

ST323-15 Multivariate Statistics

21/22

Department

Statistics

Level

Undergraduate Level 3

Module leader

Tom Berrett

Credit value

15

Module duration

10 weeks

Assessment

Multiple

Study location

University of Warwick main campus, Coventry

Description

Introductory description

This module runs in Term 1 and is an optional module intended for students in their third or fourth year of study who have previously taken preparatory modules in Statistics.

For Statistics students the pre-requisites are ST115 Introduction to Probability, ST218 Mathematical Statistics A, ST219 Mathematical Statistics B.

For Non-Statistics students the pre-requisites are ST111/112 Probability A&B and ST220 Introduction to Mathematical Statistics.

The coursework uses the statistical software package R, so basic knowledge in R such as covered in ST104 Statistical Laboratory I or ST952 Introduction to Statistical Practice is expected.

[Module web page](#)

Module aims

Multivariate data arises whenever several interdependent variables are measured simultaneously. Such high-dimensional data is becoming the rule, rather than the exception in many areas: in medicine, in the social and environmental sciences and in economics. The analysis of such multidimensional data often presents an exciting challenge that requires new statistical techniques which are usually implemented using computer packages. This module aims to give you a good and rigorous understanding of the geometric and algebraic ideas that these techniques are based

on, before giving you a chance to try them out on some real data sets.

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.

Multivariate data arises whenever several interdependent variables are measured simultaneously. Such high-dimensional data is becoming the rule, rather than the exception in many areas: in medicine, in the social and environmental sciences and in economics. The analysis of such multidimensional data often presents an exciting challenge that requires new statistical techniques which are usually implemented using computer packages. This module aims to give you a good and rigorous understanding of the geometric and algebraic ideas that these techniques are based on, before giving you a chance to try them out on some real data sets.

Learning outcomes

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

- Construct and Interpret graphical representations of multivariate data
- Carry out a principal components to summarise high dimensional data
- Perform clustering analysis to discover and characterize subgroups in the population.
- Use classification and discrimination methods to assign individuals into groups.
- Conduct inference for multivariate means, construct confidence regions, and understand their potential uses, such as for group comparisons.
- Understand any additional topics covered in the lectures. Time permitting, lectures will cover one or two additional topics such as Factor Analysis, Multidimensional Scaling, random forests, bagging, sparse multivariate methods, Gaussian graphical models, multiple testing, functional data analysis, spatial statistics, independent component analysis, compositional data analysis, canonical correlation analysis.

Indicative reading list

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*.: Pearson Prentice Hall. Upper Saddle River, NJ.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York: Springer.

Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The elements of statistical learning* (second edition). New York: Springer.

Efron, B., & Hastie, T. (2016). *Computer age statistical inference* (Vol. 5). Cambridge University Press.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.

[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 (20%)	2 sessions of 1 hour
Private study	90 hours (60%)	
Assessment	30 hours (20%)	
Total	150 hours	

Private study description

Weekly revision of lecture notes and materials, wider reading and practice exercises, working on assignments 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 D3

	Weighting	Study time
Assignment 1	10%	15 hours

Due in Term 1 Week 6.
The assignment will contain a number of questions for which solutions and / or written responses will be required.

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

Weighting**Study time**

text, diagrams, formula or equations; your ST323 Assignment 1 should not exceed 15 pages in length.

Assignment 2

10%

15 hours

Due in Term 2 Week 4.

The assignment will contain a number of questions for which solutions and / or written responses will be required.

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 ST323 Assignment 2 should not exceed 15 pages in length.

On-campus Examination

80%

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

The examination paper is the same that is provided for the advanced topics variant and will contain an additional question for students taking that variant.

~Platforms - Moodle

- Answerbook Pink (12 page)
- students may use a calculator
- Graph paper
- Cambridge Statistical Tables (blue)

Assessment group R1**Weighting****Study time**

Online Examination

100%

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

The examination paper is the same that is provided for the advanced topics variant and will contain an additional question for students taking that variant.

~Platforms - Moodle

- Online examination: No Answerbook required
- students may use a calculator
- Graph paper
- Cambridge Statistical Tables (blue)

Feedback on assessment

Marked assignments will be available for viewing at the support office within 20 working days of the submission deadline. Cohort level feedback and solutions will be provided, and students will be given the opportunity to receive feedback via face-to-face meetings.

