Faculty of Science
School of Mathematics and Statistics

MATH2871
Data Management for Statistical Analysis

Semester 2, 2014
MATH2871 – Course Outline

Information about the course

Course Authority and lecturer: Dr. Leung Lung Chan, Room RC-1036, phone 9385 7021, email leung.chan@unsw.edu.au

Consultation: Times will be announced in Week 1.

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 $90). 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 and active participation at laboratories worth 10% (see below)
- 4 hand-in assignments worth 10% each
- The final exam (2 hours) is worth 50%.

Attendance and active participation at laboratories: 10% will be given for attendance at 9 or more laboratories, and showing active participation; 5% will be awarded for attendance at 6–8 laboratories, and showing active participation.
Location and Times

- Lectures, Central Lecture Block 2, weeks 1–12, Thursday 9am–11am
- Laboratories, weeks 2–13:
  - Wednesday 11am–12pm, Red Centre Central Wing G12C
  - Wednesday 12pm–1pm, Red Centre Central Wing M020
  - Thursday 11am–12pm, Red Centre Central Wing M020

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.

Moodle

All course materials will be available on Moodle, https://moodle.telt.unsw.edu.au/login/index.php/. They are not a substitute for attendance at lectures and laboratories. 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.

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

The schedule for assignments is as follows:

<table>
<thead>
<tr>
<th>Task</th>
<th>Week avail.</th>
<th>Date Due</th>
<th>Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>Week 2</td>
<td>Week 4</td>
<td>10%</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>Week 4</td>
<td>Week 6</td>
<td>10%</td>
</tr>
<tr>
<td>Assignment 3</td>
<td>Week 6</td>
<td>Week 9</td>
<td>10%</td>
</tr>
<tr>
<td>Assignment 4</td>
<td>Week 9</td>
<td>Week 11</td>
<td>10%</td>
</tr>
</tbody>
</table>

The assignments are due at Thursday, 9:05am. 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 Moodle.

**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.

**Course contents**

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 University Edition software is available for home use. This free offering, launched a month ago, allows students an easy-to-access, easy-to-use option into the power of SAS. If you want to know more and download the software please visit the University Edition website at http://www.sas.com/en_us/software/university-edition.html

Additional References

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