THE UNIVERSITY OF SYDNEY
Semester 2, 2018

Information Sheet for **DATA1001 Foundations of Data Science**

General Information

Websites: It is important that you check both the Junior Mathematics website and the DATA1001 Learning Management System (Canvas) regularly.

Junior Mathematics webpage: `http://www.maths.usyd.edu.au/u/UG/JM/`
DATA1001 Canvas site: `https://canvas.sydney.edu.au/courses/6237`

Important announcements relating to Junior Mathematics are posted on the Junior Mathematics page. On the DATA1001 Canvas site you will find online resources and other useful links. Announcements regarding assessment tasks will be made on Canvas at various times throughout the semester.

Lectures: There is 1 lecture stream (three lectures per week), as shown on your personal timetable.

Times	Location	Lecturer(s)
8 am Mon, Tues & Wed	Eastern Avenue Auditorium (bottom level)	Dr Di Warren; Kevin Wang

Lectures run for 13 weeks. The first lectures will run in Week 1. The last lectures will run in Week 13.

Full lecture notes: Full lecture notes will be available on the DATA1001 Canvas site.

Consultation times: Consultation times will be posted on the DATA1001 Canvas site.

Labs: Labs (one per week) start in week 1. You should attend the lab given on your personal timetable. Attendance at labs will be recorded. Attendance does not count directly towards your mark, but allows tutors to monitor your progress and is taken into account with requests for special consideration. Labs use the computer software R.

Lab exercise sheets: The lab sheets will be available on the DATA1001 Canvas site.

Textbook: *Statistics* (4th Edition) – Freedman, Pisani, and Purves (2007). All students should have access to the text book, which is available in 3 forms: 1) E-text $65 (www.wileydirect.com.au/buy/statistics-4th-international-student-edition/), 2) hard copy (Co-op Bookshop), and 3) the Library.

Final Grades
Final grades are returned within one of the following bands:

> **High Distinction (HD), 85–100:** representing complete or close to complete mastery of the material; **Distinction (D), 75–84:** representing excellence, but substantially less than complete mastery; **Credit (CR), 65–74:** representing a creditable performance that goes beyond routine knowledge and understanding, but less than excellence; **Pass (P), 50–64:** representing at least routine knowledge and understanding over a spectrum of topics and important ideas and concepts in the course.

A student with a passing or higher grade should be well prepared to undertake further studies in data science, statistics, and/or mathematics which are dependent on this unit of study.

Assessment

Assessment: Your final raw mark for this unit will be calculated as follows:

10%: RQuizzes (best 10 of 13 quizzes)
10%: Project 1
10%: Project 2
10%: Project 3
60%: Exam at end of Semester 2

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% - ie 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 (full schedule is listed below). We recommend that you follow the due dates outlined on the Unit Schedule in Canvas to gain the most benefit from these quizzes.

Projects: The projects designed to develop your statistical literacy and computational ability.
- They must be submitted electronically, **as an HTML file** in Turnitin (an internet-based plagiarism-prevention service), via the DATA1001 Canvas site by the deadline. It is your responsibility to check that your assignment has been submitted correctly, otherwise it will not be marked.
- The 1st project must be completed by a group; the 2nd project can be completed as either a group or individual; and the 3rd project is completed individually.
- The better mark principle does not apply to the 3 Projects, as the 3 Projects assess different learning outcomes to the final exam. There is a 20% penalty per day applied to late submissions.

Assessment and feedback schedule:

Task	**Available**	**Deadline/date**	**Latest extension***	**Feedback**
RQuiz 1	30 Jul	10 Aug (Week 2)		On submission
RQuiz 2	30 Jul	10 Aug (Week 2)		On submission
RQuiz 3	30 Jul	17 Aug (Week 3)		On submission
RQuiz 4	30 Jul	24 Aug (Week 4)		On submission
RQuiz 5	30 Jul	31 Aug (Week 5)		On submission
RQuiz 6	30 Jul	7 Sep (Week 6)		On submission
RQuiz 7	30 Jul	14 Sep (Week 7)		On submission
RQuiz 8	30 Jul	21 Sep (Week 8)		On submission
RQuiz 9	30 Jul	5 Oct (Week 9)		On submission
RQuiz 10	30 Jul	12 Oct (Week 10)		On submission
RQuiz 11	30 Jul	19 Oct (Week 11)		On submission
RQuiz 12	30 Jul	26 Oct (Week 12)		On submission
RQuiz 13	30 Jul	2 Nov (Week 13)		On submission
Project 1	6 Aug	11:59 pm Fri 24 Aug (end of Week 4)	11:59 pm Fri 31 Aug	Mon 3 Sep
Project 2	6 Aug	11:59 pm Fri 21 Sep (end of Week 8)	11:59 pm Fri 28 Sep	Mon 1 Oct
Project 3	6 Aug	11:59 pm Fri 26 Oct (end of Week 12)	11:59 pm Fri 2 Nov	Mon 5 Nov

* Extensions for assignments are only possible for students registered with Disability Services or applying for Special Consideration or Special Arrangements.

