

Data Intensive Analysis

Programme Requirements:

Data-Intensive Analysis - MSc
MT4113 (15 credits) and MT5761 (15 credits) and MT5762 (15 credits) and MT5763 (15 credits) and MT5764 (15 credits) and ID5059 (15 credits) and Between 0 and 30 credits from Module List: CS5001 - CS5003, CS5044, CS5052 and (CS5099 (60 credits) or MT5099 (60 credits))

Compulsory modules:

MT4113 Computing in Statistics				
SCOTCAT Credits:	15	SCQF Level 10	Semester	1
Academic year:	2018/9			
Planned timetable:	12.00 noon Mon (odd weeks) and Wed, 12.00 noon - 2.00 pm Fri			
The aim of this module is to teach computer programming skills, including principles of good programming practice, with an emphasis on statistical computing. Practical work focusses on the widely-used statistical language and environment R. Practical skills are developed through a series of computing exercises that include (1) modular programming; (2) manipulating data; (3) simulating data with specific statistical properties, (4) investigating behaviour of statistical procedures under failure of statistical assumptions.				
Pre-requisite(s):	Before taking this module you must pass MT2508			
Learning and teaching methods of delivery:	Weekly contact: 1.5-hour lectures (x 10 weeks), 2-hour practical classes (x 10 weeks)			
Assessment pattern:	2-hour Written Examination = 40%, Coursework = 60%			
Re-assessment pattern:	1-hour 40 minute Written Examination = 40%, Coursework (4 new programming assignments) = 60%			
Module coordinator:	Prof L J Thomas			
Module teaching staff:	Prof L Thomas			

Mathematics & Statistics - Data Intensive Analysis - 2018/9 - August 2018

MT5761 Statistical Modelling				
SCOTCAT Credits:	15	SCQF Level 11	Semester	1
Academic year:	2018/9			
Planned timetable:	Mon, Tues, Thur, Fri 3:00 - 4:00 (lectures), Tues, Thur 4:00 - 5:00 (practicals)			
This applied statistics module covers the main aspects of linear models (LMs) and generalized linear models (GLMs). In each case the course describes model specification, various options for model selection, model assessment and tools for diagnosing model faults. Common modelling issues such as collinearity and residual correlation are also addressed, and as a consequence of the latter the Generalized Least squares (GLS) method is described. The GLM component has emphasis on models for count data and presence/absence data while GLMs for multinomial (sometimes called choice-based models) are also covered for nominal and ordinal response outcomes. The largest part of the course material is taught inside an environmental impact assessment case study with reality-based research objectives. Political and medical examples are used to illustrate the multinomial models.				
Pre-requisite(s):	Undergraduates must have passed at least one of MT4113, MT4527, MT4528, MT4530, MT4531, MT4537, MT4539, MT4606, MT4608, MT4609, MT4614.			
Anti-requisite(s)	You cannot take this module if you take MT4607 or take MT5753			
Learning and teaching methods of delivery:	Weekly contact: 4 lectures (x 5 weeks), 2 practicals (x 5 weeks)			
	Scheduled learning: 30 hours		Guided independent study: 117 hours	
Assessment pattern:	As used by St Andrews: 2-hour Written Examination = 50%, Coursework = 50%			
Re-assessment pattern:	2-hour Written Examination = 100%			
Module teaching staff:	TBC			

MT5762 Introductory Data Analysis				
SCOTCAT Credits:	15	SCQF Level 11	Semester	1
Academic year:	2018/9			
Availability restrictions:	Not available to Undergraduate students.			
Planned timetable:	Mon, Tue, Fri 2:00 - 3:30, Thur 3:30 - 5:00			
This module provides coverage of essential statistical concepts and analysis methods relevant to commercial analysis. Specifically: the different types of data and their numerical/graphical treatment; basic probability theory and concepts of inference; fundamental statistical concepts with particular emphasis on sampling issues; basic statistical models and tests; linear models; introductory computer-intensive inference. This module is a short intensive course and is a core, preliminary, requirement for the MSc in Applied Statistics and Datamining. It covers material essential for study of the more advanced statistical methods encountered in subsequent modules.				
Anti-requisite(s)	You cannot take this module if you take MT5756			
Learning and teaching methods of delivery:	Weekly contact: Four 1.5-hour lectures (x 5 weeks)			
Assessment pattern:	Coursework = 100%			
Module coordinator:	Dr C R Donovan			
Module teaching staff:	Dr D Donovan, Dr L Scott-Hayward			

Mathematics & Statistics - Data Intensive Analysis - 2018/9 - August 2018

MT5763 Software for Data Analysis				
SCOTCAT Credits:	15	SCQF Level 11	Semester	1
Academic year:	2018/9			
Availability restrictions:	Not available to Undergraduate students			
Planned timetable:	Mon, Tues, Fri 3:30 - 4:30 (lectures). Mon, Tues, Fri 4:30 - 5:30 (Practicals)			
<p>This module covers the practical computing aspects of statistical data analysis, focussing on packages most widely used in the commercial sector (R, SAS, SPSS & Excel). We cover the accessing, manipulation, checking and presentation of data (visual and numerical). We fit various statistical models to data, with subsequent assessment, interpretation and presentation. Good practice and 'reproducible research' is covered, as is computer intensive inference and big data considerations. This module is a short intensive course and is a core, preliminary, requirement for the MSc in Applied Statistics and Datamining and the MSc in Data Intensive Analysis. It covers material essential for study of the more advanced statistical methods encountered in subsequent modules.</p>				
Pre-requisite(s):	Pass in MT1007 or MT3507 or MT3508 or be taking MT5762			
Anti-requisite(s)	You cannot take this module if you take MT5756			
Learning and teaching methods of delivery:	Weekly contact: Three 2-hour lecture/practical classes (x 5 weeks)			
	Scheduled learning: 30 hours		Guided independent study: 120 hours	
Assessment pattern:	As used by St Andrews: Coursework = 100%			
Re-assessment pattern:	Coursework = 100%			
Module teaching staff:	TBC			

