

STAT6038

Regression Modelling

Semester and Year	SEMESTER 1 2017
Course URL	https://wattle.anu.edu.au/
Mode of Delivery	On campus, see http://timetable.anu.edu.au/ 3 one-hour lectures, as timetabled, and 1 one-hour tutorial starting in week 2 (see page 5 for tutorial enrolment instructions).
Prerequisites	To enrol in this course you must have completed STAT7055 (or STAT1003 or STAT1008) or be enrolled in the Master of Actuarial Studies/Practice or the Master of Statistics .
Incompatible Courses	Incompatible with STAT2008 (and STAT4038)
Course Convener and Lecturer:	Ian McDermid, Lecturer in Statistics Research School of Finance, Actuarial Studies and Statistics (RSFAS) ANU College of Business and Economics (CBE) Room 3.08 (Level 3), CBE Building 26C 2 Kingsley Street, Acton
Phone:	+61 2 612 51084 (just extension 51084 in ANU)
Email:	ian.mcdermid@anu.edu.au
Office hours for student consultation:	Wednesdays and Thursdays, 4pm to 5pm, or at other non-teaching times on Wednesdays to Fridays by prior appointment, made and confirmed via e-mail.
Research Interests	I have over 30 years of experience in statistical consulting, research and university teaching. My current research interests are in: population health and mortality; sample survey analysis and design.
Tutor(s)	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.
Student administrator:	Tracy Skinner, RSFAS Office, 9am–5pm Mondays to Fridays Room 4.48 (Level 4), CBE Building 26C 2 Kingsley Street, Acton ACT 2601
Phone:	+61 2 612 50487 (just extension 50487 in ANU)
Email:	tracy.skinner@anu.edu.au or enquiries.fas@anu.edu.au

COURSE OVERVIEW

Course Description

[STAT6038](#) 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.

Prerequisites

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 [STAT7055](#).

Unfortunately, it will NOT be possible to waive the prerequisite for this course. It is strict Research School of Finance, Actuarial Studies and 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

[STAT6038](#) and the equivalent undergraduate ([STAT2008](#)) and honours courses ([STAT4038](#)) share the same lecture content and assignments, but will have separate tutorials and different final examinations. The different cohorts of students will also be treated separately in grading and any scaling that is applied.

Learning Outcomes

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

1. Demonstrate a thorough understanding of the *R* statistical computing language, particularly the graphical capabilities.
2. Fit simple linear regression models, interpret model parameters and relate these back to the underlying research question.
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 and discuss the implications of outlying and influential data points.
6. Select and discuss a useful multiple regression model from a number of competing models.

Assessment Summary

Assessment Task	Value	Due Date	Linked Learning Outcomes
1. Wattle Quiz	5%	3pm, 31 March 2017	1, 2 & 3
2. Assignment 1	15%	3pm, 31 March 2017	1, 2 & 3
3. Assignment 2	20%	3pm, 19 May 2017	1, 4, 5 & 6
4. Final Examination	60%	Exam Period	All

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.

Feedback

Staff Feedback

You will be given individual feedback by your tutor, who will mark your assignments. Solutions to the assignments will be provided on Wattle and discussed in tutorials and/or lectures.

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.

Student Feedback

ANU is committed to the demonstration of educational excellence and regularly seeks feedback from students. One of the key formal ways students have to provide feedback is through Student Experience of Learning Support (SELS) surveys. The feedback given in these surveys is anonymous and provides the Colleges, University Education Committee and Academic Board with opportunities to recognise excellent teaching, and opportunities for improvement.

For more information on student surveys at ANU and reports on the feedback provided on ANU courses, go to

<http://unistats.anu.edu.au/surveys/selt/students/> and
<http://unistats.anu.edu.au/surveys/selt/results/learning/>

Communication

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/>) and
- ANU Programs and Courses (<http://programsandcourses.anu.edu.au/course/STAT6038>)

Support for Students

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

Resources

Class materials, including detailed lecture notes, 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. All of the related courses ([STAT2008](#), [STAT4038](#) and [STAT6038](#)) share the same Wattle web page.

Recommended Texts

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:

- Faraway, Julian J. (2015) *Linear Models with R*, 2nd Edn, CRC/Chapman & Hall.

There are multiple copies of this text on 2 hour reserve in the ANU Hancock library (Call # QA279.F37 2015) and the ANU Co-op Bookshop has copies available for purchase (\$146.00 or \$135.78 with the member's discount, though there are cheaper e-book versions or second-hand copies of the first edition readily available, which would be fine).

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

- Verzani, John (2014) *Using R for Introductory Statistics*, 2nd Edn, CRC/Chapman & Hall. (Hancock QA276.4.V47 2014, Co-op Bookshop \$104.00 or members \$96.72)

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 \$162.95 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=STAT6038>

Technology and Software

The application of modern statistical techniques requires familiarity with some 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.

