Faculty of Science
School of Mathematics and Statistics

MATH5826
STATISTICAL METHODS IN EPIDEMIOLOGY

Semester 2, 2014
MATH5826 – Course Outline

Information about the course

Course Authority:  Jake Olivier

Lecturer:  Jake Olivier
          Room RC-2051
          Phone 9385 6656
          email j.olivier@unsw.edu.au.

Consultation:  By appointment

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.

Lectures:  There will be one 3 hour lecture per week: Friday 5-8pm. Most classes will be in RC-4082, but some (to be advised) will be held in the computer labs.

Tutorials:  There are no separate tutorials.

UNSW Moodle:  Further information and course material will be provided via UNSW Moodle.

Course aims

The aim of this course is to introduce the main statistical concepts, methods and models used in the design and analysis of epidemiological studies.

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 familiarity 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, in dealing with modelling of relationships between variables. The focus here is on models which seek to explain the relationship between disease and exposure to risk factors, however. The course has links with MATH5906: Design and Analysis of Clinical Trials, which covers techniques applicable to experimental, rather than
observational studies. The course is closely related to MATH5945: Categorical Data Analysis, since the disease and exposure variables in epidemiological studies are often categorical rather than continuous in nature.

**Student Learning Outcomes**

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

- Explain the key features of epidemiological studies and what distinguishes them from other types of studies
- List the different types of epidemiological studies, explain their key features and the distinctions between them, and the advantages and disadvantages of each type of study
- Explain how the design features of the different types of studies dictate the methods of analysis appropriate to each
- Explain the different measures of disease occurrence and calculate these measures
- Explain the different measures of disease-exposure association, the differences between the measures and under what circumstances each is appropriate, and calculate these measures
- Assess the statistical significance of disease-exposure association by applying hypothesis tests and calculating confidence intervals for measures of disease-exposure association
- Explain the concepts of causality, confounding, and interaction
- Examine whether interaction and/or confounding is present in an epidemiological study
- Implement techniques to deal with interaction and/or confounding in an epidemiological study
- Explain the different types of regression models used in epidemiological studies
- Explain generalized linear models (GLMs), and indicate how logistic regression models fit into the GLM framework
- Use the theory of maximum likelihood estimation to derive estimators and standard errors for parameters in a logistic regression model
- Interpret the coefficients of a logistic regression model for different types of studies
• Check the validity of a logistic regression model using residuals and influence measures

• Explain the idea of matching in epidemiological studies and the reasons why it is used

• Derive and use statistical techniques appropriate for matched studies

• Solve theoretical problems related to epidemiological studies

• Use the computer language R to analyze data from epidemiological studies, and interpret R output.

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 epidemiological 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 mid-session test (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 epidemiological studies, 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 epidemiological 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 epidemiology, and to apply statistical methods for epidemiology in analysing real datasets and interpreting the results of those analyses.

Assignments must be YOUR OWN WORK, or severe penalties will be incurred.

You should consult the University web page on plagiarism

student.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>5pm Fri Week 5</td>
<td>Written</td>
<td>10%</td>
</tr>
<tr>
<td>Asst 2</td>
<td>Week 9</td>
<td>5pm Fri Week 11</td>
<td>Written</td>
<td>10%</td>
</tr>
</tbody>
</table>
Assignments must be submitted by 5pm (prior to class). In general, late assignments will NOT be accepted without good documented reasons (such as illness or misadventure).

Mid-session test

Duration and schedule: Take-home test, available 5pm Friday 5 September, due 9am Thursday 11 September (to be confirmed).

Rationale: The mid-session test will assess student mastery of the material covered up to that point in the lectures.

Weighting: 20% of your final mark.

Examination

Duration: Two 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 uploaded to UNSW Moodle.

Textbooks

There are no set textbooks.

References

The following reference is available in the library:


Other references will be suggested in class.
UNSW Moodle

All course materials will be available on UNSW Moodle. 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).

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

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 website at http://www.maths.unsw.edu.au/currentstudents/assessment-policies.

UNSW Occupational Health and Safety policies and expectations:
https://www.ohs.unsw.edu.au/hs_students/index.html

Student equity and diversity issues: should be directed to the Student Equity Officers in the Student Equity and Disabilities Unit: phone 9385-4734, website: http://http://www.studentequity.unsw.edu.au/

Students with Disabilities: Further information for students with disabilities is
available at
http://www.studentequity.unsw.edu.au/disability-services

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 student.unsw.edu.au/plagiarism and https://student.unsw.edu.au/conduct.
## Course schedule

An approximate schedule for the course is given below (subject to change).

<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>Definitions, background and history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 August</td>
<td>Study design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 2</td>
<td>Measures of disease occurrence: prevalence and incidence</td>
<td>Computer lab 1</td>
<td></td>
</tr>
<tr>
<td>8 August</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td>Measures of disease-exposure association: odds ratio, relative risk, excess risk</td>
<td></td>
<td>Assignment 1 distributed</td>
</tr>
<tr>
<td>15 August</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td>Assessing significance: $\chi^2$ test, c.i. for RR and OR</td>
<td>Computer lab 2</td>
<td></td>
</tr>
<tr>
<td>22 August</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td>Causality, confounding, and interaction</td>
<td></td>
<td>Assignment 1 due 5.00pm*</td>
</tr>
<tr>
<td>29 August</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 6</td>
<td>Stratification, Cochran-Mantel-Haenszel test, Woolf’s method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 September</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 7</td>
<td>Mid-session test</td>
<td>Take-home test due 9am Mon 3 Sep</td>
<td></td>
</tr>
<tr>
<td>28 August</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 8</td>
<td>Regression models for epidemiological studies: linear regression, generalized linear models</td>
<td>Computer lab 3</td>
<td></td>
</tr>
<tr>
<td>19 September</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 9</td>
<td>Logistic regression: MLEs, interpretation of coefficients for cohort and case-control studies</td>
<td></td>
<td>Assignment 2 distributed</td>
</tr>
<tr>
<td>26 September</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 10</td>
<td>Model checking for logistic regression: residuals and influence</td>
<td>Computer lab 4</td>
<td></td>
</tr>
<tr>
<td>10 October</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 11</td>
<td>Methods for matched studies</td>
<td>Assignment 2 due 5.00pm*</td>
<td></td>
</tr>
<tr>
<td>17 October</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 12</td>
<td>Poisson regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 October</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 13</td>
<td>No lecture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 October</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* BEFORE the lecture