850-8584, scurtis@efs.mq.edu.au
Tutor: Mrs Bala Pasupathy, bpasu@efs.mq.edu.au

Please note that any communication with staff via email will only be conducted using your official university email address. Consultation hours will be finalised at the end of Week 1. These will be posted on the web site.

CLASSES

Lectures
Lectures begin in Week 1. Lectures are held on Mondays between 11:00am and 1:00pm in room E4B 102.

Tutorials
1 x 2 hour tutorial. Tutorials will start in the second week and are held on Thursday between 11am and 1pm in E4B 102. 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: http://www.timetables.mq.edu.au/

REQUIRED AND RECOMMENDED TEXTS AND/OR MATERIALS

Prescribed texts
There are no prescribed texts for this unit, but the following list provides useful references, which are available in Special Reserve in the Library.

Recommended texts:
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)

Weeks 7-13 Material
Ehrenberg A S C (1982)*Primer in Data Reduction.* (QA276.12 .E37/1982)
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)
Tufte E R (2001) *The Visual Display of Quantitative Information*. (QA276.3 T8 2001)

UNIT WEB PAGE

The web page for this unit can be found at: http://www.stat.mq.edu.au/units/stat302/. Blackboard facility is available. Please check this regularly.

LEARNING OUTCOMES

The learning outcomes of this unit are 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.

In addition to the discipline-based learning objectives, all academic programs at Macquarie seek to develop students' generic skills in a range of areas. One of the aims of this unit is that students develop their skills in the following:

- problem solving skills;
- written communication skills, particularly report writing;
- critical analysis skills.

TEACHING AND LEARNING STRATEGY

The unit is taught in traditional mode; that is, on campus in standard semesters with weekly lectures and tutorials (in the computer laboratory).

Students are expected to

- attend all the lectures and the tutorials;
- hand in the solutions to weekly homework exercises to ERIC (Economic Resource & Information Centre) E4B 106;
- contact the lecturer in advance if for any reason, students cannot hand in their assessment tasks on time;
- hand in the assignments to ERIC (Economic Resource & Information Centre) E4B 106;
- Collect their marked assessment from ERIC (Economic Resource & Information Centre) E4B 106.

Refer to end of this handout for a week-by-week list of topics to be covered in this unit.

RELATIONSHIP BETWEEN ASSESSMENT AND LEARNING OUTCOMES

While attendance at classes is important it is only a small proportion of the total workload for the unit: reading, working with other students in groups, completing assignments, using the computer and private study are all part of the work involved. At Macquarie it is expected that the average student should spend four hours per week per credit point.

The assessment is based on performance in the tutorials (including homework), practical test, assignments, and the final examination.

Assessment	**Weighting**	**Due**
Three Assignments	15%	Weeks 3, 7 & 11
Homework and Tutorial Participation	10%	Each week
Practical Test	15%	Week 13 tutorial
Final Examination	60%	As timetabled

Marked assessment will be available for collection from ERIC (Economic Resource & Information Centre) E4B 106 approximately two weeks after the due date.

Late assessments will only be accepted with the agreement of the lecturer and may be subject to the deduction of some marks.

NOTE: To obtain a passing grade, both coursework and exam performance must be satisfactory.

The University Examination period in Semester 2 is from 19th November, 2008 to 5h December, 2008.

You are expected to present yourself for examination at the time and place designated in the University Examination Timetable. The timetable will be available in Draft form approximately eight weeks before the commencement of the examinations and in Final form approximately four weeks before the commencement of the examinations.
http://www.timetables.mq.edu.au/exam

The only exception to not sitting an examination at the designated time is because of documented illness or unavoidable disruption. In these circumstances you may wish to consider applying for Special Consideration. Information about unavoidable disruption and the special consideration process is available at http://www.reg.mq.edu.au/Forms/APSCon.pdf

If a Supplementary Examination is granted as a result of the Special Consideration process the examination will be scheduled after the conclusion of the official examination period. (Individual Divisions may wish to signal when the Division's Supplementary examinations are normally scheduled.)

You are advised that it is Macquarie University policy not to set early examinations for individuals or groups of students. All students are expected to ensure that they are available until the end of the teaching semester; that is, the final day of the official examination period.

PLAGIARISM

The University defines plagiarism in its rules: "Plagiarism involves using the work of another person and presenting it as one's own." Plagiarism is a serious breach of the University's rules and carries significant penalties. You must read the University's practices and procedures on plagiarism. These can be found in the *Handbook of Undergraduate Studies* or on the web at: http://www.student.mq.edu.au/plagiarism/

The policies and procedures explain what plagiarism is, how to avoid it, the procedures that will be taken in cases of suspected plagiarism, and the penalties if you are found guilty. Penalties may include a deduction of marks, failure in the unit, and/or referral to the University Discipline Committee.

UNIVERSITY POLICY ON GRADING

Academic Senate has a set of guidelines on the distribution of grades across the range from fail to high distinction. Your final result will include one of these grades plus a standardised numerical grade (SNG).

On occasion your raw mark for a unit (i.e., the total of your marks for each assessment item) may not be the same as the SNG which you receive. Under the Senate guidelines, results may be scaled to ensure that there is a degree of comparability across the

university, so that units with the same past performances of their students should achieve similar results.

It is important that you realise that the policy does not require that a minimum number of students are to be failed in any unit. In fact it does something like the opposite, in requiring examiners to explain their actions if more than 20% of students fail in a unit.

The process of scaling does not change the order of marks among students. A student who receives a higher raw mark than another will also receive a higher final scaled mark.

For an explanation of the policy see
http://www.mq.edu.au/senate/MQUonly/Issues/Guidelines2003.doc or
http://www.mq.edu.au/senate/MQUonly/Issues/detailedguidelines.doc.

STUDENT SUPPORT SERVICES

Macquarie University provides a range of Academic Student Support Services. Details of these services can be accessed at http://www.student.mq.edu.au.

COMPUTER LABS AND CONDITIONS OF USE

Please see www.efs.mq.edu.au/current/ug/resources/labs.

Unit Schedule

WEEK	TOPIC AND SUGGESTED READING	WORK DUE
1 (4 August)	Introduction & presenting data numerically	
2 (11 Aug)	Good and bad graphical displays	
3 (18 Aug)	Choosing different graphic displays	**Assignment 1**
4 (25 Aug)	Displaying multivariate data	
5 (1 Sept)	Similarities and distances	
6 (8 Sept)	Hierarchical cluster analysis	
7 (15 Sept)	K-means clustering	**Assignment 2**
	MIDSEMESTER BREAK – Two Weeks	
8 (6 Oct)	**Public Holiday**	
9 (13 Oct)	Eigenvalues and eigenvectors	
10 (20 Oct)	Principal component analysis	
11 (27 Oct)	Discriminant analysis	**Assignment 3**
12 (3 Nov)	Multiple discriminant analysis	
13 (10 Nov)	Classification and regression trees	**Test (15%)**

Homework and tutorial participation: Weeks 2 - 13