# STAT2008

## Regression Modelling

### Course Description

STAT2008 is a course in applied statistics that studies the use of linear regression techniques for examining relationships between variables. The course emphasizes the principles of statistical modelling through the iterative process of fitting a model, examining the fit to assess imperfections in the model and suggest alternative models, and continuing until a satisfactory model is reached. Both steps in this process require the use of a computer: model fitting uses various numerical algorithms, and model assessment involves extensive use of graphical displays. The R statistical computing package is used as an integral part of the course.

<table>
<thead>
<tr>
<th>Semester and Year</th>
<th>Semester 1 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course URL</td>
<td><a href="https://wattle.anu.edu.au/">https://wattle.anu.edu.au/</a></td>
</tr>
<tr>
<td>Mode of Delivery</td>
<td>On campus, see <a href="http://timetable.anu.edu.au/">http://timetable.anu.edu.au/</a></td>
</tr>
<tr>
<td>Prerequisites</td>
<td>To enrol in this course you must have completed STAT1003 or STAT1008</td>
</tr>
<tr>
<td>Incompatible courses</td>
<td>No courses listed</td>
</tr>
<tr>
<td>Course Convener and Lecturer</td>
<td>Ian McDermid, Lecturer in Statistics Research School of Finance, Actuarial Studies and Applied Statistics (RSFAS) ANU College of Business and Economics (CBE) Room 3.08 (Level 3), CBE Building 26C Kingsley Street, Acton</td>
</tr>
<tr>
<td>Phone:</td>
<td>+61 2 612 51084 (just extension 51084 inside ANU)</td>
</tr>
<tr>
<td>Email:</td>
<td><a href="mailto:ian.mcdermid@anu.edu.au">ian.mcdermid@anu.edu.au</a></td>
</tr>
<tr>
<td>Consultation hours:</td>
<td>Wednesdays, 1pm to 3pm, or at other non-teaching times on Mondays to Thursdays by prior appointment, made and confirmed via e-mail.</td>
</tr>
<tr>
<td>Bio and research interests</td>
<td>30 years’ experience in statistical consulting, research and university teaching; current research interests in population health and mortality, sample survey analysis and design.</td>
</tr>
<tr>
<td>Tutors</td>
<td>A full list of contact details for course tutors and their consultation arrangements will be posted and regularly updated on the Wattle site for this course.</td>
</tr>
<tr>
<td>Student Administrator</td>
<td>Maria Lander, RSFAS Office, 9am–5pm Mondays to Fridays (except ACT Public Holidays) Room 4.48 (Level 4), CBE Building 26C Kingsley Street, Acton ACT 2601</td>
</tr>
<tr>
<td>Phone:</td>
<td>+61 2 612 53087 (just extension 53087 inside ANU)</td>
</tr>
<tr>
<td>Email:</td>
<td><a href="mailto:maria.lander@anu.edu.au">maria.lander@anu.edu.au</a> or <a href="mailto:enquiries.fas@anu.edu.au">enquiries.fas@anu.edu.au</a></td>
</tr>
</tbody>
</table>
COURSE OVERVIEW

Course Learning Outcomes
Upon successful completion of this course, students will be able to:

1. Demonstrate a working knowledge of the R statistical computing language, particularly the graphical capabilities.
2. Fit simple linear regression models and interpret model parameters.
3. Summarise and analyse relationships between a response variable and a covariate.
4. Summarise and analyse relationships between a response variable and several covariates.
5. Assess and refine simple and multiple linear regression models based on diagnostic measures. Identify outlying and influential data points.

Research-Led Teaching
My teaching in this introductory course in statistical modelling will draw on numerous examples from my extensive experience in applied statistical research and consulting.

Requisites
This is a course in applied statistics, using numerous examples, rather than a course in mathematical statistics; but it is NOT an introductory first course in basic statistics. We assume you have already completed a course such as STAT1003 or STAT1008.