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

[Past exam papers for ST323](#)

Availability

Courses

This module is Core optional for:

- USTA-G301 Undergraduate Master of Mathematics, Operational Research, Statistics and Economics (with Intercalated
 - Year 3 of G30F Master of Maths, Op.Res, Stats & Economics (Econometrics and Mathematical Economics Stream) Int
 - Year 4 of G30F Master of Maths, Op.Res, Stats & Economics (Econometrics and Mathematical Economics Stream) Int

This module is Optional for:

- Year 3 of UCSA-G4G1 Undergraduate Discrete Mathematics
- Year 3 of UCSA-G4G3 Undergraduate Discrete Mathematics
- Year 4 of UCSA-G4G2 Undergraduate Discrete Mathematics with Intercalated Year
- USTA-G300 Undergraduate Master of Mathematics, Operational Research, Statistics and Economics
 - Year 3 of G300 Mathematics, Operational Research, Statistics and Economics
 - Year 4 of G300 Mathematics, Operational Research, Statistics and Economics

This module is Core option list A for:

- USTA-G300 Undergraduate Master of Mathematics, Operational Research, Statistics and Economics
 - Year 3 of G30B Master of Maths, Op.Res, Stats & Economics (Econometrics and Mathematical Economics Stream)
 - Year 3 of G30D Master of Maths, Op.Res, Stats & Economics (Statistics with Mathematics Stream)
- USTA-G301 Undergraduate Master of Mathematics, Operational Research, Statistics and Economics (with Intercalated
 - Year 3 of G30H Master of Maths, Op.Res, Stats & Economics (Statistics with Mathematics Stream)
 - Year 4 of G30F Master of Maths, Op.Res, Stats & Economics (Econometrics and

Mathematical Economics Stream) Int

- Year 4 of G30H Master of Maths, Op.Res, Stats & Economics (Statistics with Mathematics Stream)

This module is Core option list B for:

- Year 3 of USTA-G300 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics
- USTA-G301 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics (with Intercalated
 - Year 3 of G30G Master of Maths, Op.Res, Stats & Economics (Operational Research and Statistics Stream) Int
 - Year 4 of G30G Master of Maths, Op.Res, Stats & Economics (Operational Research and Statistics Stream) Int

This module is Option list A for:

- Year 3 of USTA-G302 Undergraduate Data Science
- Year 3 of USTA-G304 Undergraduate Data Science (MSci)
- Year 4 of USTA-G303 Undergraduate Data Science (with Intercalated Year)
- Year 4 of USTA-G300 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics
- Year 5 of USTA-G301 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics (with Intercalated
- USTA-G1G3 Undergraduate Mathematics and Statistics (BSc MMathStat)
 - Year 3 of G1G3 Mathematics and Statistics (BSc MMathStat)
 - Year 4 of G1G3 Mathematics and Statistics (BSc MMathStat)
- USTA-G1G4 Undergraduate Mathematics and Statistics (BSc MMathStat) (with Intercalated Year)
 - Year 4 of G1G4 Mathematics and Statistics (BSc MMathStat) (with Intercalated Year)
 - Year 5 of G1G4 Mathematics and Statistics (BSc MMathStat) (with Intercalated Year)
- Year 3 of USTA-GG14 Undergraduate Mathematics and Statistics (BSc)
- Year 4 of USTA-GG17 Undergraduate Mathematics and Statistics (with Intercalated Year)
- Year 3 of USTA-Y602 Undergraduate Mathematics,Operational Research,Statistics and Economics
- Year 4 of USTA-Y603 Undergraduate Mathematics,Operational Research,Statistics,Economics (with Intercalated Year)

This module is Option list B for:

- UMAA-G105 Undergraduate Master of Mathematics (with Intercalated Year)
 - Year 3 of G105 Mathematics (MMath) with Intercalated Year
 - Year 5 of G105 Mathematics (MMath) with Intercalated Year
- Year 3 of USTA-G300 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics
- USTA-G301 Undergraduate Master of Mathematics,Operational Research,Statistics and Economics (with Intercalated
 - Year 3 of G30E Master of Maths, Op.Res, Stats & Economics (Actuarial and Financial

Mathematics Stream) Int

- Year 4 of G30E Master of Maths, Op.Res, Stats & Economics (Actuarial and Financial Mathematics Stream) Int
- Year 3 of UMAA-G100 Undergraduate Mathematics (BSc)
- UMAA-G103 Undergraduate Mathematics (MMath)
 - Year 3 of G103 Mathematics (MMath)
 - Year 4 of G103 Mathematics (MMath)
- UMAA-G106 Undergraduate Mathematics (MMath) with Study in Europe
 - Year 3 of G106 Mathematics (MMath) with Study in Europe
 - Year 4 of G106 Mathematics (MMath) with Study in Europe
- Year 4 of UMAA-G101 Undergraduate Mathematics with Intercalated Year