Examination: There is one examination of 3 hours' duration during the examination period at the end of Semester 2. Further information about the exam will be posted on the DATA1001 Canvas site.

Any Questions?

Before you contact us with any enquiry, please check the Get Help/FAQ page on Canvas and the Junior Mathematics FAQ page:

Canvas: `https://canvas.sydney.edu.au/courses/6237/pages/contact-details-slash-faqs`
Junior Mathematics FAQ: `http://www.maths.usyd.edu.au/u/UG/JM/FAQ.html`

- For administrative matters, go to the *Student Services Office, Carslaw 520*. You may also email `DATA1001@maths.usyd.edu.au`. Ensure that any emails that you send to this address contain your name and SID, because anonymous emails will be ignored.
- For help with course material, see your lecturer, your tutor, or use the Ed discussion forum (`https://edstem.org`) which you can access throuh Canvas. Lecturers guarantee to be available during their indicated office hours, but may be available at other times as well. Do not post questions about RQuizzes on Ed, as these are assessable.
- If you are having difficulties with mathematics due to insufficient background, you may seek help from the *Mathematics Learning Centre, Carslaw 177*.

Learning Outcomes

Overall: 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: You will learn to:
1. Articulate the importance of statistics in a data-rich world, including current challenges such as ethics, privacy and big data.
2. Identify the study design behind a dataset and how the study design affects context specific outcomes.
3. Produce, interpret and compare graphical and numerical summaries, , using base R and ggplot (extension).
4. Apply the Normal approximation to data, with consideration of measurement error.
5. Model and explain the relationship between 2 variables using linear regression.
6. Use the box model to describe chance and chance variability, including sample surveys and the central limit theorem.
7. Given real multivariate data and a problem, formulate an appropriate hypothesis and perform a range of hypothesis tests.
8. Interpret the p-value, conscious of the various pitfalls associated with testing.
9. Critique the use of statistics in media and research papers in a wide variety of data contexts, with attention to confounding and bias.
10. Perform data exploration in a team, and communicate the findings via oral and oral reproducible reports, with interrogation.

Week-by-week outline:

Week	Lectures	Labs
Module1: Exploring Data		
1	Design of Experiments Pre reading: Introduction to DATA1001 L01: Introduction to Data Science and R L02: Controlled Experiments L03: Observational Studies	 RGuide Lab1

Continued on next page

Continued from previous page

Week	Lectures	Labs
2	Data & Graphical Summaries L04: Qualitative Data L05: Quantitative Data L06: Data Visualisation	Lab2
3	Numerical Summaries L07: Centre L08: Spread L09: Data Wrangling	Lab3
Module2: Modelling Data		
4	Normal Model L10: Normal Curve L11: Measurement Error L12: Reproducible Reports [Good Friday: Stream2]	Lab4
5	Linear Model L13: Scatter Plot L14: Correlation Coefficient L15: Regression Line	Lab5
6	Linear Model L16: Residual Plot L17: Linear Regression Summary L18: Non Linear Models	Lab6
Module3: Sampling Data		
7	Understanding Chance L19: Chance L20: More Chance L21: Binomial Formula [Anzac Day: Stream1]	Lab7
8	Chance Variability L22: Law of Averages L23: Expected Value and Standard Error L24: Normal Approximation	Lab8
9	Sample Surveys L25: Sample Surveys L26: Chance Errors in Sampling L27: Accuracy of Percentages	Lab9
Module4: Decisions with Data		
10	Hypothesis Testing L28: Hypothesis Testing L29: Percentage Test L30: Percentage Test Demo	Lab10
11	Tests for a mean L31: Accuracy of Means L32: Model for Measurement Error L33: Z and T tests	Lab11
12	Tests for a relationship L34: 2 Sample T Test L35: Chi Square Tests L36: Regression Tests	Lab12
13	Revision L37-L39: Revision	Lab13

ADDITIONAL INFORMATION

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.
For some units of study the minimum attendance requirement, as specified in the relevant table of units or the unit of study outline, may be greater than 80%.
The Associate Dean may determine that a student has failed a unit of study because of inadequate attendance.
Further details are available from the Science Undergraduate Handbook 2018: `http://sydney.edu.au/handbooks/science/coursework/faculty_resolutions.shtml` and the Science Postgraduate Handbook 2018: `http://sydney.edu.au/handbooks/science_PG`.