Unfortunately, it will NOT be possible to waive the prerequisite for this course. It is strict Research School of Finance, Actuarial Studies and Applied Statistics (RSFAS) policy that pre-requisites for all courses be enforced. In RSFAS, course lecturers do not have the ability to waive pre-requisites or action enrolment variations.

The course uses the R statistical package, which uses matrix algebra to implement the regression modelling techniques. An understanding of matrix algebra (equivalent to an introductory mathematics course such as MATH1113) would be helpful in understanding how the R routines work, but such knowledge is not a required prerequisite nor an examinable part of this course.

Co-teaching
STAT2008 and the equivalent graduate course STAT6038 share the same lecture content and assignments, but will have separate tutorials and different final examinations. The two cohorts of students will also be treated separately in grading and any scaling that is applied.

Technology, Software, Equipment
The application of modern statistical techniques requires familiarity with a statistical computing package and the assignments for this course will require some data analysis on a computer.

This course makes extensive use of the R computing package, which is freely available to download at http://www.r-project.org. Further instructions on R, including a series of “getting started” workshops will be made available on the Wattle site for this course. R is also available on all InfoCommons computers on the ANU campus. All tutorials for this course will be held in one of the InfoCommons PC computer laboratories, though you may also find it helpful to also bring a laptop with R installed to the tutorials.

You will need also need access to a scientific calculator for the Final Examination.
**Student Feedback**

All CBE courses are evaluated using Student Experience of Learning and Teaching (SELT) surveys, administered by Planning and Statistical Services at the ANU. These surveys are offered online, and students will be notified via email to their ANU address when surveys are available in each course. Feedback is used for course development so please take the time to respond thoughtfully. Course feedback is anonymous and provides the Colleges, University Education Committee and Academic Board with opportunities to recognise excellent teaching and to improve courses across the university. For more information on student surveys at ANU and reports on feedback provided on ANU courses, visit [http://unistats.anu.edu.au/surveys/selt/students/](http://unistats.anu.edu.au/surveys/selt/students/) and [http://unistats.anu.edu.au/surveys/selt/results/learning/](http://unistats.anu.edu.au/surveys/selt/results/learning/)

**Continuous Improvement**

We use feedback from students, professional bodies and staff to make regular improvements to the course. In response to this feedback, design improvements from the previous version of the course include:

- Reduced emphasis on matrix algebra (see the section on Requisites above), as many students do not have the necessary mathematical background.
- Additional introductory materials on R (see Reading Lists below), as most students consider experience in R to be extremely valuable for future employment prospects, but some students found it difficult to get started.

**COURSE SCHEDULE**

<table>
<thead>
<tr>
<th>Week</th>
<th>Summary of Lecture Topics / Activities</th>
<th>Tutorials / Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (17 Feb)</td>
<td>Introduction. Getting started with R. Simple Linear Regression (revision).</td>
<td>No tutorials in week 1</td>
</tr>
<tr>
<td>3 (3 Mar)</td>
<td>Properties of least squares estimators.</td>
<td>R Worksheets 3 &amp; 4</td>
</tr>
<tr>
<td>4 (10 Mar)</td>
<td>ANOVA. Hypothesis testing and interval estimation in a SLR context.</td>
<td>Tutorial 1</td>
</tr>
<tr>
<td>5 (17 Mar)</td>
<td>Prediction intervals. Regression diagnostics (residual plots).</td>
<td>Tutorial 1 continued</td>
</tr>
<tr>
<td>6 (24 Mar)</td>
<td>Outliers and influential observations. Scale transformations.</td>
<td>Tutorial 2</td>
</tr>
<tr>
<td>7 (31 Mar)</td>
<td>Introduction to Multiple Regression. Model interpretation and estimation.</td>
<td>Tutorial 2 continued Assignment 1 due</td>
</tr>
<tr>
<td>8 (21 Apr)</td>
<td>ANOVA. Sequential sum of squares.</td>
<td>Tutorial 3 Assignment 1 solutions</td>
</tr>
<tr>
<td>9 (28 Apr)</td>
<td>Hypothesis testing, confidence intervals and prediction for multiple regression.</td>
<td>Tutorial 3 continued</td>
</tr>
<tr>
<td>10 (5 May)</td>
<td>Model diagnostics. Outlier detection. Types of residuals.</td>
<td>Tutorial 4</td>
</tr>
<tr>
<td>11 (12 May)</td>
<td>Influence diagnostics. Multicollinearity.</td>
<td>Tutorial 4 continued</td>
</tr>
<tr>
<td>12 (19 May)</td>
<td>Model selection and criteria for comparing models.</td>
<td>Tutorial 5 Assignment 2 due</td>
</tr>
<tr>
<td>13 (26 May)</td>
<td>Stepwise procedures for model selection. Revision for Final Examination.</td>
<td>Tutorial 5 continued Assignment 2 solutions</td>
</tr>
<tr>
<td>4 to 20 Jun</td>
<td>Examination period</td>
<td>Final Examination</td>
</tr>
</tbody>
</table>
Assessment Summary

