

THE UNIVERSITY OF SYDNEY
Semester 1, 2014
Information Sheet for STAT5002

Websites:

STAT5002 webpage: <http://www.maths.usyd.edu.au/u/PG/STAT5002/>

STAT5002 Blackboard page: via <https://myuni.sydney.edu.au>

On the STAT5002 Blackboard page you will find on-line resources and other useful links. Announcements regarding assessment tasks will be made on the STAT5002 Blackboard page at various times throughout the semester.

Class times and venues

Class	Time	Location	Taken by
Lecture	6-8pm Wednesday	Carlaw: Theatre 175	Dr Hilary Green hilary@math.usyd.edu.au
Tutorial/Practical	8-9pm Wednesday	Carlaw: Room 610	

Lectures and Tutorials

All classes run for 13 weeks.

Attendance at lectures and tutorials will be monitored. Please contact Hilary should you not be able to attend any of your classes.

The lecture and tutorial notes for a given week will be available on the STAT5002 Blackboard page on Tuesdays prior to each lecture/tutorial. In general, the tutorial material will be based on the lecture material of the same week.

You should bring a printed copy of the current week's lecture and tutorial notes to your classes each week. The lecture and tutorial notes will be provided for you in class in Week 1 only.

Most tutorial problems will require the use of the computer software R.

Textbook

There is no set text for STAT5002.

Online reading and reference materials will be recommended each week via the STAT5002 Blackboard page.

Where to go for help

For administrative matters, go to the Student Services Office, Carlaw 520.

For help with the unit, see Hilary.

If you are having difficulties with mathematics due to insufficient background, you may seek help from the Mathematics Learning Centre, Carlaw 455.

Assessment

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

Assessment	Allocation	Date due
Test 1	5%	Held in Week 3 Tutorial
Assignment 1	8%	Week 6 Tutorial
Test 2	15%	Held in Week 9 Tutorial
Assignment 2	12%	Week 12 Tutorial
Examination	60%	Exam period

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.

Examination

There will be an examination of 2 hours duration to be held during the examination period at the end of semester 1. Further information about the exam will be made available at a later date on the Blackboard website.

Students must perform satisfactorily in the final examination in order to pass the unit regardless of their performance throughout the semester.

Tests

Two short tests will be held during tutorials. Further information regarding the tests will be provided during lectures.

Assignments

Information regarding both assignments will be provided during lectures and on the STAT5002 Blackboard.

Objectives:

STAT5002 is an introductory statistics course which aims to introduce:

- ✓ the freely available statistical software package R and its applications;
- ✓ techniques for summarising univariate data of by means of elementary statistics, graphs and tables;
- ✓ probability theory to provide a mathematical framework for real life data
- ✓ modelling;
- ✓ a number of discrete and continuous distributions and their applications to real life problems;
- ✓ statistical inference and show how statistical tests can provide evidence for or against a scientific question;
- ✓ bivariate data analysis, correlation, simple linear regression and related problems with applications;
- ✓ categorical data analysis, contingency tables, goodness-of-fit tests and applications;
- ✓ some techniques for multivariate data analysis including non-parametric methods
- ✓ classification and selected methods
- ✓ some data mining concepts and methods

Outcomes:

Students who successfully complete this unit should be able to:

- ✓ use a variety of statistical computing skills using R and apply them appropriately;
- ✓ provide appropriate numerical and graphical summaries of data;
- ✓ calculate the probabilities of simple events, conditional probabilities and apply Bayes' Rule.
- ✓ understand concepts relating to discrete/continuous random variables, including expected value and variance, probability functions and cumulative distribution functions;
- ✓ identify some special discrete and continuous random variables; use R to calculate probabilities and construct graphs;
- ✓ understand the central limit theorem;
- ✓ understand the concept of a confidence interval and be able to apply it to real data;
- ✓ understand the concept of hypotheses tests and p-values for finding evidence for or against simple null hypotheses, in particular one-or two-sided z-, t- tests for making inference about population means or proportions;
- ✓ apply two-sample t-tests for making inference about the difference between two population means; use the Chi-squared test for simple contingency tables and goodness-of-fit problems.
- ✓ find the correlation coefficient and fit the least squares regression line as a way of describing a linear relationship in bivariate data; check assumptions and interpret results;
- ✓ understand the concepts of multiple regression and polynomial regression analysis, and use R to analyse multivariate data; interpret results and comment on suitability of the models.
- ✓ use R to analyse multivariate data using some nonparametric methods;
- ✓ understand some methods for exploratory multivariate analysis as well Principal component analysis
- ✓ understand some methods relating to classification problems including training/test set, logistic regression, cluster analysis and decision trees.
- ✓ apply graphical and statistical methods to explore large data sets.

STAT5002 Introduction to Statistics
Weekly Outline - Semester 1, 2014

Week	Lecture and Tutorial Topics
1	Introduction to Statistics and Data Mining; Populations, samples; Graphical and Numerical summaries of Data.
2	Introduction to Probability
3	Conditional Probability; Bayes' Rule
4	Random Variables; Expected Value and Variance of Discrete RV's. Some special Discrete Distributions
5	Random Variables; Expected Value and Variance of Continuous RV's; Some special Continuous Distributions; Central Limit Theorem; Sums of (some) Distributions.
6	Bivariate and multivariate concepts and distributions in statistics; Marginal, Conditional distributions, probabilities and expectations; Covariance and correlation. Independence.
7	Elements of statistical inference; Confidence Intervals; Hypothesis Testing 1. Background; One sample (parametric) tests for means and proportions.
8	Hypothesis Testing II. Types of Errors (brief). Two sample (parametric) tests for means and proportions (or chi-sq test of independence?) Bayesian Inference: The difference between Bayesian inference and frequentist inference. Prior Distribution.
9	Linear regression models: Least squares estimates. Simple, Multiple, Polynomial examples. Comparing models, best model. Nonparametric regression.
10	Catch up / Review
11	Exploratory Multivariate Analysis: Graphical Methods (Mosaic Plots, Trellis graphs, Scatterplot matrix, RGobi,); PCA.
12	Classification: Discriminant Analysis, Classification Theory.
13	Introduction to Data Mining / Review