DATA1001: Foundations of Data Science

Semester 2, 2019 | 6 Credit points | Mode of delivery: Normal (lecture/lab/tutorial) day | Unit type: Standard

Coordinator: Garth Tarr

Faculty of Science (SCIE), Mathematics and Statistics Academic Operations

Unit description

DATA1001 is a foundational unit in the Data Science major. The unit focuses on developing critical and statistical thinking skills for all students. Does mobile phone usage increase the incidence of brain tumours? What is the public's attitude to shark baiting following a fatal attack? Statistics is the science of decision making, essential in every industry and undergirds all research which relies on data. Students will use problems and data from the physical, health, life and social sciences to develop adaptive problem solving skills in a team setting. Taught interactively with embedded technology, DATA1001 develops critical thinking and skills to problem-solve with data. It is the prerequisite for DATA2002.

Prohibitions: DATA1901 or MATH1005 or MATH1905 or MATH1015 or MATH1115 or ENVX1001 or ENVX1002 or ECMT1010 or BUSS1020 or STAT1021 or STAT1022

Pre-requisites : None

Co-requisites : None

Assumed knowledge: None

Unit aims

The aim of DATA1001 is for you to become a “statistical story-teller”. DATA1001 will develop your ability to problem solve with data, using statistical thinking and computational skills. You will also develop the associated soft skills of curiosity, communication and collaboration.

Learning outcomes

At the completion of this unit, you should be able to:

LO1. articulate the importance of statistics in a data-rich world, including current challenges such as ethics, privacy and big data
LO2. identify the study design behind a dataset and how the study design affects context specific outcomes
LO3. produce, interpret and compare graphical and numerical summaries, using base R and ggplot (extension)
LO4. apply the normal approximation to data, with consideration of measurement error
LO5. model and explain the relationship between 2 variables using linear regression
LO6. use the box model to describe chance and chance variability, including sample surveys and the central limit theorem
LO7. given real multivariate data and a problem, formulate an appropriate hypothesis and perform a range of hypothesis tests
LO8. interpret the p-value, conscious of the various pitfalls associated with testing
LO9. critique the use of statistics in media and research papers in a wide variety of data contexts, with attention to confounding and bias
LO10. perform data exploration in a team, and communicate the findings via oral and oral reproducible reports, with interrogation.

Graduate qualities

The Graduate Qualities are the qualities and skills that all University of Sydney graduates must demonstrate on successful completion of an award course. As a future Sydney graduate, the set of qualities have been designed to equip you for the contemporary world. Please see https://sydney.edu.au/students/graduate-qualities.html

Study commitment

- Typically, there is a minimum expectation of 1.5-2 hours of student effort per week per credit point for units of study offered over a full semester. For a 6 credit point unit, this equates to roughly 120-150 hours of student effort in total.

Teaching staff and contact details

Coordinator	:	Garth Tarr, garth.tarr@sydney.edu.au
Teaching staff	:	Diana Warren, diana.warren@sydney.edu.au Garth Tarr, garth.tarr@sydney.edu.au

Administrative and professional staff: DATA1001@sydney.edu.au Mathematics and Statistics Student Services Office, Carslaw 520

Learning activities schedule

Week	Topic	Learning activity type	Learning outcomes
Week 1	Design of experiments	Lecture	LO1, LO2, LO9, LO10
Week 1	RGuide - Lab 1	Computer laboratory	LO10
Week 2	Data & graphical summaries	Lecture	LO3
Week 2	Lab 2	Computer laboratory	LO2
Week 3	Numerical summaries	Lecture	LO3
Week 3	Lab 3	Computer laboratory	LO3
Week 4	Normal model	Lecture	LO4
Week 4	Lab 4	Computer laboratory	LO3
Week 5	Linear model	Lecture	LO5
Week 5	Lab 5	Computer laboratory	LO4, LO10
Week 6	Linear model	Lecture	LO5
Week 6	Lab 6	Computer laboratory	LO5
Week 7	Understanding chance	Lecture	LO6
Week 7	Lab 7	Computer laboratory	LO5
Week 8	Chance variability	Lecture	LO6
Week 8	Lab 8	Computer laboratory	LO6
Week 9	Sample surveys	Lecture	LO6
Week 9	Lab 9	Computer laboratory	LO6, LO10
Week 10	Hypothesis testing	Lecture	LO7, LO8
Week 10	Lab 10	Computer laboratory	LO6
Week 11	Tests for a mean	Lecture	LO7, LO8
Week 11	Lab 11	Computer laboratory	LO7, LO8
Week 12	Tests for a relationship	Lecture	LO7, LO8
Week 12	Lab 12	Computer laboratory	LO7, LO8
Week 13	Revision	Lecture	LO1, LO2, LO3, LO4, LO5, LO6, LO7, LO8
Week 13	Lab 13	Computer laboratory	LO7, LO8, LO10

Assessments

Assessment title	Assessment category	Assessment type	Assessment type description	Individual or group	Weight	Due date & time	Closing date	Length/duration	Learning outcomes
RQuizzes	Submitted work	Assignment	For self-learning R. Weekly online quizzes, paired with the RGuide.	Individual	10%	-		By the end of each week.	LO3, LO10
Project 1	Submitted work	Assignment	A data project based on own choice of data with written and verbal reports.	Group	10%	Week 4 30 August 2019 11.59pm		Self-directed learning till Week 4.	LO1, LO2, LO3, LO9, LO10
Project 2	Submitted work	Assignment	A data project based on own survey data with written and verbal reports.	Group	10%	Week 8 27 September 2019 11.59pm		Self-directed learning till Week 8.	LO1, LO2, LO3, LO4, LO5, LO9, LO10
Project 3	Submitted work	Assignment	A data project based on client data with written and verbal reports.	Individual	10%	Week 12 01 November 2019 11.59pm		Self-directed learning till Week 12.	LO1, LO2, LO3, LO4, LO5, LO6, LO7, LO8, LO9, LO10
Final exam	Exam	Final exam	Written Exam testing statistical thinking with given R Output.	Individual	60%	Formal exam period		2 hours	LO1, LO2, LO3, LO4, LO5, LO6, LO7, LO8, LO9, LO10

