MATH5885
LONGITUDINAL DATA ANALYSIS
Semester 1, 2013
MATH5885 – Course Outline

Information about the course

Course Authority: William Dunsmuir

Lecturer: William Dunsmuir
  Room RC-2057
  Phone 9385 7035
  email W.Dunsmuir@unsw.edu.au.

Consultation: Monday 4-5pm Office 2-057.

Credit, Prerequisites, Exclusions:

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

There are no prerequisites for this course.

There are no exclusions for this course.

Teaching Format: There will be one 3.0 hour teaching session per week: Monday 5:00-8:00pm. Classes will start with formal lectures in the assigned lecture room (for approximately 2 hours depending on the material to be covered). After a short break classes will resume in the Computer Lab RC-M020 (for the balance of time until 8pm).

Tutorials: There are no separate tutorials.

My Blackboard: Almost all course material (such as lecture notes, assignments, project descriptions, lab instructions for R and SAS work, data and some other readings) will be provided via Blackboard. Material for the following week will be posted, at the latest, by Thursday of the current week.

Course aims

Longitudinal data arise when multiple measurements of a response are collected over time for each individual in the study. The aim of this course is to introduce the main statistical concepts, methods and models used in the analysis of longitudinal data.

Relation to other mathematics and statistics courses

This course is an elective course for the Master of Statistics program and a core course for the Master of Biostatistics program. Whilst there are no formal prerequisites for the course, it does make use of techniques based on statistical theory such as maximum likelihood estimation and hypothesis testing, and some familiar-
ity with these ideas is assumed. In this sense the course is related to MATH5905, Statistical Inference, which covers the theory behind some of the techniques we will apply. The course is also related to MATH5806, Applied Regression Analysis, which deals with modelling of relationships between variables, and MATH5855, Multivariate Analysis, which covers techniques for joint modelling of multiple responses and to MATH5845 Time Series Analysis which covers techniques for modelling data collected through time on one or a few series. The focus in MATH5885 is on modelling response profiles over time, whilst accounting for the correlation that will typically exist between measurements taken on the same individual. Longitudinal data are frequently collected as part of a clinical trial or an observational study, so the techniques learned will have applications in the types of studies covered in MATH5906, Design and Analysis of Clinical Trials, and MATH5826, Statistical Methods in Epidemiology.

**Student Learning Outcomes**

By the end of this course, you should be able to:

- Explain the key features of longitudinal data and what distinguishes them from cross-sectional data
- Use exploratory data analysis techniques to visualize patterns in longitudinal data using R
- Explain the different types of correlation that can occur in longitudinal data, and choose appropriate models for correlation structures in given datasets
- Derive the likelihood function for correlated data and use it to calculate maximum likelihood estimates
- Explain the concept of residual maximum likelihood estimation (REML) and use it to calculate REML estimates
- Show how linear models incorporating correlated errors can be used to analyse longitudinal data
- Explain the theory of linear mixed effects models for analysing continuous longitudinal data, and show how this extends to generalized linear mixed effects models for analysing non-normal longitudinal data
- Explain how marginal models can be used to analyse non-normal longitudinal data via generalized estimating equations
- Explain the distinction between a “subject-specific” and “population-average” interpretation of parameters
• Classify the different types of missing data that occur in longitudinal studies and explain the implications for analysis
• Use R and SAS to fit the models studied to longitudinal datasets, and interpret the output
• Carry out estimation and inference for the models studied
• Check the validity of a model for longitudinal data
• Solve theoretical problems related to longitudinal data analysis

Relation to graduate attributes

Computing skills developed in this course will improve information literacy (Science Graduate Attribute 6).

Assignments, problems, and lab exercises will develop research, inquiry and analytical thinking abilities (Science Graduate Attribute 1).

Teaching strategies underpinning the course

To support the learning outcomes, the course will use the following teaching strategies:

• Lectures, explaining the necessary statistical concepts and theory applicable to longitudinal studies
• Computer labs, providing essential practice in applying the techniques explained in lectures
• Independent study of the course notes and readings, to reflect more deeply on ideas introduced in lectures
• Problems and assignments, giving you an opportunity to independently solve theoretical problems, analyse datasets using statistical software, reflect on aspects of the course, and evaluate your understanding
• Assessment in this course will use problem-solving tasks of a similar form to those in practice problems and computer labs, to encourage the development of the core analytical and computing skills underpinning this course.
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 computer labs and in assessment tasks, and students are expected to devote the majority of their class and study time to the solving of such tasks.

Assessment

Assessment in this course will consist of two assignments (10% each), a substantial project (20%), and a final examination (60%).

Knowledge and abilities assessed: All assessment tasks will assess the learning outcomes outlined above, specifically, the ability to explain the concepts and theory underlying statistical techniques for longitudinal data, to apply the techniques in analysing real datasets and critically interpret the results of analyses, and to solve theoretical problems related to the statistical aspects of longitudinal studies.

