STAT6038
Regression Modelling

This 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>Mode of Delivery</th>
<th>On campus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prerequisites</td>
<td>To enrol in this course you must have completed STAT1003 or STAT1008 or be enrolled in the Bachelor of Applied Data Analytics.</td>
</tr>
<tr>
<td>Incompatible Courses</td>
<td>Incompatible with STAT2008 and STAT4038</td>
</tr>
<tr>
<td>Co-taught Courses</td>
<td>STAT2008 and Stat6038. Graduate students attend joint classes with undergraduates but are assessed separately.</td>
</tr>
<tr>
<td>Course Convener:</td>
<td>Dr. Anton Westveld</td>
</tr>
<tr>
<td>Phone:</td>
<td>02 6125 5122</td>
</tr>
<tr>
<td>Email:</td>
<td><a href="mailto:anton.westveld@anu.edu.au">anton.westveld@anu.edu.au</a></td>
</tr>
<tr>
<td>Office hours for student consultation:</td>
<td>TBD</td>
</tr>
<tr>
<td>Research Interests</td>
<td>Research interests include Bayesian methodology and theory, statistical methods for interaction/relational data (network, game theoretic), statistical applications in social (economics, political science, public policy), environmental, and biological sciences.</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>Administrator</td>
<td>Ms. Tracy Skinner</td>
</tr>
<tr>
<td>Phone:</td>
<td>+61250487</td>
</tr>
<tr>
<td>Email:</td>
<td><a href="mailto:tracy.skinner@anu.edu.au">tracy.skinner@anu.edu.au</a></td>
</tr>
</tbody>
</table>

SEMESTER 1
2018
http://programsandcourses.anu.edu.au/course/STAT6038
COURSE OVERVIEW

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 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 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.

Co-teaching
STAT2008 shares the same lecture content and assignments with STAT4038 & STAT6038, however these cohorts may 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 the requirements of this course, students should have the knowledge and skills 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

<table>
<thead>
<tr>
<th>Assessment Task</th>
<th>Value</th>
<th>Due Date</th>
<th>Linked Learning Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wattle quiz</td>
<td>5%</td>
<td>March 29</td>
<td>1, 2, &amp; 3</td>
</tr>
<tr>
<td>2. Assignment 1</td>
<td>15%</td>
<td>March 29</td>
<td>1, 2, &amp; 3</td>
</tr>
<tr>
<td>3. Assignment 2</td>
<td>20%</td>
<td>May 18</td>
<td>1, 4, 5, &amp; 6</td>
</tr>
<tr>
<td>4. Final exam</td>
<td>60%</td>
<td>Exam period</td>
<td>All</td>
</tr>
</tbody>
</table>
Research-Led Teaching

My teaching in this introductory course in statistical modelling will draw on examples from my experience in 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. Additionally, general verbal comments will be provided to the whole class.

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.

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.

Announcements
Students are expected to check the Wattle site for announcements about this course, e.g. changes to timetables or notifications of cancellations.

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 Rule before the commencement of their course.
Other key policies include:

- Student Assessment (Coursework)
- Student Surveys and Evaluations

**Resources**

**Recommended Texts**
There is no prescribed text, however I may draw a fair amount of material from:


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 (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:


**Technology and Software**
The application of modern statistical techniques requires familiarity with a statistical computing package. Examples provided in lectures, tutorials, and work related to the assignments will entail the use of the statistical computer packages R and RStudio, which are freely available at [www.r-project.org](http://www.r-project.org) and [https://www.rstudio.com](https://www.rstudio.com). The program code used for examples provided in lectures and tutorials will be available on the course Wattle site. Note: students will not be able to use any statistical package during the exam.

**Examination material or equipment**
You will also need access to a scientific calculator for the Final Examination.
COURSE SCHEDULE - Rough Schedule

<table>
<thead>
<tr>
<th>Week/Session</th>
<th>Summary of Activities</th>
<th>Tutorials / Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Orientation Week</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Introduction. Getting started with R. Simple Linear Regression (revision). Parameter interpretation/estimation.</td>
<td>No tutorials in week 1</td>
</tr>
<tr>
<td>2</td>
<td>Matrix approach to linear regression. Properties of least squares estimators.</td>
<td>Intro to R Worksheet</td>
</tr>
<tr>
<td>3</td>
<td>ANOVA. Hypothesis testing and interval estimation in a SLR context.</td>
<td>Tutorial 1</td>
</tr>
<tr>
<td>4</td>
<td>Prediction intervals. Regression diagnostics (residual plots).</td>
<td>Tutorial 1 continued</td>
</tr>
<tr>
<td>5</td>
<td>Outliers and influential observations. Scale transformations.</td>
<td>Tutorial 2</td>
</tr>
<tr>
<td>6</td>
<td>Introduction to Multiple Regression. Model interpretation and estimation.</td>
<td>Tutorial 2 continued; Wattle quiz &amp; Assignment 1</td>
</tr>
<tr>
<td></td>
<td><strong>Two-Week Teaching Break</strong></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Model interpretation continued (discussion of causality)</td>
<td>Tutorial 3</td>
</tr>
<tr>
<td>8</td>
<td>ANOVA for multiple regression. Sequential sum of squares.</td>
<td>Tutorial 3 continued</td>
</tr>
<tr>
<td>9</td>
<td>Hypothesis testing, confidence intervals and prediction for multiple regression.</td>
<td>Tutorial 4</td>
</tr>
<tr>
<td>11</td>
<td>Model selection and criteria for comparing models.</td>
<td>Tutorial 5; Assignment 2</td>
</tr>
<tr>
<td>12</td>
<td>Revision for Final Examination.</td>
<td>Tutorial 5 continued</td>
</tr>
<tr>
<td></td>
<td><strong>Examination Period</strong></td>
<td>Final exam</td>
</tr>
</tbody>
</table>

ASSESSMENT REQUIREMENTS

The ANU is using Turnitin to enhance student citation and referencing techniques, and to assess assignment submissions as a component of the University’s approach to managing Academic Integrity. For additional information regarding Turnitin please visit the ANU Online website.

Any student identified, either during the current semester or in retrospect, as having used ghost writing services will be investigated under the University’s Academic Misconduct Rule.
Wattle Quiz
A short quiz will be made available on Wattle for you to complete in week 6.

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 are compulsory and 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.

Examination There is a compulsory final exam. 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/.

Assignment submission
Online Submission: Assignments are submitted using Turnitin in the course Wattle site. You will be required to electronically sign a declaration as part of the submission of your assignment. 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 assessment pieces that are not examinations or take-home 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.

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
Assignment will be returned online.

Resubmission of assignments
Assignments may not be resubmitted.
Referencing requirements

Appropriate referencing will be necessary for the assignments. For more information see: http://www.anu.edu.au/students/learning--development/academic--integrity/how--referencing--works

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.

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 licensors 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.

Tutorial Seminar Registration

Tutorial signup for this course will be done via the Wattle website. Detailed information about signup times will be provided on Wattle or during your 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.

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/