

Data
Science

Lancaster
University



MSc in Data Science

Course Handbook 2015-16

<http://www.lancaster.ac.uk/data-science/>

Computing &
Communications



Mathematics
and Statistics



Environment
Centre



Section	Contents
Preface	University Core Information for Students
1	Welcome to Data Science at Lancaster
2	Overview of Data Science at Lancaster
3	Data Science: Computing Specialism
4	Data Science: Statistical Inference Specialism
5	Submission of Coursework and Feedback
6	Student Feedback Mechanisms
7	Programme Rules and Requirements for Awards
8	Student Support, Advice and Facilities
Appendix A	Module Descriptions
Appendix B	Term Dates

Core Information for Lancaster University PG Students

This course handbook has been compiled to assist MSc Data Science students studying at Lancaster University. More generally, all core university-level postgraduate information is available from the web page:

<http://www.lancaster.ac.uk/current-students/>

and can be accessed from 'Core Information for Students' categorised under the respective student groupings:

Taught Postgraduate Core Information

The screenshot shows the Lancaster University website interface. At the top, there is a navigation menu with links for Home, Study, Research, About Us, Business, and Alumni. Below the menu is a banner with the Lancaster University logo and statistics: "12,000 We have over 12,000 students, from over 100 countries, within one of the safest campuses in the UK" and "97% 97% of Lancaster students go into work or further study within six months of graduating". A search bar is located on the right side of the banner.

The main content area features a "Current Students" heading with a sub-heading: "This page contains links to content and services for existing students. Some items are only available to logged-in users." Below this is a "Latest News" section titled "Top 10 for student satisfaction" with a "Learn more" link.

At the bottom of the page, there is a "Core Information for Students" section with a grid of links. An arrow from the grey box above points to the "Taught Postgraduates" link in this section. Other links include "Student Portal", "Undergraduate Email", "Postgraduate Email", "Term Dates", "Exam Timetables", "Library", "Undergraduates", "Postgraduate Research Students", "Study-Related", "Academic Departments & Faculties", "Complaints and Appeals", "Student Services", "Accommodation", "Careers", "Counselling Service", "Equality & Diversity Information", "Faults Reporting (Facilities)", "Graduation Ceremonies", "LancasterAnswers", "Online Documents", "Student Activities", "Lancaster University Students' Union", "Student Enterprise", "Student Societies", "Campus Facilities", "Campus Map", and "Commuting to Campus".

The footer of the page displays "Lancaster University" and the date "10/09/2015".

1. Welcome to Data Science at Lancaster

Welcome on board the MSc in Data Science!

This is the second year that we have run MSc offerings in the expanding field of Data Science, and each variety of the degree has been crafted to give students a challenging, yet rewarding, experience that prepares them for the growing job market surrounding Data Science.

In formulating the degree schemes we have drawn on our world-leading research and teaching expertise in Mathematics and Statistics, Data Mining, and Environmental Science to provide you with a cutting-edge curriculum. This will include the use of state-of-the-art big data technologies and the usage of data from a variety of domains (e.g. social media, health, ecology, business etc.). When designing the curriculum, we also sought guidance from businesses and research organisations, so you will be exposed to real-world data science problems, several of whom will provide students placements.

Ultimately the degrees are about your experience, which we hope will be challenging, yet fulfilling. To enhance this, we have provided virtual learning environments through which you can network with your peers and ask questions of your tutors and lecturers. Should you have any issues related to your studies then the course directors and support staff will be on hand to help you.

This handbook provides an overview of the courses structure and modules that are offered as part of the MSc in Data Science. Coursework submissions have been specified so as to spread your work load throughout the year, however, should you be struggling with any aspect of the work, or need more time, then please contact one of the directors as soon as possible.

We are incredibly excited about this coming year, the curriculum that you will be taught and the opportunities that your studies will provide.

Dr Chris Edwards; Dr Matthew Rowe; Dr Deborah Costain; Dr Ian Hartley



Dr. Chris Edwards
c.edwards@lancaster.ac.uk
+44 (0)1524 510329

Dr. Matthew Rowe
m.rowe@lancaster.ac.uk
+44 (0)1524 510310

Dr. Deborah Costain
d.costain@lancaster.ac.uk
+44 (0)1524 594666

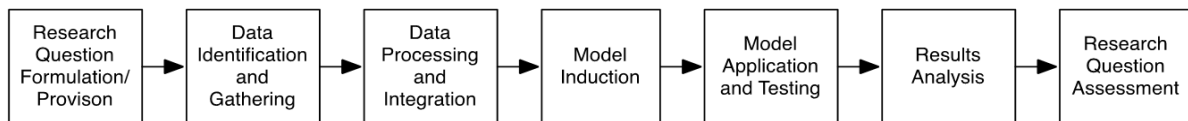
Dr. Ian Hartley
i.hartley@lancaster.ac.uk
+44 (0)1524 592566

