



**UNSW**  
THE UNIVERSITY OF NEW SOUTH WALES

**FACULTY OF SCIENCE  
SCHOOL OF MATHEMATICS AND  
STATISTICS**

**MATH2871**

**Data Management for Statistical  
Analysis**

Session 2, 2009



# MATH2871 – Course Outline

## Information about the course

**Course Authority and lecturer:** Mr. Nathan Pearce, Room RC-4091, phone 9385 6905, email [nathan@maths.unsw.edu.au](mailto:nathan@maths.unsw.edu.au)

**Consultation:** Wednesdays, 4-5pm.

**Credit, Prerequisites, Exclusions:**

This course counts for 6 Units of Credit (6UOC).

Prerequisites: MATH1041 or ECON1203 or ECON2292 or PSYC2001 or MATH1231 or MATH1241 or MATH1251 or equivalent. There are no exclusions for this course.

## Aims

The course, a collaborative venture of the School of Mathematics with SAS, aims to provide a practical introduction to the management and analysis of data. Large data sets are now found widely in business, finance, bioinformatics, government, intelligence and elsewhere. Skills in querying, cleaning, managing, displaying and analysing data, which are widely sought, will be developed in this course.

The course will provide the opportunity for SAS certification in Base Programming (as well as normal academic credit) (cost for the SAS exam: about \$80). In addition, during the summer after the course SAS will operate a work experience placement program.

## Assessment

Assessment in this course will use problem-solving tasks of a similar form to those practised in laboratory sessions, to encourage the development of the core skills underpinning this course and the development of analytical thinking. Assessment in this course will consist of

- Attendance at laboratories worth 10% (see below)
- 4 hand-in assignments worth 10% each
- The final exam (2 hours) is worth 50%.

Attendance at laboratories: 10% will be given for attendance at 9 or more laboratories; 5% will be awarded for attendance at 6–8 laboratories.

## Location and Times

- Lectures, Mechanical Engineering 303 (K-J17-303), weeks 1–12, Wednesday 10–12
- Laboratories, RC-M020, also weeks 1–12:
  - Wednesday 12–1pm
  - Wednesday 1–2pm

## Outcomes

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 lectures, in tutorials and in assessment tasks and students are expected to devote the majority of their class and study time to the solving of such tasks. New ideas and skills are first introduced and demonstrated in lectures, and then students develop these skills by applying them to specific tasks in laboratories. Computing skills are developed and practised in the weekly computing sessions.

## webCT

All course materials will be available on webCT. You should check regularly for new materials.

## 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

### 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

[www.lc.unsw.edu.au/plagiarism](http://www.lc.unsw.edu.au/plagiarism)

The schedule for assignments is as follows-

Task	Week avail.	Date Due	Weighting
Assignment 1	Week 2	Week 4	10%
Assignment 2	Week 4	Week 6	10%
Assignment 3	Week 6	Week 9	10%
Assignment 4	Week 9	Week 11	10%

The assignments are due at Wednesday, 10:05am. Late assignments will not be accepted.

Every class is different, and to accommodate this, some variation from the above schedule may be prudent. Hence the above schedule should be considered as a guide only, as it will not be strictly adhered to. In the case of assessment dates, no changes will be made without consultation with the class nor without confirmation being posted as an announcement on webCT Vista.

### **Textbooks**

Although there is no set text for this course, the following books may prove useful:

Delwiche, L. D. and Slaughter, S. J. (2008) The Little SAS Book: A Primer, 4th edition, SAS Publishing, Cary NC.

### **webCT**

All course materials, including lecture notes, will be available on webCT. You should check regularly for new materials.

## **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

## School rules and regulations

Fuller details of the general rules regarding attendance, release of marks, special consideration, additional assessment etc are available via the School of Mathematics and Statistics Web page at

<http://www.maths.unsw.edu.au/students/current/policies/studentpolicy.html>

## 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.lc.unsw.edu.au/plagiarism> and [http://www.lc.unsw.edu.au/plagiarism/plagiarism\\_STUDENTBOOK.pdf](http://www.lc.unsw.edu.au/plagiarism/plagiarism_STUDENTBOOK.pdf). Syllabus Lecture notes and laboratory exercises for this course will be made available via the webCT Vista web site, <http://vista.elearning.unsw.edu.au>. They are not a substitute for attendance at lectures and laboratories. The course content is ultimately defined by the material covered in lectures. The course will include material from the following.

1. Introduction: Programming Basics, SAS windows environment, files used by SAS, SAS data libraries.
2. SAS programs: Components, running programs, diagnosing errors.
3. Producing List Reports: PRINT procedure, sequencing and grouping observations, identifying observations.
4. Creating and Reading SAS datasets: Reading raw and Excel data files, examining errors, assigning and changing variable attributes, reading SAS datasets, Concatenating and merging datasets.
5. Producing Summary Reports: Basic reports, accumulating totals.
6. Controlling Input and Output: Outputting multiple observations, writing to multiple datasets, selecting variables and observations.
7. Data Transformations: Manipulating character and numeric values
8. Iterative Processing: DO loops, arrays
9. Combining Datasets
10. Introduction to Graphics: Bar charts, pie charts, scatterplots

## **Supply of SAS software**

SAS software is available for home use. See the General Office of the School of Mathematics on the 3rd floor of the Red Centre and to be supplied with a free DVD containing a copy of SAS 9. The copy is compatible with Windows XP, though appears to be compatible with only some versions of Windows Vista. SAS is installed on the Windows booted computers in RC-M020.