# STAT4040
## Statistical Learning

This course provides an introduction to statistical learning and aims to develop skills in modern statistical data analysis. There has been a prevalence of “big data” in many different areas such as finance, marketing, social networks and the scientific fields. As traditional statistical methods have become inadequate for analysing data of such size and complexity, this has led to the development of new statistical methods for extracting information, or “learning”, from such data. This course will cover a range of topics in statistical learning including linear regression, classification techniques, resampling methods such as the bootstrap, regularisation methods, tree based methods and unsupervised learning techniques such as clustering. As much modern data analysis requires the use of statistical software, there will be a strong computing component in this course.

<table>
<thead>
<tr>
<th>Mode of Delivery</th>
<th>On campus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prerequisites</td>
<td>As listed in Programs and Courses</td>
</tr>
<tr>
<td>Incompatible Courses</td>
<td>As listed in Programs and Courses</td>
</tr>
<tr>
<td>Co-taught Courses</td>
<td>STAT3040, Stat7040. 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>Relevant administrator</td>
<td>Ms. Maria Lander</td>
</tr>
<tr>
<td>Email:</td>
<td><a href="mailto:maria.lander@anu.edu.au">maria.lander@anu.edu.au</a></td>
</tr>
<tr>
<td>Tutor</td>
<td>Mr. Daning Bi</td>
</tr>
<tr>
<td>Email:</td>
<td><a href="mailto:daning.bi@anu.edu.au">daning.bi@anu.edu.au</a></td>
</tr>
<tr>
<td>Office hours for student consultation:</td>
<td>TBD</td>
</tr>
</tbody>
</table>

**SEMESTER**

2017

COURSE OVERVIEW

Learning Outcomes

Upon successful completion of the requirements of this course, students should have the knowledge and skills to:

- Discuss in detail the rationale behind the formulation and components of a statistical model.
- Analytically describe and implement approaches to compare and contrast statistical models in the context of a particular scientific question.
- Communicate complex statistical ideas and heuristics to a diverse audience.
- Develop, analytically describe, and implement a statistical solution to real-data research problems.
- Demonstrate an in-depth level interpretation of modeling results.
- Discuss in detail the theoretical and computational underpinnings of various statistical procedures, including common classes of statistical models.
- Demonstrate computational skills to implement various statistical procedures.

Assessment Summary

<table>
<thead>
<tr>
<th>Assessment Task</th>
<th>Value</th>
<th>Due Date</th>
<th>Date for Return of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Online quiz</td>
<td>0% of the total mark</td>
<td>Week 6</td>
<td>End of Week 6</td>
</tr>
<tr>
<td>2. Assignment</td>
<td>20% of the total mark</td>
<td>Submitted online</td>
<td>Two weeks after the due date</td>
</tr>
<tr>
<td></td>
<td></td>
<td>at the end of Week 8</td>
<td></td>
</tr>
<tr>
<td>3. Project</td>
<td>30% of the total mark</td>
<td>Submitted online</td>
<td>Two weeks after the due date</td>
</tr>
<tr>
<td></td>
<td></td>
<td>at the end of Week 12</td>
<td></td>
</tr>
<tr>
<td>4. Final exam</td>
<td>50% of the total mark</td>
<td>Exam period</td>
<td></td>
</tr>
</tbody>
</table>

Research-Led Teaching

An important component of this course is a research project, which will allow students to think creatively about potential solutions to data analytic problems.

Feedback

Staff Feedback

Students will be given feedback in the following forms in this course:
• Written comments, both individually as well as to the whole class.
• Verbal comments to the whole class.

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/

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

Required Resources
Prescribed Texts
1. G. James, D. Witten, T. Hastie, and R. Tibshirani
   An Introduction to Statistical Learning with Applications in R
   Springer
   This text is freely available at http://www-bcf.usc.edu/~gareth/ISL/

2. T. Hastie, R. Tibshirani, and J. Friedman
   The Elements of Statistical Learning: Data Mining, Inference, and Prediction (second edition)
   Springer
   This text is freely available at http://statweb.stanford.edu/~tibs/ElemStatLearn/

3. N. Silver
   The Signal and the Noise: The Art and Science of Prediction
   Allen Lane Publishing
   I have requested that the ANU Bookstore stock this text.
Examples provided in lectures, tutorials, and work related to the assignments and project will entail the use of the statistical computer packages R and RStudio, which are freely available at www.r-project.org and 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.

**Recommended Texts**

1. W. Piegorsch  
   *Statistical Data Analytics: Foundation for Data Mining, Informatics, and Knowledge Discovery*  
   Wiley

2. W. Piegorsch  
   *Statistical Data Analytics: Solution Manual*  
   Wiley

3. J. Weatherwax  
   *A Solution Manual and Notes for: An Introduction to Statistical Learning: with Applications in R: Machine Learning*  
   Amazon Digital Services LLC

**Examination material or equipment**

The permitted material for the final exam will be:

- Two sheets of A4 paper with notes on both sides
- Paper-based dictionary, no approval required (must be clear of ALL annotations)
- Calculator (Any - programmable or not)
COURSE SCHEDULE - Rough Schedule

<table>
<thead>
<tr>
<th>Week/Session</th>
<th>Summary of Activities</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Orientation Week</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Overview of Statistical Learning</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Linear Regression</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Generalised Linear Models</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Classification</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Resampling Methods</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Linear Model Selection and Regularisation</td>
<td>Quiz</td>
</tr>
<tr>
<td>7</td>
<td>Linear Model Selection and Regularisation (continued)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Two-Week Teaching Break</strong></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Moving Beyond Linearity</td>
<td>Assignment</td>
</tr>
<tr>
<td>9</td>
<td>Tree-Based Methods</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Support Vector Machines</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Unsupervised Learning</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Unsupervised Learning</td>
<td>Project</td>
</tr>
<tr>
<td></td>
<td>Examination Period</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.

Students may choose not to submit assessment items through Turnitin. In this instance you will be required to submit, alongside the assessment item itself, copies of all references included in the assessment item.

Assessment Tasks

**Assessment Task 1:** Quiz  
**Details of task:** There will be a compulsory online quiz. Results will be made available at the end of Week 6.  
**Value:** 0% of the total mark.

**Assessment Task 2:** Assignment  
**Details of task:** There will be a compulsory take-home problem set. Students should attempt all of the questions, showing appropriate mathematical and computational details, as well as discuss the results. Students should write-up the results using R-markdown within RStudio.  
**Value:** 20%
Assessment Task 3: Group Project
Details of task: In groups of 1-4 (based on the number of students in the course), you will take part in a compulsory data prediction competition, in conjunction with broader data analyses. Students are required to communicate their findings in a formal written report (using R-markdown within RStudio).
Value: 30%

Assessment Task 4: Final Exam
Details of task: There will be a compulsory final exam covering the material in the course.
Writing time: 3 hours
Value: 50%

Examination(s)
There is a compulsory final exam.

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.

No submission of assessment tasks without an extension after the due date will be permitted. If an assessment task is not submitted by the due date, a mark of 0 will be awarded.

Returning assignments
Assignment will be returned online.

Resubmission of assignments
Assignments may not be resubmitted.
Referencing requirements

Appropriate referencing will be necessary for the presentations. 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/