2. Overview of Data Science at Lancaster

2.1 Data Science at Lancaster

Data are now being generated at an unprecedented rate and scale. Whether they come from environmental sensors, social media streams, businesses, or public sector organisations, there is a need to process these data sets and extract information, thereby allowing business and research questions to be answered. This is the job of the data scientist whose role involves: research question formation, gathering and synthesis of data, data mining and visualisation, statistical modelling and inference, forecasting and prediction and dissemination.

A data scientist is thus a highly skilled individual with the ability to: articulate a research question; gather, process and model data at large-scale (developing scalable algorithms for performing inference and modelling complex and heterogeneous data structures) and to disseminate research findings in context. This process is sequential, often cyclic in nature, and flows to form the data science pipeline:



The Lancaster programme combines interdisciplinary teaching from world-leading departments: the School of Computing and Communications (SCC); the Department of Mathematics and Statistics and the Lancaster Environment Centre (LEC) and Lancaster University Management School (LUMS).

Building upon a common ‘core’ set of modules the Data Science Masters schemes allow for subject specialism:

- **MSc: Data Science**
 - Statistical Inference pathway
 - Computing pathway
- **MSc: Data Science for the Environment.**

The *Computing specialism* of the Data Science MSc is aimed at students with a background in computer science who want to develop strong quantitative, data handling and analytical skills, along with skills that support the engineering of systems to support data science. The *Statistical Inference* specialism of the Data Science MSc is aimed at students with a background in mathematics and statistics and who want to develop their statistical, computing and analytical skills for the extraction, synthesis, processing and analysis of large and complex data.

Note that, subject to module pre-requisites, students may choose modules across specialisms in order to fit their background and career aspirations. Course details, by specialism, are provided in Section 3 and Section 4.

The production of data that are relevant to environmental issues is increasing at a phenomenal rate. Technological advances in fields such as remote sensing, imaging and forecasting are increasingly producing data from more sources and with higher resolution than we have been able to handle until recently. There is a need for data scientists who are able to organise, analyse and disseminate such data in order to answer questions which will help us manage a range of scenarios, from management of natural resources to predicting population change. This is the aim of the MSc: *Data Science for the Environment*.

“This new initiative represents a significant investment in post graduate teaching at Lancaster, positioning us among the world leaders in the education of highly qualified people with the relevant skills necessary to make a real impact in the world, in fields from big business and engineering to big data and environmental science”.

Professor Mark Smith, Vice-Chancellor, 2014

2.2 Admissions Criteria

We expect to recruit students who are interested in a career at the interface of computing, statistics and their application.

For the MSc in Data Science you should hold, or expect to obtain, a BSc in Mathematics and, or, Statistics or a BSc in Computing. For the MSc in Data Science for the Environment you should hold, or expect to obtain, a BSc in Ecology or Environmental Sciences. Admission to the Master's programme requires a minimum of a 2:1.

For Data Science for the Environment, A' level mathematics is the expected standard.

For applicants whose first language is not English a recognised English qualification is required:

- IELTS: 6.5 (with at least 6.0 in each skill).

Overseas students will require a visa to be able to study with us in the UK.

2.3 Making an Application

You should apply online using the [My Applications website](#)

Supporting documentation includes:

- Your degree transcripts and certificates, including certified English translations if applicable
- Two references
- If English is not your first language, you should also enclose copies of your English language test results

You also need to complete a personal statement to help us understand why you wish to study your chosen degree. You should also state your intended specialism/focus.

If you are a current Lancaster University student, you will not need to provide your Lancaster degree transcript and only need to provide one reference.

The postgraduate section of Lancaster University's website has plenty of [Information](#) about applying to Lancaster

3. MSc Data Science: Computing Specialism

Entry Requirements: 2:1 degree in a subject relevant to computer science.

Key Contact: Dr Matthew Rowe, School of Computing and Communications: m.rowe@lancaster.ac.uk

3.1 Computing pathway: overview

Underpinning the data scientist role are the technologies that enable the processing of data at large-scale, often using parallel processing paradigms.

The Computing specialism provides the training to understand how these technologies function and how they are implemented within both enterprise and research environments. Students will get hands-on experience of building, from scratch, large-scale systems that enable data science questions to be answered, using technologies such as Hadoop, Spark, Giraph, and HBase.

