



DATA1001 : Foundations of Data Science

Semester 1, 2019 | 6 Credit points | Mode of delivery: Normal (lecture/lab/tutorial) day | Unit type: Standard

Coordinator: Prof. Qiying Wang

Faculty of Science , School of Mathematics and Statistics

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.

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". You 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.** Give 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. For more information go to sydney.edu.au/students/graduate-qualities

- GQ1** Depth of disciplinary expertise
- GQ2** Critical thinking and problem solving
- GQ3** Communication (oral and written)
- GQ4** Information & digital literacy
- GQ5** Inventiveness
- GQ6** Cultural competence
- GQ7** Interdisciplinary effectiveness
- GQ8** Integrated professional, ethical and personal identity



GQ9 Influence

Study commitment

For a 6 credit point unit, this equates to roughly 120-150 hours of student effort in total.

Teaching staff and contact details

Coordinator: Prof. Qiying Wang, qiying.wang@sydney.edu.au
Teaching staff: Dr. Di Warren, diana.warren@sydney.edu.au
 A/Prof. Thomas Bishop, thomas.bishop@sydney.edu.au
 Dr. Floris Van Ogtrop, floris.vanogtrop@sydney.edu.au

Note: DATA1001 is being taught concurrently with ENVX1002.

Administrative and professional staff: DATA1001@sydney.edu.au

Weekly schedule

| Week | Topic | Learning activity type | Learning outcomes |
|--------|---|------------------------|-------------------|
| Week 1 | Design of Experiments 1. Introduction to Data Science and R 2. Controlled experiments 3. Observational Studies | Lecture | LO2 |
| Week 1 | RGuide - Lab1 | Computer laboratory | LO10 |
| Week 2 | Data & Graphical Summaries 1. Qualitative Data 2. Quantitative Data 3. Data Visualisation | Lecture | LO3 |
| Week 2 | Lab 2 | Computer laboratory | LO2 |
| Week 3 | Numerical Summaries 1. Centre 2. Spread 3. Data Wrangling | Lecture | LO3 |
| Week 3 | Lab 3 | Computer laboratory | LO3 |
| Week 4 | Normal Model 1. Normal Curve 2. Measurement Error 3. Reproducible Reports | Lecture | LO4 |
| Week 4 | Lab 4 | Computer laboratory | LO3 |
| Week 5 | Linear Model 1. Scatter Plot 2. Correlation Coefficient 3. Regression Line | Lecture | LO5 |
| Week 5 | Lab 5 | Computer laboratory | LO4 |
| Week 6 | Linear Model 1. Residual Plot 2. Linear Regression Summary 3. Non Linear Models | Lecture | LO5 |
| Week 6 | Lab 6 | Computer laboratory | LO5 |
| Week 7 | Understanding Chance 1. Chance 2. More Chance 3. Binomial formula | Lecture | LO6 |
| Week 7 | Lab 7 | Computer laboratory | LO5 |
| Week 8 | Chance Variability 1. Law of Averages | Lecture | LO6 |

| | | | |
|---------|---|---------------------|-----|
| | 2. Expected value and Standard error 3. Normal Approximation | | |
| Week 8 | Lab 8 | Computer laboratory | LO6 |
| Week 9 | Sample Surveys 1. Sample Surveys 2. Chance Errors in Sampling 3. Accuracy of Percentages | Lecture | LO6 |
| Week 9 | Lab 9 | Computer laboratory | LO6 |
| Week 10 | Hypothesis testing 1. Hypothesis testing 2. Percentage test 3. Percentage test demo | Lecture | LO7 |
| Week 10 | Lab 10 | Computer laboratory | LO6 |
| Week 11 | Tests for a mean 1. Accuracy of Means 2. Model for Measurement Error 3. Z and T tests | Lecture | LO7 |
| Week 11 | Lab 11 | Computer laboratory | LO7 |
| Week 12 | Tests for a relationship 1. 2 Sample T Test 2. Chi Square Tests 3. Regression tests | Lecture | LO7 |
| Week 12 | Lab 12 | Computer laboratory | LO7 |
| Week 13 | Revision | Lecture | |
| Week 13 | Lab 13 | Computer laboratory | LO7 |

