ST117-15 Introduction to Statistical Modelling

24/25

Department
>Statistics

Level
>Undergraduate Level 1

Module leader
>Martyn Parker

Credit value
>15

Module duration
>10 weeks

Assessment
>Multiple

Study location
>University of Warwick main campus, Coventry

Description

Introductory description

This module is an introduction to statistical thinking, modelling, and inference. At its core are the concepts of a statistical model and the associated likelihood function, as well as their manipulation to obtain rigorous inferences.

This module is core for students with their home department in Statistics and available to students from other departments for whom it is a listed option. It will be useful for all subsequent modules on statistics.

This module is NOT available as an unusual option. Students from outside the Statistics dept who are interested in a first year statistics module should consider taking ST121 Statistical Laboratory.

Pre-requisites:
Statistics students: ST118 Probability 1
Non-Statistics students: ST120 Introduction to Probability

[Module web page](#)

Module aims

To introduce the students to statistical thinking, formal reasoning under uncertainty, and the specification of a statistical model.

To build a foundation for likelihood-based statistical inference.

To connect mathematical models and inferences to real-world results, as well as provide practice in communicating them effectively.

To introduce computational tools and concepts necessary for modern data science.

To consider how collection, choice, or preprocessing of data sets influences the results of statistical analyses.

To convey basic ethical concepts arising with the generation, interpretation, and dissemination of data and information.

Outline syllabus

This is an indicative module outline only to give an indication of the sort of topics that may be covered. Actual sessions held may differ.

This module introduces the inherently interdisciplinary field of modern statistics. It covers the specification of appropriate statistical models for a variety of data sets, the mathematical underpinnings of model-based statistical inference, and the interpretation and communication of inference outcomes. The R computational language is introduced and used as a toolkit for examples throughout.

Learning outcomes

By the end of the module, students should be able to:

- Describe an appropriate probabilistic model and associated likelihood function for a simple data set.
- Calculate estimates of unknown parameters and their associated uncertainty based on a simple model and observed data.
- Compare model predictions and observed data graphically.
- Describe the modelling assumptions underlying a simple statistical model.
- Interpret model output to inform decisions or further experiments.
- Know the R programming environment well enough to write simple scripts to accomplish computational or data visualisation tasks.
- Discuss ethical aspects of a data collection/selection, their statistical analysis, and the interpretation and communication of the results.
- Apply team working processes to integrate ideas from various team members, break complex tasks into parts, and critically appraise each others contributions.

Indicative reading list

See Talis Aspire link.

[View reading list on Talis Aspire](#)

Interdisciplinary

The module includes dataset analysis arising from a range of areas. Studying these datasets requires utilising statistical knowledge and the ability to communicate outcomes across multiple domains and discplines.

Subject specific skills

- Select and apply appropriate mathematical and/or statistical techniques.
- Create structured and coherent arguments communicating them in written form.
- Construct and develop logical mathematical arguments with clear identification of assumptions and conclusions.
- Communicate subject-specific information effectively and coherently.
- Analyse problems, abstracting their essential information formulating them using appropriate mathematical language to facilitate their solution.
- Select and apply appropriate statistical programming language (for example, R) for exploratory data analysis.
- Understand major aspects of data collection, generation, and quality, and how this influences analyses and conclusions.

Transferable skills

- Critical thinking: extracting patterns from incomplete data and using them to form evidence-based conclusions.
- Problem solving: use of logical reasoning to build arguments grounded in evidence and with explicit underlying assumptions.
- Self-awareness: monitoring of your own learning and seeking feedback.
- Teamwork: collaboration with with members of your team in seminars, during self-study and during the completion of extended tasks.
- Communication: verbal discussion of ideas in seminars and among peers; written communication in assignments and the final project with members of your team.
- Information literacy: evaluation of data and uncertainty in a model-based way.
- Digital literacy: use of computational tools to understand and visualise data, and to produce reports.
- Professionalism: self-motivation, taking charge of your own learning, and prioritising effectively.
- Ethics: reflect on professional responsibilities as a statistician in conjunction with the generation and dissemination of information.
- Time management: dividing a long-term project into distinct phases and employing project management techniques to ensure timely delivery.