The Computing specialism consists of a series of **taught modules** (120 credits) followed by the completion of a **dissertation** (60 credits). The taught course component consists of 8 modules and can be decomposed as follows:

A '**core**' set of five compulsory modules (**75 credits**) spanning modern data science fundamentals:

- SCC460 Data Science Fundamentals (Michaelmas term & Lent term)
- SCC403 Data Mining (Michaelmas term)
- SCC461 Programming for Data Scientists (Michaelmas term)
- CFAS440 Statistical Methods and Modelling (Michaelmas term)
- CFAS406 Statistical Inference (Michaelmas term)

A set of three '**compulsory**' specialism specific modules (**45 credits**):

- SCC401 Elements of Distributed Systems (Lent)
- SCC413 Applied Data Mining (Lent)
- SCC411 Systems Architecture and Integration (Lent)

All module descriptions are, in alpha-numeric order, contained in Appendix A. Modes of assessment and credit weightings for all modules follow in Table 1, Section 3.2.

The dissertation component consists of a substantial project applying data scientific methods to address a substantive research question. This will be undertaken during the summer term and will typically incorporate a two-month industrial or research organisation placement.

Note that the above is the designated Computing pathway, however, subject to module pre-requisites, students can tailor their programme of study to reflect expertise and aspirations.

3.2 Computing specialism: course scheduling and assessment

Course Scheduling

The taught courses run during weeks 1 to 20. Core modules are taught in weeks 1 to 10 (Michaelmas term), except for Data Science fundamentals which follows its own structure over both terms 1 and 2. Compulsory specialism specific modules are taught in weeks 11 to 20 (Lent term). Term dates for 2015/16 are provided in Appendix B.

Core Module Assessment

All Computing specialism modules are assessed via 100% coursework (Section 3.2, Appendix A). Marks are ratified by the Board of Examiners which meets in October. Credit for a module is given if the overall module mark is 50% or more.

Dissertation Assessment

The dissertation component carries a weighting of 60 credits. Each dissertation will be double-marked, and a provisional mark agreed between the two markers. A copy of each dissertation and a brief report (including a provisional mark) agreed between the two internal markers will then be sent to the external examiner in advance of the final meeting of the Board of Examiners in October. The deadline for submission of the dissertation is Friday 9th September.

Table 1. Computing specialism: course structure, module weightings and mode of assessment

1. A set of five 'core' modules (75 credits) in common with Statistical Inference specialism and the Data Science for the Environment programme except that the statistics modules studied are tailored according to background

Module number	Title	Weeks scheduled	Coursework %	Exam %	Credit Weighting
SCC460	Data Science Fundamentals	1 - 20	100	NA	15
SCC403	Data Mining	1 - 10	100	NA	15
SCC461	Programming for Data Scientists	1 - 10	100	NA	15
CFAS440	Statistical Methods and Modelling	1 - 5	100	NA	15
CFAS406	Likelihood Inference	6 - 10	100	NA	15

2. A set of three 'compulsory' (45 credits) computing specialism specific modules:

Module number	Title	Weeks scheduled	Coursework %	Exam %	Credit Weighting
SCC401	Elements of Distributed Systems	11 - 13	100	NA	15
SCC413	Applied Data Mining	14 - 20	100	NA	15
SCC411	Systems Architecture and Integration	14 - 20	100	NA	15

SCC modules: taught by School of Computing and Communications except for SCC460 and SCC461 which are cross-disciplinary; CFAS modules: taught by the Department of Mathematics and Statistics

3. A dissertation (60 credits) with an associated industrial placement/research organisation.

4. MSc Data Science: Statistical Inference Specialism

Entry Requirements: 2:1 in Mathematics and/or Statistics.

Key Contact: Dr Deborah Costain, Mathematics and Statistics: d.costain@lancaster.ac.uk

4.1 Statistical Inference specialism overview

The Statistical Inference specialism is aimed at students with a background in mathematics and statistics and who want to develop their statistical, computing and analytical skills for the extraction, synthesis, processing and analysis of large and complex data.

The programme encompasses data science fundamentals but with additional focus upon statistical modelling and inference of large and complex data structures and the underpinning theories. More specifically, the course provides a thorough 'core' training in statistical theory; data analysis and computing via a distinctive blend of leading-edge methodology and practical techniques (including Bayesian and computational methods and data mining) and 'optional' modules spanning, for example, genomics, longitudinal data analyses, time-to-event data, extremal analysis, spatial data analyses and forecasting. The modules reflect inter-departmental research expertise and will prepare students for particular career options in areas with a growing demand for data scientists.