<table>
<thead>
<tr>
<th>Task</th>
<th>Title</th>
<th>Value</th>
<th>Due Date</th>
<th>Linked Learning Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Assignment 1</td>
<td>20%</td>
<td>3pm, 2 April 2015</td>
<td>1, 2 &amp; 3</td>
</tr>
<tr>
<td>2</td>
<td>Assignment 2</td>
<td>20%</td>
<td>9am, 22 May 2015</td>
<td>1, 4, 5 &amp; 6</td>
</tr>
<tr>
<td>3</td>
<td>Final Examination</td>
<td>60%</td>
<td>Exam period</td>
<td>All</td>
</tr>
</tbody>
</table>

Assignment 1 (Simple Linear Regression) and Assignment 2 (Multiple Regression)
Detailed assignment specifications will be handed out at least three weeks prior to the due dates. Assignments will involve using R to analyse data from a case study, then organising and editing the R output and preparing a written report on your analyses.

Assignment Submission
Assignments reports should be submitted in hard copy to the appropriate box at the RSFAS School Office by the due date. Submitted assignments must include a completed version of the cover sheet provided on Wattle. Please keep a copy of the assignment for your records.

Extensions and Penalties
 Extensions will not generally be granted unless there is a compelling reason. If you apply for an extension a decision will be made within 5 working days of your application and you will be notified within 3 working days of the decision being made.

Late assignments will NOT be accepted. Late submission will result in a score of zero for that assessment task.

Returning Assignments and Resubmission
Assignments will be marked to a commonly agreed marking schedule by your tutor and returned to you in tutorials.

Assignment solutions will be discussed in the tutorials in the week following the due date or when marked assignments are returned and there will be no resubmission of assignments.

Referencing Requirements
All submitted assessment should be your own work, except where group work has been agreed and included in the detailed assignment specifications. All outside sources should be clearly referenced as per the ANU College of Business and Economics standards: [http://cbe.anu.edu.au/students/student-information/examinations-assessment/](http://cbe.anu.edu.au/students/student-information/examinations-assessment/).

Final Examination
Permitted materials and other conditions for the Final Examination will be discussed with students and the outcome advised on Wattle. The Final Examination will be centrally timetabled and the details released via [http://timetable.anu.edu.au/](http://timetable.anu.edu.au/).

Scaling
Your final mark for the course will be based on the raw marks allocated for each assignment or examination. However, your final mark may not be the same number as produced by that formula, as marks may be scaled. Any scaling applied will preserve the rank order of raw marks (i.e. if your raw mark exceeds that of another student, then your scaled mark will exceed or equal the scaled mark of that student), and may be either up or down.
SUPPORT SERVICES AND POLICIES

The University offers a number of support services for students. Information on these is available online from http://www.anu.edu.au/students.

ANU has educational policies, procedures and guidelines, which are designed to ensure that staff and students are aware of the University's academic standards, and implement them. You can find the University's education policies and an explanatory glossary at: http://policies.anu.edu.au/.

Students are expected to have read the Code of Practice for Student Academic Integrity before the commencement of their course.

