ST231-10 Linear Statistical Modelling with R

24/25

Department
Statistics
Level
Undergraduate Level 2
Module leader
Elke Thonnes
Credit value
10
Module duration
10 weeks
Assessment
Multiple
Study location
University of Warwick main campus, Coventry

Description

Introductory description

This module builds from ideas introduced in Year 1 Statistical Modelling and embeds them into the framework of linear models. Linear regression models are widely used in statistical practice and aim to explain or predict a continuous response variable using a collection of explanatory variables. Students will learn the theoretical background of such models, how to fit linear models to a given data set using R and how to interpret and evaluate the results.

Pre-requisistes:

- First Year Statistics Core (including ST117 Introduction to Statistical Modelling, ST118 Probability 1, and ST119 Probability 2) or equivalents.

AND

- Statistics students. Both ST229 Probability for Mathematical Statistics and ST230 Mathematical Statistics; or
- External students: ST232/ST233 Introduction to Mathematical Statistics.

Leads to

- ST340 Programming for Data Science
- ST344 Professional Practice of Data Analysis
- ST346 Generalised Linear Models for Regression and Classification.

Other third-year statistics modules.

Module web page

Module aims

1. Introduce the application of statistical modelling and statistical model exploration.
2. Use of R software and its use as a tool for statistical modelling, specifically for working with linear models in a variety of different scenarios.

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 theory of normal linear models and their practical application in R.

1. Normal linear models: definition and model assumptions.
2. Estimators for normal linear models and their sampling distributions.
3. Diagnostics and model building.
4. Confidence intervals and t-tests for normal linear models.
5. F-tests and analysis of variance; model selection and diagnostics.
6. Variable selection.

Learning outcomes

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

- Define a (normal) linear model and describe its modelling assumptions;
- Derive the properties of estimators for normal linear models; compute confidence intervals and perform hypothesis tests for normal linear models;
- Fit, diagnostically check, improve and compare regression models in R;
- Interpret and critically evaluate various linear models;
- Communicate solutions to problems accurately with structured and coherent arguments.

Indicative reading list

Sheather, S (2009) A modern approach to regression with R. Springer Science and Business Media.

View reading list on Talis Aspire

Research element

Students complete guided exploration of data sets as part of the coursework which provides a foundation for applied statistics research in later years.

Interdisciplinary

While not explicitly interdisciplinary, students are exposed to dataset from a variety of application contexts.

Subject specific skills

Demonstrate facility with advanced mathematical and probabilistic methods.

Demonstrate knowledge of key mathematical and statistical concepts, both explicitly and by applying them to the solution of mathematical problems.

Select and apply appropriate mathematical and/or statistical techniques

Create structured and coherent arguments communicating them in written form.

Select and applyappropriate computational techniques in a statistical programming language (for example, R)to build and evaluate linear models.

Transferable skills

Problem solving skills: The module requires students to solve problems presenting their conclusions as logical and coherent arguments.

Written communication skills: Students complete written assessments that require precise and unambiguous communication in the manner and style expected in mathematical sciences.

Verbal communication skills: Students are encouraged to discuss and debate formative assessment and lecture material within small-group tutorials sessions.

Students can continually discuss specific aspects of the module with the module leader. This is facilitated by statistics staff office hours.

Team working and working effectively with others: Students are encouraged to discuss and debate formative assessment and lecture material within small-group tutorials sessions.

Professionalism: Students work autonomously by developing and sustain effective approaches to learning, including time-management, organisation, flexibility, creativity, collaboratively and intellectual integrity.

Study

Study time

Type	Required	Optional
Lectures	20 sessions of 1 hour (20%)	2 sessions of 1 hour
Practical classes	10 sessions of 1 hour (10%)	
Private study	45 hours (45%)	
Assessment	25 hours (25%)	
Total	100 hours	

Private study description

Weekly revision of lecture notes and materials, wider reading and practice exercises, working on problem sets and preparing for examination.

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 D1

	Weighting	Study time
Lab. Report	20%	18 hours

You will use the statistical programming language R to carry out calculations and fit models on provided data sets in response to a set of questions. You will present, discuss, and evaluate the results. The length of the report will not exceed 18 pages, including figures, tables, code and R output,

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.

Set of short lab reports.	10%	5 hours

There will be approximately weekly problem sets. Each set will contain a number of individual questions based on the material delivered in the lectures. Problem sheets are supported by practical classes, including analytical, computational tasks and computer-based work. Assessment is based on solutions to the problems and engagement with in-class practical classes.

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 assessment.

Linear Statistical Modelling with R	70%	2 hours

	Weighting	Study time
examination		

You will be required to answer all questions on this examination paper.

- Answerbook Pink (12 page)
- Students may use a calculator
- Cambridge Statistical Tables (blue)

Assessment group R1

	Weighting	Study time
In-person Examination - Resit	100%	

You will be required to answer all questions on this examination paper.

- Answerbook Pink (12 page)
- Students may use a calculator
- Cambridge Statistical Tables (blue)

Feedback on assessment

Individual feedback will be provided on problem sheets by class tutors.

Solutions and cohort level feedback will be provided for the examination

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.

Past exam papers for ST231

Availability

Courses

This module is Core for:

- USTA-G302 Undergraduate Data Science
  - Year 2 of G302 Data Science
  - Year 2 of G302 Data Science
- USTA-GG14 Undergraduate Mathematics and Statistics (BSc)
  - Year 2 of GG14 Mathematics and Statistics

    - Year 2 of GG14 Mathematics and Statistics
- USTA-Y602 Undergraduate Mathematics,Operational Research,Statistics and Economics
    - Year 2 of Y602 Mathematics,Operational Research,Stats,Economics
    - Year 2 of Y602 Mathematics,Operational Research,Stats,Economics

This module is Option list B for:

- UCSA-G4G1 Undergraduate Discrete Mathematics
    - Year 2 of G4G1 Discrete Mathematics
    - Year 2 of G4G1 Discrete Mathematics
- Year 2 of UCSA-G4G3 Undergraduate Discrete Mathematics