Study

Study time

Type	Required	Optional
Lectures	30 sessions of 1 hour (20%)	2 sessions of 1 hour
Seminars	10 sessions of 1 hour (7%)	
Private study	22 hours (15%)	
Assessment	88 hours (59%)	
Total	150 hours	

Private study description

Weekly revision of lecture notes, work on problem sheets, study for quizzes, participate in activities, preparation of the final project.

Costs

No further costs have been identified for this module.

Assessment

You do not need to pass all assessment components to pass the module.

Assessment group A1

	Weighting	Study time
Exercise sheets	35%	18 hours

Three exercise sheets supported by seminars, including both analytical and computational tasks. Theproblem sheets will contain a number of questions for which solutions and / or written responses will be required. The preparation and completion time noted below refers to the amount of time in hours that a well-prepared student who has attended lectures and carried out an appropriate amount of independent study on the material could expect to spend on this assignment.

	Weighting	Study time
Multiple Choice Quizzes	25%	12 hours

Multiple choice quizzes which will take place during the term that the module is delivered.

	Weighting	Study time
Final project	36%	50 hours

A written report on a project completed over an extended period, based on a project outline provided to the student.

The scope of the project spans the whole module syllabus and may include: specification of an appropriate likelihood from a description of an experiment, justified choice of estimators or statistical methods to answer a research question, execution of a statistical analysis in R, production of appropriate visualisations to assess inference results and model fit, and communication of the context and results of the analysis to a non-specialist audience. Should not

	Weighting	**Study time**

exceed 10 pages in length, including legible figures, displayed equations, and appropriate code snippets; the word limit is indicative of these expectations. Further details on the structure (e.g. word counts in sections) will be specified in written instructions.

	Weighting	Study time
Activity 1	2%	4 hours

Short example of data visualisation or aspects of data analysis to share in a very short oral presentation with other students.

	Weighting	Study time
Activity 2	2%	4 hours

Short example of data visualisation or aspects of data analysis to share in a very short oral presentation (a few minutes) with other students.

Assessment group R1

	Weighting	**Study time**
Reassessment as an individual project	100%	

This is an individual project replacing any parts of the module that need to be reassessed.

Feedback on assessment

Individual feedback will be provided on problem sheets by class tutors, and on the final project by the lecturer. A cohort-level summary will also be available for the project. Students are actively encouraged to make use of office hours to build up their understanding, and to view all their interactions with lecturers and class tutors as feedback.

Availability

Post-requisite modules

If you pass this module, you can take:

- ST236-10 Python for data-analytic tasks

Courses

This module is Core for:

- USTA-G302 Undergraduate Data Science
  - Year 1 of G302 Data Science
  - Year 1 of G302 Data Science
- Year 1 of USTA-G304 Undergraduate Data Science (MSci)
- Year 1 of USTA-G300 Undergraduate Master of Mathematics,Operational

Research,Statistics and Economics
- Year 1 of USTA-G1G3 Undergraduate Mathematics and Statistics (BSc MMathStat)
- USTA-GG14 Undergraduate Mathematics and Statistics (BSc)
  - Year 1 of GG14 Mathematics and Statistics
  - Year 1 of GG14 Mathematics and Statistics
- USTA-Y602 Undergraduate Mathematics,Operational Research,Statistics and Economics
  - Year 1 of Y602 Mathematics,Operational Research,Stats,Economics
  - Year 1 of Y602 Mathematics,Operational Research,Stats,Economics

This module is Option list B for:

- Year 1 of UECA-GL12 Undergraduate Mathematics and Economics (with Intercalated Year)