Other key policies include:

- Student Assessment (Coursework)
- Student Surveys and Evaluations

READING LISTS

Class materials, including detailed lecture notes, power-point slides, instructions on how to access R, lecture demonstrations, tutorials, assignments and other relevant materials, will be made available on the class web page on Wattle (https://wattle.anu.edu.au), the University's online learning environment. It is essential that you visit the class web page regularly.

To log on to Wattle, you need to have your ANU ID (student number) and password. In order to access the class web page within Wattle, you will need to be formally enrolled in the course. The two courses (STAT2008 and STAT6038) share the same Wattle web page.

Prescribed Text

As we have a lot of detailed course material already available, there is NO prescribed text for this course. However, I recommend the following text for students who would like an even more detailed discussion of the course contents:


There are multiple copies of this text on 2 hour reserve in the ANU Hancock library (Call # QA279.F37) and the ANU Co-op Bookshop has copies available for purchase ($136.99 or $127.40 with the member’s discount, though there are cheaper new or second copies of the first edition readily available which would be acceptable).

For students who would like additional help getting started with R, I also recommend:


For students with a good mathematical background who would like further details on the matrix algebra implemented in R, I also recommend:

- Clarke, Brenton R. (2008) Linear Models: The Theory and Application of Analysis of Variance, Wiley. (Hancock QA279.C55 2008, Co-op Bookshop $162.55 or members $151.55; though you will probably have to ask the bookshop to order you a copy)

I will recommend other references during the course and possibly also make additional material available in the library e-Reserve. Use the following link to find course-related material in the ANU library: http://library.anu.edu.au/search/r?SEARCH=STAT2008
TUTORIAL REGISTRATION

Enrolment in tutorials will be completed online using the CBE Electronic Teaching Assistant (ETA). Enrolments will open shortly following the end of the first lecture on Tuesday 17 February 2015. To enrol, follow these instructions:

2. You will see the Student Login page. To log into the system, enter your University ID (your student number) and password (your ISIS password) in the appropriate fields and hit the Login button.
3. Read any news items or announcements.
4. Select "Sign Up!" from the left-hand navigation bar.
5. Select your courses from the list. To select multiple courses, hold down the control key. On PCs, this is the Ctrl key; on Macs, it is the key. Hold this key down while selecting courses with the mouse. Once courses are selected, hit the SUBMIT button.
6. A confirmation of class enrolments will be displayed. In addition, an email confirmation of class enrolments will be sent to your student account.
7. For security purposes, please ensure that you click the LOGOUT link on the confirmation page, or close the browser window when you have finished your selections.
8. If you experience any difficulties, please contact the RSFAS School Office (see page 1 for contact details).
9. Students will have until 5pm on Wednesday 25 February 2015 to finalise their enrolment in tutorials. After this time, you will be unable to change your tutorial enrolment.

COMMUNICATION

You will be given individual feedback by your tutor, who will mark your assignments. You are also welcome to ask questions of me or any of the class tutors at consultations or during classes. If you wish to ask me questions immediately following a lecture, please wait for me outside the lecture theatre, so that I can clean-up and log-off in preparation for the next class that will be using the same venue.

I am also happy to answer SHORT questions on the course material sent via email or posted on Wattle. If you send me a question via email, I will (unless you specifically ask me not to) post your question (anonymously) and my answer on Wattle for the benefit of all students.

Email

If I, or anyone in the School, College or University administration, need to contact you, we will do so via your official ANU student email address, which you need to check regularly. Information about your enrolment and fees from the Registrar and Student Services’ office will also be sent to this email address.

Announcements

Students are expected to check the Wattle site for announcements about this course, e.g. changes to timetables or notifications of cancellations. Notifications of emergency cancellations of lectures or tutorials will be posted on the door of the relevant room.

Other Course URLs

More information about this course may be found on:

- CBE website ([http://cbe.anu.edu.au/students/student-information/college-courses/](http://cbe.anu.edu.au/students/student-information/college-courses/)) and