

Subject code: 620-159

Subject name: Data Analysis 1

Credit points: 12.5

Coordinator: Karen Baker

Semesters of offer: 2

Prerequisites: Study score of 25 or more in VCE Mathematical Methods 3/4 or equivalent.

Mode of delivery: Lectures, tutorials and computer laboratory classes

Contact hours: 36 one-hour lectures (three per week), 11 one-hour tutorials (one per week), 11 one-hour computer laboratory classes (one per week)

Estimated total time commitment: 120 hours

Description: This subject lays the foundations for an understanding of the fundamental concepts of probability and statistics required for data analysis. Students should develop expertise in some of the statistical techniques commonly used in the design and analysis of experiments, and will gain experience in the use of a major statistical computing package. They should develop skills in collecting random samples, data description, basic statistical inference including parametric and nonparametric tests to compare population proportions and means, data manipulation and statistical computing. The methods will be illustrated using applications from science, engineering and commerce. Descriptive statistics, data manipulation and the implementation of the statistical procedures covered in lectures will be reinforced in the computer laboratory classes.

Sampling; introduction to experimental design; review of simple probability; estimation; confidence intervals; hypothesis testing including types of errors and power; inferences about means and proportions based on single and independent samples; matched pairs designs; introduction to nonparametric methods; contingency tables; regression; and analysis of variance.

Assessment: Up to 25 pages of written assignments 10% (due during semester); two 45-minute computer laboratory test 10% (held during semester); a 3-hour written examination 80% (held in the examination period).

Prescribed texts: J. M. Utts and R. F. Heckard, Mind on Statistics, 3rd edition, Duxbury, USA, 2005.

Notes: Students may only gain credit for one of [07]620-152, 620-159, [07]620-160, 620-168 or 316-130.

Students who have completed [07]620-202, [07]620-270 or [07]620-272 may not enrol in this subject for credit.

Subject objectives: Students completing this subject will:

- Understand the importance of random samples and experimental design in scientific research.
- Understand some fundamental concepts of statistical inference relating to confidence intervals and hypothesis testing.
- Use quantitative and graphical methods to describe a set of data.
- Develop expertise in the use of some common statistical techniques.
- Become familiar with a major statistical computing package.

Generic skills: In addition to learning specific skills that will assist students in their future careers in science, they will have the opportunity to develop generic skills that will assist them in any future career path. These include

- problem-solving skills: the ability to engage with unfamiliar problems and identify relevant solution strategies;
- analytical skills: the ability to construct and express logical arguments and to work in abstract or general terms to increase the clarity and efficiency of analysis;
- collaborative skills: the ability to work in a team;
- time management skills: the ability to meet regular deadlines while balancing competing commitments.
- computer skills: the ability to use an appropriate computing package.

Week-by-week outline:

1. Introduction: Populations and samples. Parameters and statistics. Difference between descriptive and inferential statistics.
2. Principles of experimental design and sampling: designed experiments, observational studies, random samples, randomisation. Examples of good and bad experimental designs.
3. Probability and Random Variables: review of simple probability, random variables and their distributions, the normal and binomial distributions.
4. Statistics as random variables. Random samples. Sampling distributions of means and proportions. The central limit theorem and the normal approximation to the binomial distribution.
5. Confidence intervals for proportions and means from single samples. Paired samples.
6. Testing hypotheses about proportions and means from single samples. P-value. One and two-sided tests.
GOOD FRIDAY/EASTER BREAK (nominal position)
7. Types of errors in hypothesis testing. Power, sample size.
ANZAC DAY (nominal position)

8. Nonparametric tests for single samples: Sign test, Wilcoxon signed rank test.
 9. Comparing populations using independent samples: Means and proportions. Nonparametric methods.
 10. Comparing more than two population means: One-way analysis of variance.
 11. Relationships between variables: Contingency tables, Simple linear regression.
 12. Design of research: sampling and experimental design in practice.
-