Examination Equipment

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

COURSE SCHEDULE

Week	Summary of Lecture Topics / Activities	Tutorials / Assessment
1 (20 February)	Introduction. Getting started with <i>R</i> . Simple Linear Regression (revision). Parameter interpretation/estimation.	No tutorials in week 1
2 (27 February)	Matrix approach to linear regression. Properties of least squares estimators.	<i>Intro to R</i> Worksheet
3 (6 March)	ANOVA. Hypothesis testing and interval estimation in a SLR context.	Tutorial 1
4 (14 March)	Prediction intervals. Regression diagnostics (residual plots).	Tutorial 1 continued
5 (20 March)	Outliers and influential observations. Scale transformations.	Tutorial 2
6 (27 March)	Introduction to Multiple Regression. Model interpretation and estimation.	Tutorial 2 continued Quiz & Assignment 1 due
7 (18 April)	ANOVA for multiple regression. Sequential sum of squares.	Tutorial 3 Assignment 1 solutions
8 (24 April)	Hypothesis testing, confidence intervals and prediction for multiple regression.	Tutorial 3 continued
9 (1 May)	Model diagnostics. Outlier detection. Types of residuals.	Tutorial 4
10 (8 May)	Influence diagnostics. Multicollinearity.	Tutorial 4 continued
11 (15 May)	Model selection and criteria for comparing models.	Tutorial 5 Assignment 2 due
12 (22 May)	Stepwise procedures for model selection. Revision for Final Examination.	Tutorial 5 continued Assignment 2 solutions
(1 to 17 June)	Examination period	Final Examination

Tutorial/Seminar Registration

Tutorial signup for this course will be done via the Wattle site. Detailed information about signup times are provided on Wattle and will be available 15 minutes after the first lecture. When tutorials are available for enrolment, follow these steps:

1. Log on to Wattle, and go to the course site.
2. Click on the link "Tutorial enrolment".
3. On the right of the screen, click on the tab "Become Member of" for the tutorial class you wish to enter.
4. Confirm your choice.

If you need to change your enrolment, you will be able to do so by clicking on the tab "Leave group..." and then re-enrol in another group. You will not be able to enrol in groups that have reached their maximum number. Please note that enrolment in ISIS must be finalised for you to have access to Wattle.

ASSESSMENT REQUIREMENTS

Assessment Task	Value	Due Date	Estimated Return Date
1. Wattle Quiz	5%	3pm, 31 March 2017	On submission
2. Assignment 1	15%	3pm, 31 March 2017	Week 7, starting 18 April 2017
3. Assignment 2	20%	3pm, 19 May 2017	Week 12, starting 22 May 2017
3. Final Examination	60%	Exam Period	Results released 29 June 2017

Wattle Quiz

An (optional) short quiz will be made available on Wattle for you to complete in week 6, starting 3pm on Tuesday, 27 March 2017 and closing at 3pm on Friday, 31 March 2017. Marks for this quiz will be redeemable on Assignment 1; which will be worth 20%, if your marks for Assignment 1 are better than your results on the quiz.

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 and late submission of assessment pieces are covered by the Student Assessment (Coursework) Policy and Procedure.

The Course Convener may grant extensions for the assignments but not for examinations. If you need an extension, you must request it in writing on or before the due date. If you have documented and appropriate medical evidence that demonstrates you were not able to request an extension on or before the due date, you may be able to request it after the due date.

Extensions will generally NOT be granted unless there is a compelling reason. No extensions will granted after the solutions have been released/discussed and/or marked assignments have been returned to other students (the due date may NOT be extended after the estimated return date shown above).

No submission of assignments after the due date (without an extension) will be permitted. If an assignment is not submitted by the due date, a mark of 0 (zero) will be awarded.

Note this is an applied statistics course and the assignments represent an opportunity for you to show that you can correctly apply the statistical modelling techniques. As a result, the assignments are compulsory and assignment marks are NOT redeemable on the final examination.

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 first teaching week following the due date or when marked assignments are returned and there will be no resubmission of assignments.

Assessment Requirements continued

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/>.

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/>.

Scaling

Your final mark for the course will be based on the **raw** marks allocated for each of your assessment items. 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 the scaled mark of that student), and may be either up or down.

Policies

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 [Academic Misconduct Rules 2014](#) before the commencement of their course.

Other key policies include:

- [Student Assessment \(Coursework\)](#)
- [Student Surveys and Evaluations](#)

Privacy Notice

The ANU has made a number of third party, online, databases available for students to use. Use of each online database is conditional on student end users first agreeing to the database licensor's terms of service and/or privacy policy. Students should read these carefully.

In some cases student end users will be required to register an account with the database licensor and submit personal information, including their: first name; last name; ANU email address; and other information.

In cases where student end users are asked to submit 'content' to a database, such as an assignment or short answers, the database licensor may only use the student's 'content' in accordance with the terms of service – including any (copyright) licence the student grants to the database licensor.

Any personal information or content a student submits may be stored by the licensor, potentially offshore, and will be used to process the database service in accordance with the licensors terms of service and/or privacy policy.

If any student chooses not to agree to the database licensor's terms of service or privacy policy, the student will not be able to access and use the database. In these circumstances students should contact their lecturer to enquire about alternative arrangements that are available.