Assessment summary

- **RQuizzes**: The RQuizzes are designed to help you learn R, in preparation for each lab. The RQuizzes will be held on the DATA1001 Canvas site. Each RQuiz consist of 5 randomised questions. The best 10 of your 13 RQuizzes will count, making each worth 1%. The better mark principle will be used for the total marks on the RQuizzes so do not submit an application for Special Consideration or Special Arrangements if you miss a quiz. The better mark principle means that the total quiz mark counts if and only if it is better than or equal to your exam mark. If your total quiz mark is less than your exam mark, the exam mark will be used for that portion of your assessment instead.
- **Projects**: The data projects are designed to develop your statistical literacy and computational ability. They must be submitted electronically, as an HTML file via the DATA1001 Canvas site by the deadline. It is your responsibility to check that your project has been submitted correctly, otherwise it will not be marked. The better mark principle does not apply to the projects, as they assess different learning outcomes to the final exam. Late submissions will receive a penalty.
- **Final exam**: There is one examination of 2 hours' duration during the examination period.

Prescribed readings

All readings for this unit can be accessed on the Library eReserve link available on Canvas.

- **Required textbook:** Statistics (4th Edition) – Freedman, Pisani, and Purves (2007). An e-text version is available.

Other resources

Other resources can be found on Canvas.

Late penalties

The Assessment Procedures 2011 provide that any written work submitted after 11:59pm on the due date will be penalised by 5% of the maximum awardable mark for each calendar day after the due date. If the assessment is submitted more than ten calendar days late, a mark of zero will be awarded.
However, a unit of study may prohibit late submission or waive late penalties only if expressly stated below.

Special consideration

A special consideration application can be made for short-term circumstances beyond your control, such as illness, injury or misadventure, which affect your preparation or performance in an assessment.

If you are eligible for special consideration, you must submit an online application and supporting documents within three working days of the assessment, unless exceptional circumstances apply.

Assessment grading

The University awards common result grades, set out in the Coursework Policy 2014 (Schedule 1).

As a general guide, a high distinction indicates work of an exceptional standard, a distinction a very high standard, a credit a good standard, and a pass an acceptable standard.

Result name	Mark range	Description
High distinction	85 - 100	Representing complete or close to complete mastery of the material.
Distinction	75 - 84	Representing excellence, but substantially less than complete mastery.
Credit	65 - 74	Representing a creditable performance that goes beyond routine knowledge and understanding, but less than excellence.
Pass	50 - 64	Representing at least routine knowledge and understanding over a spectrum of topics and important ideas and concepts in the course.
Fail	0 - 49	When you don't meet the learning outcomes of the unit to a satisfactory standard.

For more information see sydney.edu.au/students/guide-to-grades.

Educational integrity

While the University is aware that the vast majority of students and staff act ethically and honestly, it is opposed to and will not tolerate academic dishonesty or plagiarism and will treat all allegations of dishonesty seriously.

All written assignments submitted in this unit of study will be submitted to the similarity detection software known as Turnitin. Turnitin searches for matches between text in your written assessment task and text sourced from the Internet, published works and assignments previously submitted to Turnitin. If such matches indicate evidence of plagiarism or other forms of dishonesty to your teacher, they are required to report your work for further investigation.

Further information on academic honesty, academic dishonesty, and the resources available to all students can be found on the academic integrity pages on the current students website: https://sydney.edu.au/students/academic-integrity.html

Work, health and safety requirements

We are governed by the Work Health and Safety Act 2011, Work Health and Safety Regulation 2011 and Codes of Practice. Penalties for non-compliance have increased. Everyone has a responsibility for health and safety at work. The University's Work Health and Safety policy explains the responsibilities and expectations of workers and others, and the procedures for managing WHS risks associated with University activities.

Other requirements

- **Attendance:** Unless otherwise indicated, students are expected to attend a minimum of 80% of timetabled activities for a unit of study, unless granted exemption by the Associate Dean.
- **Lab attendance:** You should attend the lab given on your personal timetable. Attendance at labs will be recorded. Your attendance will not be recorded unless you attend the labs in which you are enrolled. While there is no penalty if 80% attendance is not met we strongly recommend you attend labs regularly to keep up with the material and to engage with the lab questions. Since there is no assessment associated with the labs do not submit an application for Special Consideration or Special Arrangements for missed labs.

Site visit guidelines

There are no site visit guidelines for this unit.

Additional costs

There are no additional costs for this unit.

Closing the loop

No changes have been made since this unit was last offered.

Links to policies and other information for students

The following information and services are available to all students:

- Student administration: sydney.edu.au/study/student-administration.html
- Wellbeing and support: sydney.edu.au/students/health-wellbeing.html
- Study resources: sydney.edu.au/students/learning-services.html
- Expectations of student conduct: sydney.edu.au/students/student-responsibilities.html
- Learning and Teaching Policy: sydney.edu.au/policies/
- Academic appeals: sydney.edu.au/students/academic-appeals.html
- Libraries: sydney.edu.au/students/libraries.html
- Financial support: sydney.edu.au/students/financial-support.html

Other relevant information

More information can be found on Canvas.

Other links

Science student portal (canvas.sydney.edu.au/courses/7114)

Mathematics and Statistics student portal (canvas.sydney.edu.au/courses/7913)