Assessment criteria: The main criteria for marking assessment tasks involving explanation of theory and solution of theoretical problems will be clear and logical presentation of correct solutions. In the case of assessment tasks involving the application of techniques to the analysis of real datasets, the main criteria will be selection and justification of appropriate analysis methods; clear, logical, and well-documented computer code; well-organised output giving evidence of successful implementation; correct interpretation of results; and clear, complete, and fully justified conclusions.

Assignments

Rationale: Assignments will give students an opportunity to practice solving theoretical problems related to longitudinal data analysis, and to apply statistical methods in analysing real datasets and interpreting the results of those analyses.

Assignments and projects 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

Schedule and weighting:

<table>
<thead>
<tr>
<th>Task</th>
<th>Date Avail.</th>
<th>Date Due</th>
<th>Form of Submission</th>
<th>Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asst 1</td>
<td>Week 3</td>
<td>5:00pm Mon Week 5</td>
<td>Written</td>
<td>10%</td>
</tr>
<tr>
<td>Asst 2</td>
<td>Week 7</td>
<td>5:00pm Mon Week 9</td>
<td>Written</td>
<td>10%</td>
</tr>
</tbody>
</table>
Assignments must be submitted by 5:00pm (at the start of lecture class). In general, late assignments will NOT be accepted without good documented reasons (such as illness or misadventure).

Project

Rationale: The project will develop student abilities in modelling a real and substantive set of longitudinal data and will ensure that students are fluent in applying the modelling and software skills learnt in the course so that they can reach appropriate conclusions and write a clear summary of their findings.

Duration: The final project is due at the start of lectures in Week 11 so that I can provide feedback and a final mark by Week 12 lecture. A brief verbal progress presentation to groups of class members will be required in week 8. Further details of the project requirements and the allocation of marks for the progress and final reports will be available in week 4 of the course.

Weighting: 20% of your final mark.

Examination

Duration: 2 hours.

Rationale: The final examination will assess student mastery of the material covered in the lectures.

Weighting: 60% of your final mark.

Further details about the final examination will be available in class closer to the time.

Additional resources and support

Lecture notes

Lecture notes (including computer lab exercises and practice problems) will be distributed via Blackboard by the end of the week preceding the lecture.

Textbooks

There are no set textbooks.
References

Students in the past have found the following references to be helpful and the course material is strongly linked to these, particularly the first listed:


Other references will be suggested in class.

Blackboard

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

Computer laboratories

Computer laboratories (RC-M020 and RC-G012) 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).

UNSW Library website:

http://info.library.unsw.edu.au/web/services/services.html

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. In response to the 2009 student evaluations of the course the mid-session test has been replaced by a substantial project and some elements of course content and structure have been modified.

Administrative matters

Important Information for Postgraduate Course Work Students

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/students/current/policies/studentpolicy.html.

UNSW Occupational Health and Safety policies and expectations:


Student equity and diversity issues

should be directed to the Student Equity Officers (Disability) in the Student Equity and Diversity Unit (9385-4734).

Students with Disabilities

Further information for students with disabilities is available at https://my.unsw.edu.au/student/atoz/

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 https://my.unsw.edu.au/student/academiclife/assessment/StudentMisconduct.html.
Course schedule

An approximate schedule for the course is given below.

<table>
<thead>
<tr>
<th>Date of class</th>
<th>Topic</th>
<th>Computer lab</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Characteristics of longitudinal data</td>
<td>Lab 1</td>
<td></td>
</tr>
<tr>
<td>4 March</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 2</td>
<td>Exploratory data analysis</td>
<td>Lab 2</td>
<td></td>
</tr>
<tr>
<td>11 March</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td>Summary statistics and other simple approaches</td>
<td>Lab 3</td>
<td>Assignment 1 available</td>
</tr>
<tr>
<td>18 March</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td>Correlation models</td>
<td>Lab 4</td>
<td>Project Details</td>
</tr>
<tr>
<td>25 March</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mid-session break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td>Linear models for correlated data</td>
<td>Lab 5</td>
<td>Assignment 1 due 5:00pm</td>
</tr>
<tr>
<td>8 April</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 6</td>
<td>Maximum likelihood estimation and REML</td>
<td>Lab 6</td>
<td></td>
</tr>
<tr>
<td>15 April</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 7</td>
<td>Linear mixed models</td>
<td>Lab 7</td>
<td>Assignment 2 available</td>
</tr>
<tr>
<td>22 April</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 8</td>
<td>Linear mixed models</td>
<td>Lab 8</td>
<td>Project Presentation</td>
</tr>
<tr>
<td>29 April</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 9</td>
<td>Generalized linear mixed models</td>
<td>Lab 9</td>
<td>Assignment 2 due 5:00pm</td>
</tr>
<tr>
<td>6 May</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 10</td>
<td>Generalized linear mixed models</td>
<td>Lab 10</td>
<td></td>
</tr>
<tr>
<td>13 May</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 11</td>
<td>Generalized LMMs Missing Data</td>
<td>Lab 11</td>
<td>Project Final Report</td>
</tr>
<tr>
<td>20 May</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 12</td>
<td>Missing Data, Review</td>
<td>Lab 12</td>
<td></td>
</tr>
<tr>
<td>27 May</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>