The Statistical Inference specialist pathway of the MSc in Data Science consists of a series of **taught modules** (120 credits) followed by the completion a **dissertation** (60 credits). The taught course component consists of nine modules which can be decomposed as follows:

A '**core**' set of five compulsory modules (**75 credits**) spanning modern data science fundamentals:

- SCC460 Data Science Fundamentals (Michaelmas & Lent term)
- SCC403 Data Mining (Michaelmas term)
- SCC461 Programming for Data Scientists (Michaelmas term)
- MATH552 Generalised Linear Modelling (Michaelmas term)
- MATH551 Likelihood Inference (Michaelmas term)

A '**compulsory specialist**' module (**15 credits**) in Bayesian Inference for Data Science (MATH555);

A set of three '**optional**' modules (**30 credits**) chosen from 10 available which span a range of specialist/advanced statistical methods relevant to the study design, analysis and interpretation of observational and experimental data:

- MATH562 Extreme Value Theory
- MATH563 Design and Analysis of Clinical Trials
- MATH564 Principles of Epidemiology
- CHIC565 Environmental Epidemiology
- MATH566 Longitudinal Data Analysis
- MATH572 Genomics: Technologies and Data Analysis
- MATH573 Survival and Event History Analysis
- MSCI523 Forecasting
- MSCI526 Data Mining for Marketing, Sales and Finance
- MSCI524 Optimisation & Heuristics

All module descriptions are, in alpha-numeric order, contained in Appendix A. Modes of assessment and credit weightings follow in Section 4.2, Table 2.

The dissertation component consists of a substantial project applying data scientific methods to address a substantive research question. This will be undertaken during the summer term and will typically incorporate a two-month industrial or research organisation placement.

4.2 Statistical Inference Course Structure; weightings and assessment strategy

Course Scheduling and Assessment

The taught courses run during weeks 1 to 20. Core modules are taught in weeks 1 to 10 (Michaelmas term) except for Data Science Fundamentals (SCC460) which runs over weeks 1 to 20. Optional and specialism specific modules are taught in weeks 11 to 20 (Lent term). **All Exams are held in May/June 2016.** Term dates for 2015/16 are provided in Appendix B.

Assessment

Testing of knowledge and understanding is achieved through a range of assessment methods. The credit weighting and balance between coursework and examination varies between modules, as shown in Section 4.3. Marks are ratified by the Board of Examiners which meets in June and October. Credit for a module is given if the overall module mark is 50% or more.

End of Module Tests

For Math551 (Likelihood Inference) and Math552 (Generalised Linear Models) 30% of the course work assessment (i.e. 15% of total module mark) will be the result of a progress test. **Tests are typically held during the tutorial session in the final week of the module. The aim is to consolidate module material.**

Exam Timetable

All examinations for modules studied in academic year 15/16 will be timetabled during the period May 2016 and June 2016. Provisional exam results will be made available to students in late June 2016. Final results will be communicated by post by late November 2016.

Dissertation Assessment

The dissertation component carries a weighting of 60 credits at level 7. Each dissertation will be double-marked, and a provisional mark agreed between the two markers. You will be asked to attend a viva in person or via Skype to defend your work. A copy of each dissertation and a brief report (including a provisional mark) agreed between the two internal markers will then be sent to the external examiner in advance of the final meeting of the Board of Examiners in October. The deadline for submission of the dissertation is Friday 9th September 2016.

Table 2. Assessment arrangements for individual taught modules

1. A set of five 'core' modules (75 credits) in common with the Computing specialism and the Data Science for the Environment programme except that the statistics modules are tailored according to background.

Module	Title	Weeks Scheduled	Coursework	Exam %	Credit Weighting
SCC460	Data Science Fundamentals	1 – 20	100	NA	15
SCC403	Data Mining	1 – 10	100	NA	15
SCC461	Programming for Data Scientists	1 – 10	100	NA	15
MATH551	Likelihood inference	1 – 5	50	50	15
MATH552	Generalised linear models	1 – 5	50	50	15

2. A 'compulsory' Bayesian Inference module (15 credits)

Module	Title	Weeks Scheduled	Coursework	Exam %	Credit Weighting
MATH555	Bayesian inference for Data Science	11 – 20	50	50	15

