



UNSW
SYDNEY

Faculty of Science

UNSW Mathematics and Statistics
UNSW Business School
UNSW Computer Science and Engineering

Course Outline

DATA1001

Introduction to Data Science and Decisions

Term 2, 2019

1. Staff

Position	Name	Email	Room
Course Authority – weeks 1-7	Prof Denzil Fiebig	d.fiebig@unsw.edu.au	Bus-444
Course Authority – weeks 8-10	Dr Pavel Krivitsky	p.krivitsky@unsw.edu.au	RC-TBA
Lecturer – weeks 1-3	Prof Denzil Fiebig	d.fiebig@unsw.edu.au	Bus-444
Lecturer – weeks 4-6	Dr Wenjie Zhang	wenjie.zhang@unsw.edu.au	CSE-508
Lecturer – weeks 8-10	Dr Pavel Krivitsky	p.krivitsky@unsw.edu.au	RC-TBA

Staff consultation times will be announced in Week 1 on Moodle.

2. Administrative matters

Contacting the Student Services Office

Please visit the School of Mathematics and Statistics web-site for a wide range of information on School Policies, Forms and Help for Students by visiting the “**Student Services**” page.

For information on Courses, please go to “Current Student”, “Undergraduate and/or Postgraduate” “**Courses Homepage**” for information on all **course offerings**.

The “Student Notice Board” can be located by going to the “Current Students” page; Notices are posted regularly for your information here. Please familiarise yourself with the information found in these locations. The School web page is found at: <http://www.maths.unsw.edu.au>

If you cannot find the answer to your queries on the web pages you are welcome to contact the Student Services Office directly.

By email to ug.mathsstats@unsw.edu.au

- By phone: 9385 7011
- Or in person to the Red Centre building, level 3, room 3072/3088

3. Course information

Units of credit: 6

Teaching times and locations: see the link on the Handbook web pages:

<https://www.handbook.unsw.edu.au/undergraduate/courses/2019/data1001/?q=data>

Course summary

This course will be taught as a number of distinct, but related, topics covering the fundamentals of Data Science as it is applied in Mathematics and Statistics, Computer Science, and Economics. The

course will be pitched at a level that is accessible for students as a General Education elective and it will also form a platform for students wishing to undertake further studies in Data Science. The course will provide an introduction to topics such as data analytics, data mining, Bayesian statistics, statistical software, econometrics, machine learning, business forecasting.

Course aims

The aim of this course is to provide students with an overview over the variety of analytical, technical and communication skills a Data Scientist should have, in order to process, analyse and extract impactful information from complex datasets.

Course learning outcomes

The Course Learning Outcomes are what you should be able to DO by the end of this course if you participate fully in learning activities and successfully complete the assessment items.

The Learning Outcomes in this course also help you to achieve some of the overall Program Learning Goals and Outcomes for all undergraduate coursework students in the participating Schools. Program Learning Goals are what we want you to BE or HAVE by the time you successfully complete your degree. You demonstrate this by achieving specific Program Learning Outcomes - what you are able to DO by the end of your degree.

Outcome 1: Statistics & Machine Learning Literacy

Students will be able to

- identify appropriate statistical methods for describing data & making inferences about population parameters
- use statistical reasoning to aid in decision making
- apply appropriate statistical methods to samples of data
- apply the framework of Bayesian reasoning
- apply machine learning concepts to numeric, categorical, textual and graph data

Outcome 2: Data Literacy

Students will be able to

- realize data models as relational database schemas
- formulate queries via SQL constructs
- understand the characteristics and challenges of big data
- solve a big data problem using cloud computing techniques

Outcome 3: IT Literacy

Students will be able to

- use R and SQL to pre-process and manipulate datasets
- formulate queries via SQL constructs
- Use R to apply appropriate statistical methods

Outcome 4: Learning autonomy

DATA1001 is one of the first courses in the program Data Science and Decisions, and it is delivered at a General Education level. Hence its students will start off at varying levels of mathematical ability, and many students will need to adapt to the fast-paced format of lectures at a university.

Students will hence learn how to

- find additional study material for both the theoretical and computational parts covered in the course, and
- manage their time in general, to be able to keep up with DATA1001 and their other courses.

