# MONASH University
## Information Technology

**Unit Guide**

# FIT5197
# Statistical data modelling

# Semester 1, 2020

*We acknowledge and pay respects to the Traditional Owners and Elders - past, present and emerging - of the lands and waters on which Monash University operates.*

The information contained in this unit guide is correct at time of publication. The University has the right to change any of the elements contained in this document at any time.

*Last updated: 24 Feb 2020*

*Status: Approved*

# Table of contents

# Unit handbook information

## Synopsis

This unit explores the statistical modelling foundations that underlie the analytic aspects of Data Science. Motivated by case studies and working through examples, this unit covers the mathematical and statistical basis with an emphasis on using the techniques in practice. It introduces data collection, sampling and quality. It considers analytic tasks such as statistical hypothesis testing and exploratory versus confirmatory analysis. It presents basic probability distributions, random number generation and simulation as well as estimation methods and effects such as maximum likelihood estimators, Monte Carlo estimators, Bayes theorem, bias versus variance and cross validation. Basic information theory and dependence models such as regression and log-linear models are also presented, as well as the role of general modelling such as inference and decision making, and predictive models.

## Location(s) and mode(s) of delivery

Suzhou (On-campus)
Clayton (On-campus)

## Workload requirements

Minimum total expected workload equals 144 hours per semester comprising:

(a) Contact hours for on-campus students:

- Two hours/week lectures
- Two hours/week laboratories

(b) Contact hours for Monash Online students:

- Two hours/week online group sessions.
- Online students generally do not attend lecture, tutorial and laboratory sessions, however should plan to spend equivalent time working through resources and participating in discussions.

(c) Additional requirements (all students):

- A minimum of 8 hours per week of personal study (22 hours per week for Monash online students) for completing lab/tutorial activities, assignments, private study and revision, and for online students, participating in discussions.

See also Unit timetable information

# Class Timetable

Follow this link to view the scheduled times for your unit's learning activities for this teaching period.

# Unit relationships

## Prerequisites

MAT9004 and ( FIT9133 or FIT9136 or FIT9131 )

## Prohibitions

None

## Co-requisites

None

## Enrolment rules

<p>Prerequisites:</p> <p>For students enrolled in C6007: None. </p>

# Staff details

## Chief Examiner

Dr Levin Kuhlmann

## Campus Lecturer(s)

## Clayton

**Name:** Dr Levin Kuhlmann
**Email:** Levin.Kuhlmann@monash.edu
**Consultation hours:** Will be advised via Moodle


**Name:** Mr Dan Nguyen
**Email:** Dan.Nguyen2@monash.edu
**Consultation hours:** Will be advised via Moodle


## Suzhou

**Name:** Assoc Professor Vincent Lee
**Email:** Vincent.CS.Lee@monash.edu
**Consultation hours:** Will be advised via Moodle

# Academic overview

## Learning outcomes

On successful completion of this unit, you should be able to:

1. Perform exploratory data analysis with descriptive statistics on given datasets;

2. Construct models for inferential statistical analysis;

3. Produce models for predictive statistical analysis;

4. Perform fundamental random sampling, simulation and hypothesis testing for required scenarios;

5. Implement a model for data analysis through programming and scripting;

6. Interpret results for a variety of models.

## Teaching approach

The teaching approach is lectures with a reduced element of "flipped" classroom. Students will be introduced to topics via online ePub and video material. Lectures are where concepts are demonstrated interactively with the class. Students will gain practical experience via tutorial classes with activities and tutor-led discussion.

## Live streaming

If you have been allocated to a lecture activity as a livestream, you will participate in this activity online. A link to the livestreams will be available in your Moodle unit.