3. Three 'self-selected' optional modules (30 credits) from:

Module	Title	Weeks Scheduled	Coursework %	Exam %	Credit Weighting
MATH562	Extreme value theory	11 – 12	50	50	10
MATH563	Clinical trials	11 – 12	50	50	10
MATH564	Principles of epidemiology	13 – 14	50	50	10
CHIC565	Environmental epidemiology	19 – 20	50	50	10
MATH566	Longitudinal data analysis	15 – 16	50	50	10
MATH572	Genomics: technologies and data analysis	11 – 20	100	NA	10
MATH573	Survival and event history analysis	17 – 18	50	50	10
MSCI523	Forecasting	11 – 20	100	NA	10
MSCI526	Data Mining for Marketing, Sales and Finance	11 – 20	100	NA	10
MSCI534	Optimisation & Heuristics	11 – 20	30	70	10

SCC modules taught by School of Computing and Communications; MATH modules: taught by the Department of Mathematics and Statistics; MSCI modules taught by Lancaster University Management School; CHIC taught by CHICAS research group.

4. A dissertation (60 credits) with an associated industrial / research organisation placement.

5. Submission of Coursework and Feedback

Each module has its own procedure for handling coursework and the module tutor will provide information to you regarding the submission of coursework. One paper copy is submitted and also an electronic copy is uploaded online through Moodle (<https://modules.lancs.ac.uk/>). This allows us to check all coursework electronically for plagiarism. Please ensure that you check and are aware of the coursework deadlines and submission procedures. Coursework **MUST NOT** be handed directly to the module tutors since all coursework submissions need to be recorded by the administrating PG Office.

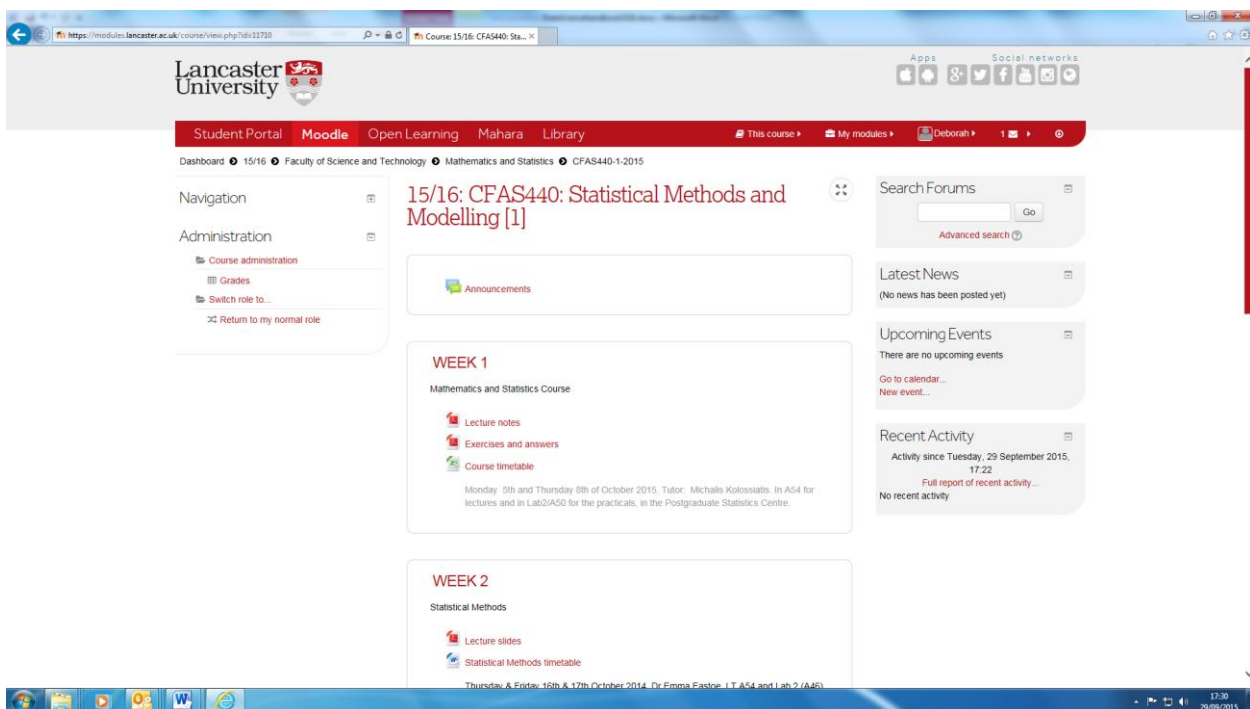
All hard-copy coursework assignments are to be submitted to the **relevant** Masters Submission Box located administrating department:

1. **SCCXXX modules: School of Computing and Communications** Data Science submission box (contact Debbie Stubbs d.stubbs@lancaster.ac.uk)
- **MATHXXX/CFASXXX modules: Mathematics and Statistics Dept.** Data Science submission box (contacts MATHXXX: Jane Hall j.hall2@lancaster.ac.uk; CFASXXX Angela Mercer a.j.mercer@lancaster.ac.uk)
2. **LECXXX: Lancaster Environment Centre** Master's submission box

Precise details will be provided by the module tutor. An assignment cover sheet must signed and attached to the front of your coursework to confirm that the work is your own.