MT5764 Advanced Data Analysis				
SCOTCAT Credits:	15	SCQF Level 11	Semester	2
Academic year:	2018/9			
Planned timetable:	Mon 12:00-1:00 Weeks 2, 4, 5, 8, 10 Tues, Thur 12:00-2:00, Weeks 1-10 (lectures) Tues 2:00 - 3:00 Weeks 2-9 (practicals)			
<p>This module covers modern modelling methods for situations where the data fails to meet the assumptions of common statistical models and simple remedies do not suffice. This represents a lot of real world data. Methods covered include: nonlinear models; basic splines and Generalised Additive Models; LASSO and the Elastic Net; models for non-independent errors and random effects. Pragmatic data imputation is covered with associated issues. Computer intensive inference is considered throughout. Practical applications build sought-after skills in R and the commercial packages SAS.</p>				
Pre-requisite(s):	Undergraduates must pass MT4607 or MT5753 or MT5761			
Anti-requisite(s)	You cannot take this module if you take MT5757			
Learning and teaching methods of delivery:	Weekly contact: 2.5 hours of lectures lectures (Weeks 1 - 10) and 8 practicals over the semester.			
Assessment pattern:	2-hour Written Examination = 60%, Coursework = 40%			
Re-assessment pattern:	2-hour Written Examination = 100%			
Module teaching staff:	TBC			

Mathematics & Statistics - Data Intensive Analysis - 2018/9 - August 2018

ID5059 Knowledge Discovery and Datamining				
SCOTCAT Credits:	15	SCQF Level 11	Semester	2
Academic year:	2018/9			
Planned timetable:	11.00 am Mon (odd weeks), Wed and Fri			
Contemporary data collection can be automated and on a massive scale e.g. credit card transaction databases. Large databases potentially carry a wealth of important information that could inform business strategy, identify criminal activities, characterise network faults etc. These large scale problems may preclude the standard carefully constructed statistical models, necessitating highly automated approaches. This module covers many of the methods found under the banner of Datamining, building from a theoretical perspective but ultimately teaching practical application. Topics covered include: historical/philosophical perspectives, model selection algorithms and optimality measures, tree methods, bagging and boosting, neural nets, and classification in general. Practical applications build sought-after skills in programming (typically R, SAS or python).				
Learning and teaching methods of delivery:	Weekly contact: Lectures, seminars, tutorials and practical classes.			
Assessment pattern:	2-hour Written Examination = 60%, Coursework = 40%			
Re-assessment pattern:	2-hour Written Examination = 60%, Existing Coursework = 40%			
Module coordinator:	Dr T W Kelsey			
Module teaching staff:	Dr T Kelsey, Dr R Hoffmann			

One of:

CS5099 Dissertation in Computer Science				
SCOTCAT Credits:	60	SCQF Level 11	Semester	Full Year
Academic year:	2018/9			
Planned timetable:	To be arranged.			
This module is an individually supervised MSc project on a topic in Computer Science. It results in a dissertation of no more than 15,000 words. Typically the dissertation comprises a review of related work, the extension of old or development of new ideas, software implementation and testing, analyses and evaluation. Students are required to give a presentation of their work.				
Pre-requisite(s):	Requires admission to dissertation phase of msc and permission of the head of school			
Anti-requisite(s)	You cannot take this module if you take CS5098			
Learning and teaching methods of delivery:	Weekly contact: Meeting with supervisor.			
	Scheduled learning: 0 hours		Guided independent study: 0 hours	
Assessment pattern:	As used by St Andrews: Coursework = 100%			
Module teaching staff:	TBC Module coordinator(s): Director of Postgraduate Teaching - Computer Science (dopgt-cs@st-andrews.ac.uk)			

Or:

MT5099 Dissertation for MSc Programme/s			
SCOTCAT Credits:	60	SCQF Level 11	Semester Full Year
Academic year:	2018/9		
Planned timetable:	At times to be arranged with the supervisor.		
Student dissertations will be supervised by members of the teaching staff who will advise on the choice of subject and provide guidance throughout the progress of the dissertation. The completed dissertation must be no more than 15,000 words.			
Learning and teaching methods of delivery:	Weekly contact: Individual supervision		
	Scheduled learning: 0 hours	Guided independent study: 0 hours	
Assessment pattern:	As used by St Andrews: Dissertation = 100%		
Re-assessment pattern:	No Re-Assessment Available		
Module coordinator:	Dr J D Mitchell		

Optional modules are available - see the pdf online called Computer Science optional modules 2018-2019