- For information on how to participate in your lectures via live streaming, you can review this guide created by the library: https://guides.lib.monash.edu/learning-tools/video

- If you have any technical issues please contact the service desk: https://www.monash.edu/esolutions/contact
- If you need more information on timetabling you can visit their site: https://www.monash.edu/timetables/fix-problems

# Assessment summary

Monash Online: In-semester assessment: 100%

On-campus: Examination (2 hours and 10 minutes): 50%; In-semester assessment: 50%

| Assessment task | Value | Due date |
| --- | --- | --- |
| Assignment 1a, 1b, 1c, 1d, 1e, 1f | 6 x 2.5 % per assignment = 15% | Week 4 (Ass. 1a), 5 (Ass. 1b), 7 (Ass. 1c), 8 (Ass. 1d), 9 (Ass. 1e), 12 (Ass.1f) |
| Assignment 2 | 20% | Week 6 |
| Assignment 3 | 15% | Week 10 |
| Examination | 50% | To be advised |

# Unit schedule

For units with on-campus classes, teaching activities are normally scheduled to start on the hour (teaching will commence on the hour and conclude 10 minutes prior to the scheduled end time).

| Week | Activities | Assessment |
|---|---|---|
| 0 | | No formal assessment or activities are undertaken in week 0 |
| 1 | Introduction to Modelling for Data Science and to R | Students should do the Academic Integrity Awareness Training Tutorial Activity. |
| 2 | Probabilities and Bias | |
| 3 | Expectations | |
| 4 | Distributions | Assignment 1a |
| 5 | Statistical Inference | Assignment 1b |
| 6 | Hypothesis Testing | Assignment 2 |
| 7 | Linear Regression | Assignment 1c |
| 8 | Classification | Assignment 1d |
| 9 | Simulation | Assignment 1e |
| 10 | Clustering and Validation | Assignment 3 |
| 11 | Further Modelling | |
| 12 | Revision | Assignment 1f |
| | SWOT VAC | No formal assessment is undertaken during SWOT VAC |
| | Examination period | LINK to Assessment Policy:http://policy.monash.edu.au/policy-bank/academic/education/assessment/assessment-in-coursework-policy.html |

*Unit Schedule details will be maintained and communicated to you via your learning system.

# Assessment requirements

# Faculty Unit Assessment Hurdles

*To pass a unit which includes an examination as part of the assessment*, a student must obtain, unless otherwise approved and published:

- 45% or more in the unit's examination, and
- 45% or more in the unit's total non-examination assessment, and
- an overall unit mark of 50% or more.

If a student does not achieve 45% or more in the unit examination or the unit non-examination total assessment, and the total mark for the unit is:

- equal to or greater than 50%, then a mark of 45-NH will be recorded for the unit.
- less than 50% then the actual mark for the unit will be recorded.

*To pass a unit with 100% in-semester assessment*, a student must obtain, unless otherwise approved and published:

- an overall unit mark equal to or greater than 50%.

# Assessment tasks

**Assessment title: Assignment 1a, 1b, 1c, 1d, 1e, 1f**
**Learning outcomes:** 2, 3, 4, 6
**Type:** Individual assessment task
**Details of task:** Answer questions by providing short answers about the following:

Assignment 1a: Descriptive Statistics & Probability

Assignment 1b: Expectation & Entropy

Assignment 1c: Maximum Likelihood, Distributions & Inference

Assignment 1d: Hypothesis testing & Confidence Intervals

Assignment 1e: Regression & Classification

Assignment 1f: Simulation & Modelling

**Value:** 6 x 2.5 % per assignment = 15%
**Criteria for marking:**

- Correctness of result;
- program commentary/explanation,
- demonstration of conceptual/mathematical understanding.

You may be subject to an interview as part of this assessment which will be used to assess the quality of your understanding and may influence your grade.

**Due date:** Week 4 (Ass. 1a), 5 (Ass. 1b), 7 (Ass. 1c), 8 (Ass. 1d), 9 (Ass. 1e), 12 (Ass.1f)
**Estimated return date:** within one week

---

**Assessment title: Assignment 2**
**Learning outcomes:** 1,2,5,6
**Type:** Individual assessment task
**Details of task:** Students are required to answer questions about Bayes theorem, probabilities, expectations, properties of distributions and statistical inference. These will also involve writing and submitting a short computer program that evaluates some of the answers provided. Further, they will be required to write and submit a program that visualizes sample data and their relation to probability models.
**Value:** 20%
**Criteria for marking:**

- Correctness of result;
- program commentary/explanation,
- demonstration of conceptual/mathematical understanding.

You may be subject to an interview as part of this assessment which will be used to assess the quality of your understanding and may influence your grade.