Submitting your assignments online

Each module you take will have its own Moodle page. You can view all of you Moodle page access by accessing <https://modules.lancs.ac.uk/> or logging into your student portal <https://portal.lancs.ac.uk/>. Your assignments should be submitted online via Moodle. The example below illustrates this.



The screenshot shows a Moodle course page for '15/16: CFAS440: Statistical Methods and Modelling [1]'. The page is viewed through a web browser with the URL <https://modules.lancaster.ac.uk/course/view.php?id=11730>. The page features a navigation menu on the left with options like 'Course administration', 'Grades', and 'Return to my normal role'. The main content area displays 'WEEK 1' and 'WEEK 2' sections, each containing links for 'Lecture notes', 'Exercises and answers', and 'Course timetable'. A search bar and 'Search Forums' section are visible on the right. The page is branded with the Lancaster University logo and includes social media icons. The bottom of the page shows a Windows taskbar with the date 29/09/2015 and time 17:30.

Note all clickable links are highlighted in red. Files etc. can be downloaded and coursework submissions uploaded etc.

5.2 Late submission of assignments and penalties

There are strict rules in place for the late submission of assignments;

- **unsanctioned late work submitted up to 3 days after the deadline will incur a penalty of 10% of the available marks for the work**
- **work submitted after 3 days, without an agreed extension, will be marked at 0%.**

5.3 Plagiarism

The University has established an institutional framework for dealing with plagiarism: [Plagiarism Framework Secretariat](#); and is a member of the JISC Plagiarism Detection Service (Turnitin), which searches for matching text between a paper and available material on the Internet. The LEC PG Office reserves the right to assess some or all of your work submitted electronically using this service.

You commit plagiarism if you try to pass off the work of any other person, whether a published author, an internet source or a fellow-student, as your own. To copy passages, illustrations or even sentences from a book without acknowledgement is gross plagiarism; to copy from another student's essay or dissertation is a particularly grave offence. Never use another's actual words (or a close paraphrase of them) without putting them in quotation marks and giving an exact reference to your source.

Serious plagiarism may result in a zero mark, and if repeated - or if particularly severe - may result in your being excluded from further courses within the University (this matter would be decided by the Standing Academic Committee). These strict rules are in place to ensure that the work you undertake as part of your course has real value to you both educationally and in giving you a sense of personal achievement.

Feedback and Notification of Marks

Feedback on assessed work will typically be provided within four weeks of submission (excluding vacations and unforeseen staff absences). Once coursework has been marked and the marks recorded by the PG Office, students will be informed and can then collect coursework from the administrating PG Office.

Exam papers are marked in line with university timescales. It is University policy not to return examination scripts to students.

Students may also view their coursework marks via the 'Student Portal' once they have been processed by the administrating PG Office.

It should be remembered that until the External Exam Board has met, (October each year) any marks given to students are provisional and may be subject to change. The External Exam Board does not usually meet until 6 weeks after the end of the programme.

As per the [University regulations](#) there is no appeal against academic judgement.

6. Student Feedback Mechanisms

Module Evaluation

You will be contacted by email at the end of each module and asked to complete a module evaluation form online. Evaluation is more than feedback of the good and bad elements of a module. It provides continual information for us to improve the modules we offer. Consequently it is very important that all students actively participate in the evaluation process. We ask that you respond to the request to submit your feedback as soon as possible. **Please note that feedback is anonymous.**

Student Representatives

Student representatives are invited to sit on the departmental PGT Staff-Student Consultative Committees. This is the official mechanism for communication between PGT students and the PG team. The committees are formed in accordance with University Regulations, which require staff-student consultation prior to programme changes.

Two representatives will be required, one from each of the Data Science specialisms. Volunteers will be invited to express an interest at the beginning of the academic year.

The committees meet at least once per term as a consultative body to discuss current programmes and future proposed changes. Dates for the Committees and the minutes of each meeting will be made available on Moodle by the Course Coordinators.

