

Masters in Data-Intensive Analysis

Programme Requirements

Taught Element, and PG Diploma in Data-Intensive Analysis:

120 credits:

- MT4113
- MT5753
- MT5756
- MT5757
- ID5059
- up to 30 credits from CS5001, CS5002, CS5003, CS5044, CS5052

MSc:

120 credits from Taught Element plus CS5099 **or** MT5099

For all Masters degrees there are exit awards available that allow suitably-qualified candidates to receive a Postgraduate Certificate or Postgraduate Diploma.

Compulsory modules:

MT4113 Computing in Statistics				
SCOTCAT Credits:	15	SCQF Level 10	Semester:	1
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.				
Programme module type:	Compulsory for Data-Intensive Analysis MSc programme.			
Learning and teaching methods and delivery:	Weekly contact: 1.5-hour lectures (x 10 weeks), 2-hour practical classes (x 10 weeks)			
Assessment pattern:	Written Examination = 40% (2 x 50-minute class tests), Coursework = 60%			
Module Co-ordinator:	Dr E Rexstad			
Lecturer(s)/Tutor(s):	Dr E Rexstad, Dr L Thomas			

Computer Science - Data-Intensive Analysis MSc - 2016/7 - August 2016

MT5753 Statistical Modelling				
SCOTCAT Credits:	20	SCQF Level 11	Semester:	1
Planned timetable:	2.00 pm			
<p>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.</p>				
Programme module type:	Compulsory for Applied Statistics and Datamining MSc Programme. Compulsory for Data-Intensive Analysis MSc Programme. Optional for Statistics MSc Programme.			
Anti-requisite(s):	MT4607	Required for:	MT5757	
Learning and teaching methods and delivery:	Weekly contact: 6 hours lectures, 1.5 hours tutorials and 6 hours practicals (x 5 weeks).			
Assessment pattern:	2-hour Written Examination = 50%, Coursework = 50%			
Module Co-ordinator:	Dr M L MacKenzie			
Lecturer(s)/Tutor(s):	Dr M L MacKenzie, TBC			

MT5756 Data Analysis				
SCOTCAT Credits:	20	SCQF Level 11	Semester:	1
Planned timetable:	2.00 pm			
<p>This module provides coverage of essential statistical concepts, data manipulation and analysis methods, and software skills in commercial analysis packages. Specifically: the different types of data and their numerical/graphical treatment; data entry/import/export, basic probability theory and concepts of inference; fundamental statistical concepts with particular emphasis on sampling issues; basic statistical models and tests; introductory computer-intensive inference. The widespread commercial statistical packages SAS, SPSS are introduced and utilised with Excel for most analyses. The statistical programming language R is also given brief attention. 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.</p>				
Programme module type:	Compulsory for Applied Statistics and Datamining MSc Programme. An exemption may be granted to an appropriately qualified entrant. Compulsory for Data-Intensive Analysis MSc Programme.			
Learning and teaching methods and delivery:	Weekly contact: Lectures, tutorials and practicals for 4 weeks.			
Assessment pattern:	2-hour Written Examination = 60%, Coursework = 40%			
Module Co-ordinator:	Dr C R Donovan			
Lecturer(s)/Tutor(s):	Dr C R Donovan			

Computer Science - Data-Intensive Analysis MSc - 2016/7 - August 2016

MT5757 Advanced Data Analysis				
SCOTCAT Credits:	20	SCQF Level 11	Semester:	2
Planned timetable:	12.00 noon Mon (even weeks), Tue and Thu			
<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>				
Programme module type:	Compulsory for Applied Statistics and Datamining MSc Programme. Compulsory for Data-Intensive Analysis MSc Programme. Optional for Statistics MSc Programme.			
Learning and teaching methods and delivery:	Weekly contact: 2.5 lectures (weeks 1 - 10) and 8 tutorials over the semester.			
Assessment pattern:	2-hour Written Examination = 60%, Coursework = 40%			
Module Co-ordinator:	Dr M L MacKenzie			
Lecturer(s)/Tutor(s):	Dr M L MacKenzie, Dr H Worthington			

