ST221-12 Linear Statistical Modelling

20/21

Department
Statistics
Level
Undergraduate Level 2
Module leader
Ric Crossman
Credit value
12
Module duration
10 weeks
Assessment
Multiple
Study location
University of Warwick main campus, Coventry

Description

Introductory description

This module runs in the second half of term 2 and first of term 3. It is available for students on a course where it is a listed option and as an Unusual Option to students who have completed the prerequisite modules. It is strongly recommended for any students intending to do substantial data analysis.

Students wishing to pursue the integrated Masters MMORSE are expected to take ST221 in Year 2. Data Science students will find it highly relevant for their third year project. ST221 may form part of the criteria for determining places on ST modules with capped numbers such as ST340 Programming for Data Science and ST344 Professional Practice of Data Analysis.

Pre-requisites for Statistics students: ST115 Introduction to Probability, ST218 Mathematical Statistics A and ST219 Mathematical Statistics B (taken concurrently).
Pre-requisites for Non-Statistics students: ST111/ST112 Probability A & B and ST220 Introduction to Mathematical Statistics. Basic knowledge in R such as covered in ST104 Statistical Laboratory I will be useful.

Results from the coursework from this module may be partly used to determine exemption eligibility in the computer based assessment components of the Institute and Faculty of Actuaries modules CS1, CS2, CM1 and CM2. (Independent application to the IFoA may be required.)

[Module web page](#)

Module aims

To introduce the ideas and methods of statistical modelling and statistical model exploration. To introduce students to the application 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.

1. Introduction to the R software. Some useful methods of examining large data sets. The use of this package to obtain important summary features in different data structures.
2. A review of the simple linear regression. Distributions of estimators and residuals.
3. An introduction to multiple regression. Estimators of these models. How the study of residuals can inform and refine model choice. How to use R to check the plausibility of such a statistical model and how to use diagnostic plots in combination with the theory of model refinement.
4. Introduction of polynomial regression and various ANOVA models. The coding and interpretation of these models using R.
5. An introduction to linear models for time series and generalized linear models for frequency data.

Learning outcomes

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

- Make use of the language R to explore data sets with appropriate graphs and summary statistics.
- Make use of R to fit appropriate linear models to data sets.
- Understand how various linear models can be proposed, estimated, diagnostically checked, compared and criticised.

Indicative reading list

[View reading list on Talis Aspire](#)

Subject specific skills

TBC

Transferable skills

TBC

Study

Study time

Type	Required	Optional
Lectures	30 sessions of 1 hour (25%)	2 sessions of 1 hour
Practical classes	4 sessions of 1 hour (3%)	
Private study	50 hours (42%)	
Assessment	36 hours (30%)	
Total	120 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.

Students can register for this module without taking any assessment.

Assessment group D2

	Weighting	Study time	Eligible for self-certification
Assignment 1	15%	18 hours	Yes (extension)

Due in week 10 of term 2.
You will use the R program 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 number of words 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. 500 words is equivalent to one page of text, diagrams, formula or equations; your ST221 Assignment 1 should not exceed 18 pages in length.

	Weighting	Study time	Eligible for self-certification
Assignment 2	15%	18 hours	Yes (extension)

	Weighting	Study time	Eligible for self-certification

Due in week 3 of term 3.

You will use the R program 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 number of words 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. 500 words is equivalent to one page of text, diagrams, formula or equations; your ST221 Assignment 2 should not exceed 18 pages in length.

	Weighting	Study time	Eligible for self-certification
Online Examination	70%		No

The examination paper will contain four questions, of which the best marks of THREE questions will be used to calculate your grade.

~Platforms - Moodle

- Answerbook Green (8 page)
- Students may use a calculator
- Cambridge Statistical Tables (blue)

Assessment group R

	Weighting	Study time	Eligible for self-certification
Online Examination - Resit	100%		No

The examination paper will contain four questions, of which the best marks of THREE questions will be used to calculate your grade.

~Platforms - Moodle

- Answerbook Green (8 page)
- Students may use a calculator
- Cambridge Statistical Tables (blue)

Feedback on assessment

Reports will be marked and feedback returned to students within 20 working days.

Solutions and cohort level feedback will be provided for the examination.

Past exam papers for ST221

Availability

Courses

This module is Core for:

- Year 2 of USTA-G1G3 Undergraduate Mathematics and Statistics (BSc MMathStat)
- Year 2 of USTA-GG14 Undergraduate Mathematics and Statistics (BSc)

This module is Optional for:

- Year 2 of USTA-G302 Undergraduate Data Science
- Year 2 of USTA-G304 Undergraduate Data Science (MSci)

This module is Option list A for:

- Year 2 of USTA-G300 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics
- Year 2 of USTA-Y602 Undergraduate Mathematics,Operational Research,Statistics and Economics

This module is Option list B for:

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