7. Programme Rules and Requirements for Awards

Module Rules

The pass mark for taught Masters degrees is 50% per module, with credit for a module being awarded when the overall mark for the module is 50% or greater. The mark for each unit module is derived from a stipulated combination of written examination and/or coursework (Table 1, Table 2).

Awards

The overall mark awarded for the taught courses is calculated from the average of the marks gained in the taught modules, weighted for the credit rating of each module. The overall mark awarded for the dissertation is taken as the average mark of the first and second markers. The following criteria will be applied for the awards of Masters degrees:

Award of MSc in Data Science

To qualify for the Masters' degree in Data Science you must achieve a total of 180 credits: 120 credits from the taught course component and 60 credits from the Dissertation. Credit for a module is given if the overall module mark is 50% or more.

Condonation

Notwithstanding this requirement, candidates shall be eligible for an award by compensation/condonation in respect of up to a maximum of 45 credits of a taught Masters programme provided that:

- a) no single module mark falls below 40%;
- b) the candidate's weighted mean or modal mark is 50% or greater;

Higher awards:

MSc in Data Science with 'Merit'

In addition to the requirements for the MSc in Statistics an 'MSc in Statistics with Merit' shall be awarded when the weighted average mark over the programme is 60% or more.

MSc in Data Science with 'Distinction'

In addition to the requirements for the award of and MSc in Statistics an 'MSc in Statistics with Distinction' shall be awarded when the average weighted mark over the programme (as detailed in 7.1.1) is 70% or more.

Borderline cases

Where the overall average falls within 2 percentage points of the range (68%, 58% or 48% respectively) or in cases where most credits are in the class above the mean, the Exam Board will have discretion to decide which of the alternative awards to recommend.

Re-sits

A student who fails to achieve a mark of 50% for a module/element in the MSc programme is entitled to one opportunity for reassessment in each failed module/element. **A mark of not more than 50% can be awarded for modules re-taken.**

The form of the reassessment is at the absolute discretion of the Examination Board, save that the form of reassessment must allow the student a realistic chance of achieving 50% in the re-sit.

Notification of Final Degree Marks

The External Exam Board meets in late October to recommend awards. Final marks will be released to students as soon as possible thereafter. Please note that the University Regulations state that written confirmation of results, provisional and final, may not be released to students who are in debt to the University.

The Board of Examiners

The Board of Examiners consists of the:

- Heads of Departments;
- Course Directors;
- Academic staff of the university who have lectured a contributory module(s) during the year in question;
- External Examiner for the MSc.

The External Examiner for the period 2015-2016 is Prof. Niall Adams, Imperial College London.

Graduation

The Postgraduate [Graduation Ceremony](#) will be in December 2016. Information regarding Graduation will be sent to you from the University Ceremonies Office.

Please note that it is essential that you keep your contact details address up-to-date in order to receive the relevant graduation mailings.

8. Student Support, Advice and Facilities

Who to go to for support & advice

If you are having problems with the course – either fitting the workload into your schedule or if you have any questions about our policies & procedures, please contact us. Support can come from a number of sources. Here's who to contact if you have problems with:

Course content: The module tutor, your fellow students or the Course Director.

Assignment submission:

- **SCC modules:** contact Debbie Stubbs d.stubbs@lancaster.ac.uk in the School of Computing and Communications teaching office
- **MATH modules:** contact Jane Hall j.hall2@lancaster.ac.uk in the Department of Mathematics and Statistics
- **CFAS modules:** contact Angela Mercer aj.mercer@lancaster.ac.uk in the Department of Mathematics and Statistics

Some useful contact numbers are:

Name	Role	Telephone No. (01524)
SCC		
Matthew Rowe	Course Director (Computing specialism)	510310
Debbie Stubbs	Postgraduate Coordinator (Computing specialism)	510515
Qiang Ni	Director of PG Teaching	510328
Maths & Statistics		
Deborah Costain	Course Director (Statistical Inference specialism)	594666
Jane Hall	Postgraduate Coordinator (Statistical Inference specialism)	593964
Angela Mercer	CFAS module coordinator	593064
General contacts		
Library	Library Enquiry desk	592516
LUSU	Lancaster University Students' Union	593765
Sports' Centre	Enquiry Desk	510600
The Base	Student Support	592525

Moodle

Each postgraduate programme is supported by its own online page which uses Moodle as its base. This system is used by students from departments all across the University so you are not alone. You should familiarise yourself with this as much as possible as it is a key resource for students on the course. To find the page for your course, go to the main Moodle site from your Student Portal. Go to <https://portal.lancs.ac.uk/>. You will need your University username and password to get access to the site.