ID5059 Knowledge Discovery and Datamining				
SCOTCAT Credits:	15	SCQF Level 11	Semester:	2
Planned timetable:	11.00 am Mon (odd weeks), Wed and Fri			
<p>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 the commercial packages SAS and SPSS.</p>				
Programme module type:	Compulsory for Applied Statistics and Datamining Postgraduate Programme. Compulsory for Data-Intensive Analysis MSc Programme. Optional for all Postgraduate Programmes.			
Anti-requisite(s):	MT5759			
Learning and teaching methods and delivery:	Weekly contact: Lectures, seminars, tutorials and practical classes.			
Assessment pattern:	2-hour Written Examination = 60%, Coursework = 40%			
Module Co-ordinator:	masters-coord-cs@st-andrews.ac.uk			

Compulsory module for MSc:

CS5099 Dissertation in Computer Science				
SCOTCAT Credits:	60	SCQF Level 11	Semester:	Summer
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.				
Programme module type:	Optional for MSc in Advanced Computer Science, in Artificial Intelligence, in Data-Intensive Analysis, in Human Computer Interaction, in Networks and Distributed Systems, and Software Engineering Postgraduate Programmes.			
Pre-requisite(s):	Admission to dissertation phase of MSc and permission of the Head of School	Anti-requisite(s):	CS5098	
Learning and teaching methods and delivery:	Weekly contact: Meeting with supervisor.			
Assessment pattern:	Coursework = 100%			
Module Co-ordinator:	masters-coord-cs@st-andrews.ac.uk			

OR

MT5099 Dissertation for MSc Programme/s				
SCOTCAT Credits:	60	SCQF Level 11	Semester:	Whole Year
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 of not more than 15,000 words must be submitted by the end of August.				
Programme module type:	Compulsory for MSc Programmes in Mathematics and in Statistics. Optional for Data-Intensive Analysis MSc Programme.			
Learning and teaching methods and delivery:	Weekly contact: Individual supervision			
Assessment pattern:	Dissertation = 100%			
Module Co-ordinator:	Dr J D Mitchell			

Optional modules:

CS5001 Object-Oriented Modelling, Design and Programming				
SCOTCAT Credits:	15	SCQF Level 11	Semester:	1
Planned timetable:	Variable			
This module introduces and revises object-oriented modelling, design and implementation up to the level required to complete programming assignments within other MSc modules. Students complete a number of practical exercises in laboratory sessions.				
Programme module type:	Compulsory for Advanced Computer Science, Artificial Intelligence, Human Computer Interaction, Networks and Distributed Systems, Software Engineering and Erasmus Mundus Dependable Software Systems Postgraduate Programmes. Either CS5001 or CS5002 is compulsory for Computing and Information Technology Postgraduate Programmes. Optional for Data-Intensive Analysis or Management and Information Technology Postgraduate Programmes.			
Anti-requisite(s):	CS5002	Required for:	CS5011, CS5021, CS5031	
Learning and teaching methods and delivery:	Weekly contact: Lectures, tutorials and practical classes.			
Assessment pattern:	Coursework = 100%			
Module Co-ordinator:	masters-coord-cs@st-andrews.ac.uk			

CS5002 Programming Principles and Practice				
SCOTCAT Credits:	15	SCQF Level 11	Semester:	1
Planned timetable:	Variable			
This module introduces computational thinking and problem solving skills to students who have no or little previous programming experience. It covers general programming concepts used in the development of software applications, such as data structures, functions, choice, iteration, recursion and input/output. An easy-to-learn programming language is used to illustrate these concepts, and programming skills are reinforced through practical assignments.				
Programme module type:	Either CS5001 or CS5002 is compulsory for Computing and Information Technology Postgraduate Programmes. Optional for Data-Intensive Analysis or Management and Information Technology Postgraduate Programmes.			
Anti-requisite(s):	CS5001	Required for:	CS5003	
Learning and teaching methods and delivery:	Weekly contact: Lectures, tutorials and practical classes.			
Assessment pattern:	Coursework = 100%			
Module Co-ordinator:	masters-coord-cs@st-andrews.ac.uk			