4. Learning and teaching activities

Approach to Learning and Teaching in the Course

The philosophy underpinning this course and its Teaching and Learning Strategies are based on “Guidelines on Learning that Inform Teaching at UNSW”. These guidelines may be viewed at: www.guidelinesonlearning.unsw.edu.au. Specifically, the lectures, labs and assessment have been designed to appropriately challenge students and support the achievement of the desired learning outcomes. A climate of inquiry and dialogue is encouraged between students and teachers and among students (in and out of class). The lecturers and tutors aim to provide meaningful and timely feedback to students to improve learning outcome.

Weekly Lectures & Tutorial Schedule

Please note that Lectures commence in week 1 and run to week 9 according to your myUNSW *timetable*. Labs run from week 1 to week 10. Monday 10th June (Queen’s Birthday) is a public holiday. If your lab is scheduled for this day, you should make a special arrangement to attend another lab that week.

	Monday	Tuesday	Wednesday	Thursday	Friday
Lectures				Weeks 1-10 15:00-18:00 Law Th G04	
Lab M09A	Weeks 1-10 09:00-11:00 RedC G12A				
Lab T12A		Weeks 1-10 12:00-14:00 RedC G12C			
Lab W13A			Weeks 1-10 13:00-15:00 RedC G12C		
Lab W16A			Weeks 1-10 16:00-18:00 RedC G12C		

Lab F09A					Weeks 1-10 09:00-11:00 RedC G12C
Lab F12A					Weeks 1-10 12:00-14:00 RedC G12A
Lab F16A					Weeks 1-10 16:00-18:00 RedC G12C

Term Lecture Schedule

The schedule below is an approximation. Its order and contents may vary. Lectures are held every Thursday.

LECTURE SCHEDULE	
Week	Topic
Week 1	Course overview, data, displaying data, describing bivariate distributions, summary measures, introduction to R
Week 2	Linear regression, introduction to econometrics, data collection, selection bias, designing experiments, more R
Week 3	Introduction to data mining, data and decisions, business and economic research problems
Week 4	Introduction, ERD, relational model Assignment 1 Due
Week 5	SQL
Week 6	Big Data Management. Advanced topics (graph / social network / spatial data analysis)
Week 7	NO LECTURE Assignment 2 Due
Week 8	Statistical concepts and methods I
Week 9	Statistical concepts and methods II
Week 10	Statistical concepts and methods III Assignment 3 Due

Learning Activities and Teaching Strategies

The examinable content of the course is defined by the references given in the Lecture Schedule, the content of Lectures, and the content of the Lab Program.

Lectures

The purpose of lectures is to provide a logical structure for the topics that make up the course; to emphasize the important concepts and methods of each topic; and to provide relevant examples to which the concepts and methods are applied.

Labs

Labs begin in Week 1 and are an integral part of the subject. Lab presentations, discussions, solutions to problems are designed to help students deepen their understanding and practise learnt material.

Out-of-Class Study

While students may have preferred individual learning strategies, it is important to note that most learning will be achieved outside of class time. Lectures can only provide a structure to assist your study, and lab time is limited.

An “ideal” strategy (on which the provision of the course materials is based) might include:

1. Read the relevant chapter(s) of the text and relevant lecture slides **before the lecture**. This will give you a general idea of the topic area.
2. Attend lectures. Here the context of the topic in the course and the important elements of the topic are identified. The relevance of the topic should be explained.
3. Attempt lab questions before attending the lab class. This helps you identify issues that can be clarified or resolved in the lab class.

UNSW Moodle

The School of Mathematics and Statistics uses the Learning Management System called Moodle. To log into Moodle, use your zID and zPass at the following URL: <http://moodle.telt.unsw.edu.au>

Here you will find copies of: Course Outline, Lecture Slides; Lab Questions; Data sets required for the lab questions; How to Use R; examples of R programs; and Announcements. Students should consult this website at least once a week as it contains important information about the course.

It will be assumed that all students have seen Announcements posted on the course website.

Computing

Students will have hands on experience to statistics and machine learning methods during lab times. It will be advantageous if students bring their own portable computing devices (i.e. laptops) to the labs. For assignment work, students without their own laptops may use the computer labs in the Ground Floor of the Red Centre during times at which no other courses have booked these labs.