Assessments

| Assessment title | Assessment category | Assessment type | Description of assessment type | Individual or group | Length / duration | Weight | Due date & time | Closing date | Learning outcomes |
|------------------|---------------------|--|---|---------------------|-------------------|--------|-------------------------|--------------|-------------------|
| RQuizzes | Submitted work | Tutorial quiz, small test or online task | Learning R | Individual | | 10% | Each Friday by 11:59pm | | L10 |
| Project 1 | Submitted work | Assignment | Analysis & presentation of own sourced data | Group | | 10% | Week 4: Friday 11:59pm | | LO1,LO10 |
| Project 2 | Submitted work | Assignment | Analysis & presentation of own survey data | Group | | 10% | Week 8: Friday 11:59pm | | LO1,LO10 |
| Project 3 | Submitted work | Assignment | Analysis & presentation of institutional data | Group | | 10% | Week 12: Friday 11:59pm | | LO1,LO10 |
| Final Exam | Exam | Final exam | | Individual | 2 Hrs | 60% | Exam period | | all |

Overview of assessments

Below are brief assessment details. Further information can be found in the Canvas site for this unit.

- **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%. You cannot apply for special consideration for the RQuizzes. The better mark principle will apply for the total 10% - i.e. if your overall exam mark is higher, then your 10% for RQuizzes will come from your exam. Work through the relevant section of the RGuide and then make up to 2 attempts at the RQuiz. The RQuizzes for each week should be completed before coming to each lab class, however the deadline for completion for all RQuizzes is 11:59 pm Friday following each lab class.

- **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 3 Projects, as they assess different learning outcomes to the final exam. There is a 20% penalty per day applied to late submissions.
- **Final exam:** There is one examination of 2 hours' duration during the examination period at the end of Semester 1. Further information about the exam will be posted on the DATA1001 Canvas site.

Readings

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

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

Other resources

- Please check the Canvas site for this unit for any information.

Late penalties

All assignments must be submitted by the due date. Students are expected to manage their time and to prioritise tasks to meet deadlines. Assessment items submitted after the due date without an approved extension using a special consideration or special arrangement form or request will incur penalties.

If you encounter a problem submitting your work on time, you may be able to arrange a simple extension. A simple extension is an informal arrangement between you and your unit of study coordinator. You may be able to receive an extension of up to two working days for non-examination tasks, as outlined in clause 66A of the Coursework Policy 2014. If you need an extension for a longer period, you may be eligible to apply for special consideration. sydney.edu.au/students/simple-extensions

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. sydney.edu.au/students/special-consideration-and-arrangements

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 detecting software program known as Turnitin. Turnitin searches for matches between text in your written assessment task and text sourced from the Internet, published works and assignments that have previously been submitted to Turnitin. If such matches indicate evidence of plagiarism to your teacher, they are required to report your work for further investigation.

Further information on academic honesty and the resources available to all students can be found on the Academic Integrity page of the current students' website: sydney.edu.au/students/academic-dishonesty

Work, health and safety requirements for this unit

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.

General Laboratory Safety Rules

- No eating or drinking is allowed in any laboratory under any circumstances
- A laboratory coat and closed-toe shoes are mandatory
- Follow safety instructions in your manual and posted in laboratories
- In case of fire, follow instructions posted outside the laboratory door
- First aid kits, eye wash and fire extinguishers are located in or immediately outside each laboratory
- As a precautionary measure, it is recommended that you have a current tetanus immunisation. This can be obtained from University Health Service: unihealth.usyd.edu.au/

Other requirements for this unit

Attendance requirement: 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.

Site visit guidelines

- Please check the Canvas site for this unit for any information. There are no site visit guidelines for this unit.

Additional costs

- Please check the Canvas site for this unit for any information. There are no additional costs for this unit.

Closing the loop

- Please check the Canvas site for this unit for any information. No changes have been made since this unit was last offered.

Links to policies and other information for students

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

Other relevant information

- Please check the Canvas site for this unit for any information.

Other links



- Science student portal: canvas.sydney.edu.au/courses/7114
- Mathematics and Statistics student portal: canvas.sydney.edu.au/courses/7913