**Due date:** Week 6
**Estimated return date:** within two weeks

---

**Assessment title: Assignment 3**
**Learning outcomes:** 3,4,5,6
**Type:** Individual assessment task
**Details of task:** The assignment is an individual assignment.  Students are required to write and submit a programming exercise of comparable difficulty to producing the following three tasks:

(1) a regression model for a multivariate data set and computing one or more relevant coefficients

(2) testing a null hypothesis and/versus an alternative on a data set, and performing a confidence interval estimation

(3) random number generation

Supporting explanations and visualisations may be required.

**Value:** 15%
**Criteria for marking:**

- Correctness of result;
- program commentary/explanation,
- demonstration of conceptual/mathematical understanding.

You may be subject to an interview as part of this assessment which will be used to assess the quality of your understanding and may influence your grade.

**Due date:** Week 10
**Estimated return date:** within two weeks

# Examination

*This unit may employ electronic assessment for the final exam. Further details will be provided to you by Week 4 of Semester.*

**Value:** 50%
**Length:** 2 Hours and 10 minutes
**Type:** closed-book
**Electronic devices allowed:** Nonprogrammable calculators are allowed.
**Learning outcomes assessed:** 2, 3, 4, 6

**Additional information:** Rulers for drawing graphs are also allowed in the exam

## Extensions and penalties

Submission must be made by the due date otherwise penalties will be enforced.

You must negotiate any extensions formally with your campus unit lecturer via the in-semester special consideration process: http://www.monash.edu.au/exams/special-consideration.html

If a completed documented application for special consideration (see, e.g., https://www.monash.edu/exams/changes/special-consideration [under in-semester]) for extraordinary circumstances (including documented illness) is completed and submitted in a timely manner within the deadlines specified by university policy, then an extension might be granted. Otherwise, point penalties are applied up to 5% per day of late submissions (including weekends, public holidays, etc.) or as subsequently specified. Submissions 10 or more days after the deadline are not acceptable, and a mark of 0 might be returned. Your mark can not become lower than 0.

Please also understand that if work is submitted late then it can delay our ability to return your work.

# Returning assignments

Students can expect assignments to be returned within two weeks of the submission date or after receipt, whichever is later.

# Resubmission of assignments

Re-submission allowed only when the lecturer specifically requests it.

# Referencing requirements

To build your skills in citing and referencing, and using different referencing styles, see the online tutorial Academic Integrity: Demystifying Citing and Referencing at http://www.lib.monash.edu/tutorials/citing/

**Additional information:**
Students should use APA reference style, as explained at http://intranet.monash.edu.au/infotech /resources/students/style-guide/referencing.html

# Assignment submission

It is a University requirement (http://www.policy.monash.edu/policy-bank/academic/education /conduct/student-academic-integrity-managing-plagiarism-collusion-procedures.html) for students to submit an assignment coversheet for each assessment item. Faculty Assignment coversheets can be found at https://www.monash.edu/it/current-students/resources-and-support/student-forms. Please check with your Lecturer on the submission method for your assignment coversheet (e.g. attach a file to the online assignment submission, hand-in a hard copy, or use an electronic submission).

**Please note:**

1. It is your responsibility to retain copies of your assessments.
2. Assessments submitted without an assignment coversheet will not be marked.


**Online submission:** If Electronic Submission has been approved for your unit, please submit your work via the learning system for this unit, which you can access via links in the my.monash portal.

Assessment work should be submitted via Moodle.  Students should at all times be mindful of policies on Academic Integrity.

Please keep a copy of tasks completed for your records.


# Feedback to you

Graded assignments without comments
Solutions to tutes, labs and assignments
Examination feedback after results publication
Informal feedback on progress in labs/tutes

# Learning resources

Monash Library Unit Reading List (if applicable to the unit): http://monash.rl.talis.com/index.html
Research and Learning Online: www.monash.edu/rlo

## Required resources

Students generally must be able to complete the requirements of their course without the imposition of fees that are additional to the student contribution amount or tuition fees. However, students may be charged certain incidental fees or be expected to make certain purchases to support their study. For more information about this, refer to the Higher Education Administrative Information for Providers, Chapter 18, Incidental Fees at http://education.gov.au/help-resources-providers.

Please check with your lecturer before purchasing any required resources. Limited copies of prescribed texts are available for you to borrow in the library, and prescribed software is available in student labs.