5. Assessment

- You will be able to view your final exam timetable once Exams Central has finalised the timetable. Please visit the web page: <https://my.unsw.edu.au/student.unsw.edu.au/exams> for details.
- It is very important that you understand the University's rules for the conduct of Examinations and the penalties for **Academic Misconduct Guide**. This information can be accessed through myUNSW at: <https://student.unsw.edu.au/exams> NB: In recent years there have been cases where severe penalties have been imposed for misconduct in relation to tests and exams in Maths courses.
- Assessment criteria: UNSW assesses students under a standards based assessment policy. For how this policy is applied within the School of Mathematics and Statistics, please visit the web site: <http://www.maths.unsw.edu.au/currentstudents/assessment-policies>
- If you are unwell / miss your **final examination**, please refer to the Special Consideration Policy by visiting the website: <https://student.unsw.edu.au/special-consideration>

Formal Requirements

To be eligible for a passing grade in this course, students must achieve a composite mark of at least 50 per cent.

Assessment Details

Assessment Tasks	Weight	Length	Due Date
3 Assignments	15% per assignment	≤ 4 pages	Weeks 4, 7, 10
Final Exam	55%	2 hours	University Exam Period
Total	100%		

Work commitments, holiday or travel or wedding plans, are NOT valid excuses for failing to complete any of the assessment tasks.

Late Submission of Assignments

20% of the value of the submission will be deducted for each day (24 hours). This rule is applicable to both hardcopy and softcopy submissions. Work submitted more than five days late will not be marked. If you delay submission, it is your responsibility to hand the assignment to your tutor. Staff members other than your tutor will NOT accept your project reports.

When Sickness Affects Your Submission

If you are unable to hand in your assignment or course project because of sickness, you must apply for special consideration. Applications for special consideration must be **lodged online through**

myUNSW (Log into myUNSW and go to My Student Profile tab > My Student Services channel > Online Services > Special Consideration).

Work commitments, holiday or travel or wedding plans are NOT valid excuses for failing to submit your assignments or course project.

Final Examination

The final exam will be held in the University examination period and will be 2 hours long. The final exam will cover the entire course. Further information on the content and structure of the Final Exam will be provided towards the end of session.

Calculator Information

For end of Term UNSW exams, students must supply their own calculator. Only calculators on the UNSW list of approved calculators may be used in the end of semester exams. Before the exam period, calculators must be given a “UNSW approved” sticker, obtained from the School of Mathematics and Statistics Office, and other student or Faculty centres. The UNSW list of calculators approved for use in end of semester exams is available at: <https://student.unsw.edu.au/exams>

6. Expectations of students

School Policies

The School of Mathematics and Statistics has adopted a number of policies relating to enrolment, attendance, assessment, plagiarism, cheating, special consideration etc. These are in addition to the Policies of The University of New South Wales. Individual courses may also adopt other policies in addition to or replacing some of the School ones. These will be clearly notified in the Course Initial Handout and on the Course Home Pages on the Maths Stats web site.

Students in courses run by the School of Mathematics and Statistics should be aware of the School and Course policies by reading the appropriate pages on the Maths Stats web site starting at:

<http://www.maths.unsw.edu.au/currentstudents/assessment-policies>

The School of Mathematics and Statistics will assume that all its students have read and understood the School policies on the above pages and any individual course policies on the Course Initial Handout and Course Home Page. Lack of knowledge about a policy will not be an excuse for failing to follow the procedure in it.

7. Course Evaluation and Development

Each year feedback is sought from students and other stakeholders about the courses offered in the School and continual improvements are made based on this feedback. UNSW's myExperience Survey Tool is one of the ways in which student evaluative feedback is gathered. You are strongly encouraged to take part in the feedback process.

8. Academic integrity, referencing and plagiarism

Academic integrity is fundamental to success at university. Academic integrity can be defined as a commitment to six fundamental values in academic pursuits: honesty, trust, fairness, respect, responsibility and courage.¹ At UNSW, this means that your work must be your own, and others' ideas should be appropriately acknowledged. If you don't follow these rules, plagiarism may be detected in your work.

Further information about academic integrity and **plagiarism** can be located at:

- The *Current Students* site <https://student.unsw.edu.au/plagiarism>, and
- The *ELISE* training site <http://subjectguides.library.unsw.edu.au/elise/presenting>

The *Conduct and Integrity Unit* provides further resources to assist you to understand your conduct obligations as a student: <https://student.unsw.edu.au/conduct>.

