

# ST412-15 Multivariate Statistics with Advanced Topics

**21/22**

**Department**

Statistics

**Level**

Undergraduate Level 4

**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 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.

Students will be given selected advanced research material for independent study and examination.

## 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 and canonical correlation analysis 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.
- Assess multivariate normality and do multivariate tests for comparing means across groups
- 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.
- Understand by independent study an additional advanced topic in multivariate statistics.

## 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

Study of advanced topic, 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.

## Weighting

## Study time

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 ST412 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 ST412 Assignment 2 should not exceed 15 pages in length.

On-campus Examination

80%

The examination will contain one compulsory question on the advanced topic and four additional questions of which the best marks of TWO questions will be used to calculate your grade.

~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 will contain one compulsory question on the advanced topic and four additional questions of which the best marks of TWO questions will be used to calculate your grade.

~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 ST412](#)

---

## Availability

### Anti-requisite modules

If you take this module, you cannot also take:

- ST323-15 Multivariate Statistics

## 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:

- TMAA-G1PE Master of Advanced Study in Mathematical Sciences
  - Year 1 of G1PE Master of Advanced Study in Mathematical Sciences
  - Year 1 of G1PE Master of Advanced Study in Mathematical Sciences
- Year 1 of TMAA-G1P9 Postgraduate Taught Interdisciplinary Mathematics
- Year 1 of TMAA-G1PD Postgraduate Taught Interdisciplinary Mathematics (Diploma plus MSc)
- Year 1 of TMAA-G1P0 Postgraduate Taught Mathematics
- Year 1 of TMAA-G1PC Postgraduate Taught Mathematics (Diploma plus MSc)
- Year 1 of TSTA-G4P1 Postgraduate Taught Statistics
- 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 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)

This module is Option list B for:

- Year 4 of USTA-G304 Undergraduate Data Science (MSci)
- Year 4 of UCSA-G4G3 Undergraduate Discrete Mathematics
- 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

This module is Option list D for:

- 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