Online Components: This unit of study requires regular use of the University's Learning Management System (LMS). This could be either Blackboard or Canvas. Internet access is required to use the LMS.

Assessment Submission: Assessment tasks must be submitted by the due date. Submission will be online through the LMS unless instructed otherwise.

Compliance Statement All students must submit a signed statement of compliance with each piece of work submitted to the University for assessment, presentation or publication. A statement of compliance certifies that no part of the work constitutes a breach of the Academic Honesty in Coursework Policy 2015: `https://sydney.edu.au/policies/showdoc.aspx?recnum=PDOC2012/254&RendNum=0`.
This will be completed as part of the Turnitin assignment submission.

Late Submissions Work not submitted by the due date is subject to a penalty of 20% per day late.

Academic Dishonesty and Plagiarism Academic honesty is a core value of the University. Therefore, all students are required to act honestly, ethically and with integrity. Academic dishonesty is defined as any dishonest or unfair action taken in order to gain academic advantage. It also includes knowingly assisting another student to do this.
The University will not tolerate academic dishonesty or plagiarism, and will treat all allegations of academic dishonesty and plagiarism seriously.
Plagiarism is defined as presenting another persons work as ones own by presenting, copying or reproducing it without appropriate acknowledgement of the source.
Plagiarism includes presenting work for assessment, publication, or otherwise, that includes:

a. phrases, clauses, sentences, paragraphs or longer extracts from published or unpublished work (including from the internet) without appropriate acknowledgement of the source; or

b. the work of another person, without appropriate acknowledgement of the source and in a way that exceeds the boundaries of legitimate co-operation.

Further information is available in the Academic Honesty in Coursework Policy 2015: `https://sydney.edu.au/policies/showdoc.aspx?recnum=PDOC2012/254&RendNum=0`.

Similarity Detection Software Students should be aware that the University has authorised and mandated the use of the text-based similarity detecting software called Turnitin for all text-based written assignments. 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 for analysis.
Further information regarding plagiarism detection is available in the Academic Honesty in Coursework Policy 2015: `https://sydney.edu.au/policies/showdoc.aspx?recnum=PDOC2012/254&RendNum=0`.

Academic Honesty Education Module (AHEM) All students commencing their study at the University of Sydney are required to complete the Academic Honesty Education Module. You will find the AHEM in your Learning Management System.

Special Consideration In the event of serious illness or misadventure which affects your preparation or performance in an assessment task, you may be eligible for Special Consideration. Further information is available at: https://sydney.edu.au/students/special-consideration-and-arrangements.html. You should *not* submit an application for Special Consideration or Special Arrangements for this unit of study

- if you are absent from a tutorial and there is no assessment associated with the missed tutorial, or
- if you miss a quiz, since the better mark principle applies.

The assessment category for the assignments is "Submitted Work".

Student Feedback: The Unit of Study Survey
At the completion of each Unit of Study, students are asked via email to complete an online survey to provide feedback on their experiences in that Unit of Study. This feedback is invaluable when reviewing curriculum design and implementation styles.

University Work, Health and Safety Policy: 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 (http://www.unihealth.usyd.edu.au/).
For more details please refer to Emergencies and safety on campus: https://sydney.edu.au/students/emergencies-and-safety-on-campus.html

Student Support Services:
A guide for new students:
https://sydney.edu.au/students/browse.html?category=new-students&topic=getting-started
Counselling and mental health support:
https://sydney.edu.au/students/counselling-and-mental-health-support.html
Disability Support:
https://sydney.edu.au/students/disability-support.html
International Student Support:
https://sydney.edu.au/students/support-for-international-students.html
Learning Services / Study Skills Support:
https://sydney.edu.au/students/learning-services.html
Student IT and online learning:
https://sydney.edu.au/students/browse.html?category=student-it-and-online-learning&topic=online-learning
Academic Writing:
https://sydney.edu.au/students/writing.html