CS5003 Masters Programming Projects				
SCOTCAT Credits:	15	SCQF Level 11	Semester:	2
Planned timetable:	Variable			
This module reinforces key programming skills gained in CS5002, by means of a series of coursework assignments posed as small programming projects. These are designed to offer increasing depth and scope for creativity as the module progresses.				
Programme module type:	Compulsory for Computing and Information Technology Postgraduate Programme. Optional for Advanced Computer Science, Artificial Intelligence, Data-Intensive Analysis, Dependable Software Information Technology, Human Computer Interaction MSc Programmes, DEEng in Computer Science			
Pre-requisite(s):	CS5002	Anti-requisite(s):	IS5108	
Learning and teaching methods and delivery:	Weekly contact: Lectures, tutorials and practical classes.			
Assessment pattern:	Coursework = 100%			
Module Co-ordinator:	masters-coord-cs@st-andrews.ac.uk			

CS5044 Information Visualisation and Visual Analytics				
SCOTCAT Credits:	15	SCQF Level 11	Semester:	2
Planned timetable:	To be arranged.			
This module provides an introduction to information visualisation and visual analytics. It focuses on the question of how to utilise visual representations to make information accessible for exploration and analysis. The module covers basic principles of visualisation design and interaction principles. It introduces a range of visualisation techniques and tools, and discusses how these can be effectively applied in various scenarios for communication, exploration and analysis, and how to evaluate information visualisations in different contexts. Skills in designing, developing, and evaluating information visualisations are reinforced through practical assignments. There are no pre-requisites for this module but students should have basic programming skills (e.g. in Java or JavaScript).				
Programme module type:	Optional for all Postgraduate Programmes			
Learning and teaching methods and delivery:	Weekly contact: 3-hour lecture (x 11 weeks), 1-hour seminar (x 8 weeks)			
Assessment pattern:	2-hour Written Examination = 40%, Coursework = 60%			
Module Co-ordinator:	masters-coord-cs@st-andrews.ac.uk			

Computer Science - Data-Intensive Analysis MSc - 2016/7 - August 2016

CS5052 Data-Intensive Systems				
SCOTCAT Credits:	15	SCQF Level 11	Semester:	2
Planned timetable:	Mon, Tue (time to be confirmed)			
<p>The era of big data is upon us - the volume, velocity and variety of enterprise and scientific data are growing at an exponential rate and will continue to do so for the foreseeable future. This module presents the programming paradigms, algorithmic techniques and design principles for large-scale distributed systems, such as those utilised by companies such as Google, Amazon and Facebook. This module is different in scope from CS4103 (distributed systems) as it focuses primarily on building and utilising large-scale clusters.</p> <p>The module will cover: distributed systems architecture, replication and fault tolerance, storage, coordination, scheduling algorithms, cluster computing, cloud computing, virtualisation, programming models (e.g., MapReduce), stream processing, decentralised systems (e.g., Chord), incentive-based systems (e.g., BitTorrent), and social computing (e.g., crowd sourcing techniques). This module will draw from the latest research in both academia and industry.</p>				
Programme module type:	Optional for Data-Intensive Analysis MSc Programme. Optional for Networks and Distributed Systems MSc			
Pre-requisite(s):	CS5001			
Learning and teaching methods and delivery:	Weekly contact: 3 hours of lectures (x 11 weeks), 1-hour seminar (x 4 weeks), 1-hour practical class (x 3 weeks)			
Assessment pattern:	2-hour Written Examination = 60%, Coursework = 40%			
Module Co-ordinator:	Dr A Barker			
Lecturer(s)/Tutor(s):	Dr A Barker			