Further information about academic integrity and **plagiarism** can be located at:

- The *Current Students* site <https://student.unsw.edu.au/plagiarism>, and
- The *ELISE* training site <http://subjectguides.library.unsw.edu.au/elise/presenting>

The *Conduct and Integrity Unit* provides further resources to assist you to understand your conduct obligations as a student: <https://student.unsw.edu.au/conduct>.

9. Readings and resources

Text books

The lecturers of DATA1001 do not know of any single textbook which covers all of the material in this course, hence there is no single required, or even recommended, textbook. More advanced treatments of the topics covered in this course are presented in the following textbooks:

For computation in R and basic statistics:

- The R Software: Fundamentals of Programming and Statistical Analysis, by Pierre Lafaye de Micheaux, Remy Drouilhet and Benoit Liquet

A more comprehensive statistics textbook:

- Business Statistics, by Sharpe, De Veaux and Velleman

For databases:

- Fundamentals of Database Systems, by Ramez Elmasri and Shamkant Navathe

For statistical and machine learning:

¹ International Center for Academic Integrity, 'The Fundamental Values of Academic Integrity', T. Fishman (ed), Clemson University, 2013.

- An Introduction to Statistical Learning, by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani
- Pattern Recognition and Machine Learning, by Christopher Bishop

10. Additional support for students

- The Current Students Gateway: <https://student.unsw.edu.au/>
- Academic Skills and Support: <https://student.unsw.edu.au/academic-skills>
- Disability Support Services: <https://student.unsw.edu.au/disability-services>
- UNSW IT Service Centre: <https://www.it.unsw.edu.au/students/index.html>

Applications for Special Consideration

The University policy on Special Consideration applies to this course. Its details can be found on the webpage at

<https://student.unsw.edu.au/special-consideration>

Important Notes

- If you suffer from a chronic or ongoing illness that has, or is likely to, put you at a serious disadvantage, then you should contact the Disability Support Services who provide confidential support and advice. Their web site is: <https://student.unsw.edu.au/disability>
Disability Support Services (DSS) may determine that your condition requires special arrangements for assessment tasks. Once the School has been notified of these we will make every effort to meet the arrangements specified by DSS. The Disability support person in the School of Mathematics and Statistics is Markie Lugton Red Centre East Level 3 Room 3072 Ph: 9385-7011

University Statement on Plagiarism

This statement has been adapted from statements by the St James Ethics Centre, the University of Newcastle, and the University of Melbourne.

Plagiarism is the presentation of the thoughts or work of another as one's own. Examples include:

- Direct duplication of the thoughts or work of another, including by copying work, or knowingly permitting it to be copied. This includes copying material, ideas or concepts from a book, article, report or other written document (whether published or unpublished), composition, artwork, design, drawing, circuitry, computer program or software, web site, Internet, other electronic resource, or another person's assignment without appropriate acknowledgement
- Paraphrasing another person's work with very minor changes keeping the meaning, form and/or progression of ideas of the original;
- Piecing together sections of the work of others into a new whole;
- Presenting an assessment item as independent work when it has been produced in whole or part in collusion with other people, for example, another student or a tutor; and,
- Claiming credit for a proportion a work contributed to a group assessment item that is greater than that actually contributed.
- Submitting an assessment item that has already been submitted for academic credit elsewhere may also be considered plagiarism.
- The inclusion of the thoughts or work of another with attribution appropriate to the academic discipline does not amount to plagiarism.

The Learning Centre website is the central University online resource for staff and student information on plagiarism and academic honesty. It can be located at: <http://www.lc.unsw.edu.au/academic-integrity-plagiarism>

The Learning Centre also provides substantial educational written materials, workshops, and tutorials to aid students, for example, in:

- Correct referencing practices;
- Paraphrasing, summarising, essay writing, and time management;
- Appropriate use of, and attribution for, a range of materials including text, images, formulae and concepts.

Individual assistance is available on request from The Learning Centre.

Students are also reminded that careful time management is an important part of study and one of the identified causes of plagiarism is poor time management. Students should allow sufficient time for research, drafting, and the proper referencing of sources in preparing all assessment items.