R, RStudio or Jupyter Notebook capable of running R.

**Prescribed textbook:**  We do not follow this book closely, and only cover some of it.

S. M. Ross (2014). *Introduction to Probability and Statistics for Engineers and Scientists, 5th ed.* (5th) Academic.

A number of these books are available in the Library

## Technological requirements

Students may use Windows, Linux or Mac environments for this subject. R and RStudio or Jupyter Notebook capable of running R must be used for programming assignments.  Non-programmable calculators can also be used for tutorials and exams. Rulers for drawing graphs are also allowed in the exam.

## Recommended resources

Agresti, Franklin and Klingenberg (2018, 4th edition). Statistics: the art and science of learning from data, Pearson.
G. James, D. Witten, T. Hastie and R. Tibshirani, ``An Introduction to Statistical Learning with Applications in R'' (http://www-bcf.usc.edu/~gareth/ISL).

# Bring your own device

*Please note:* This is a bring your own device unit. You will be expected to bring a web-connected device to class to access software. You can access specialised software on the [MoVE website](#) via [move.monash.edu.](#) For more information, visit [monash.edu/move.](#)

# Previous student evaluations of unit

One of the formal ways students have to provide feedback on teaching and their learning experience is through the Student Evaluation of Teaching and Units (SETU) survey. The feedback is anonymous and provides the Faculty with evidence of aspects that students are satisfied with and areas for improvement.

**Previous student evaluations of this unit**
In response to previous SETU results of this unit, the following changes have been made:

- Slides reorganised with some more examples.
- Tutorials will have more basic material to cover with the tutor doing worked examples.
- Assignment 1 consists of (almost) weekly quizzes to provide an opportunity to complete questions in a similar way to that done in the exam.

If you wish to view how previous students rated this unit, please go to https://www.monash.edu/ups/setu/about/setu-results/unit-evaluation-reports.

If you would like to know more about SETU, please go to www.monash.edu/ups/setu.

# Other information

## Policies

Monash has educational policies, procedures and guidelines, which are designed to ensure that staff and students are aware of the University's academic standards, and to provide advice on how they might uphold them. You can find Monash's Education Policies at:
http://www.policy.monash.edu/policy-bank/academic/education/index.html

## Student Academic Integrity Policy

https://www.monash.edu/__data/assets/pdf_file/0008/801845/Student-Academic-Integrity_Managing-Plagiarism-and-Collusion-Procedures.pdf

## Special Consideration

For information on applying for special consideration, please visit: http://www.monash.edu/exams/changes/special-consideration

## Graduate Attributes Policy

http://www.monash.edu/__data/assets/pdf_file/0009/786969/Course-Design-Policy.pdf

## Student Charter

http://www.monash.edu/students/policies/student-charter.html

## Student Services

The University provides many different kinds of services to help you gain the most from your studies. Further information is available at
http://www.monash.edu/students.

For Malaysia see http://www.monash.edu.my/Student-services, and for South Africa see http://www.monash.ac.za/current/.

## Monash University Library

The Monash University Library provides a range of services, resources and programs that enable you to save time and be more effective in your learning and research.
Go to http://www.monash.edu/library or the library tab in my.monash portal for more information.

At Malaysia visit the Library and Learning Commons at http://www.lib.monash.edu.my/.

At South Africa visit http://www.lib.monash.ac.za/.

## Disability Support Services

Students who have a disability, ongoing medical or mental health condition are welcome to contact Disability Support Services.

Disability Support Services also support students who are carers of a person who is aged and frail or has a disability, medical condition or mental health condition.

Disability Advisers visit all Victorian campuses on a regular basis.

- Website: monash.edu/disability
- Telephone: 03 9905 5704 to book an appointment with an Adviser, or contact the Student Advisor, Student Community Services at 03 55146018 at Malaysia
- Email: disabilitysupportservices@monash.edu
- Drop In: Level 1, Western Annexe, 21 Chancellors Walk (Campus Centre) Clayton Campus, or Student Community Services Department, Level 2, Building 2, Monash University, Malaysia Campus

# Additional Notes:

Students should make sure to have the prerequisites - especially the mathematical prerequisites - for this subject.

Students should at all times be mindful of policies on Academic Integrity.