Moodle is used for:

- Gaining access to your Module pages,
- Viewing your timetable for the year and receiving notifications about changes to the timetable,
- Asking questions to tutors or fellow students,
- Getting access to additional materials and resources,
- Accessing and submitting your coursework,
- Checking course deadlines etc.,
- Giving us feedback about the course,
- Discussing issues about work

Email

All students will be given a Lancaster University email address, in the form yourname@lancaster.ac.uk. Please note that any contact we make with you will be through your Lancaster email address and therefore you must access this e-mail account on a daily basis. Failure to check your Lancaster account regularly does not constitute an excuse for missing important information, dates etc.

Computing Facilities

There are numerous open access [PC Labs](#) located around campus. The PC labs provide a wide range of software, printers (colour and monochrome) and scanning facilities. All lab PCs are connected to the campus network and internet.

Information Systems Services (ISS) also provides other IT services to students, including [IT workshops and courses](#).

It is also possible to access University services [remotely](#) e.g. from home, or via a smart phone.

The [ISS Service Desk](#) can be contacted if you require any general computing-related assistance.

Learning Zone

[The Learning Zone](#) is located centrally on Alexandra Square and is accessible 24-7. It provides relaxed surroundings for students to work within and bookable 'pods' for meetings, presentations and group work.

University Library facilities

Lancaster University Library is a valuable reference resource. Many of the main texts for the module on your programmes are available from here. Your registration with the library should have been completed when you registered with the University.

Using the Sports Facilities on Campus

As student of the University, you are entitled to student membership of the Sports Centre for the duration of your course. Please note however, that membership runs annually according to the academic year (October – September) so, depending on when you are attending your course, it may/may not be cost-effective to apply for membership. Membership details are available from the Sports Centre itself <http://sportscentre.lancs.ac.uk/>. Details of opening times / classes etc. can be found in the leaflet in your induction pack (or ring the Sports Centre itself on 01524 510600).

The Students' Union & Student Support Office

LUSU is a body that represents all student views to the University, providing professional, academic and other advice for students. Students registering at Lancaster automatically become members of the Students' Union. There are no financial obligations associated with membership, though you can withdraw from the union if you wish, by completing an opt-out form.

You can also apply online for your NUS "Purple Card" at <http://card.lusu.co.uk/members> and then collect this from LUSU in Bowland College, next to Alexandra Square.

Academic Support for Students in the Faculty of Science and Technology

Academic Support offers a number of drop-in sessions open to all students, plus bookable courses for International students, as well as one-to-one consultations. More information about this service can be found at: http://www.lancs.ac.uk/sci-tech/academic_support/

Academic support for students in FST s provided via a faculty-based service and the Student Learning Advisor for the Faculty of Science & Technology is Robert Blake: fststudyadvice@lancaster.ac.uk. Support is provided for any student seeking advice on effective scientific writing and study practices. Additional support is available for students who are non-native speakers of English or those with learning difficulties such as dyslexia. If you have a previously identified learning difficulty, or you suspect your performance is being hampered by a problem that has previously never been diagnosed, you are encouraged to contact Academic Support.

Careers Advice

The University's [Careers Service](#), offers an extensive service tailored to your needs. Their professional staff includes specialists in careers information, employer liaison, event management and careers guidance. They work closely with other staff within the university, the Students Union, professional bodies and a broad range of national and international employers to provide a variety of opportunities to help you progress your career goals. Careers are located in the Base, just off Alexander Square.

[TARGETconnect](#) is an online system administered by Careers and provides students with access to student and graduate vacancies, details of careers events, an appointment booking system to see a careers adviser and the online careers query system. Careers information including online psychometric testing and video resources are available [online](#).

Student Based Services

[Student Based Services](#) provide information, advice and guidance covering different areas.

We hope you have an enjoyable and productive time at Lancaster, but we recognise that sometimes problems can affect your ability to study. Please do not forget that it is your responsibility to seek help if you are experiencing difficulties. The University will do whatever is possible to assist you, provided that we are aware of your problems. These may be personal, financial or academic. If you find yourself getting into difficulties we strongly urge you to consult the PG Administrator in the first instance. In addition, Student Based Services provide information, advice and guidance covering different areas of student welfare.



'The Base' is situated on A-Floor of University House in Alexandra Square and is a one-stop enquiry desk for all Student Based Services. Staff there will be able to make appointments with specialist staff where needed. Details on the various student based services can be found on the links below.

[Student Registry](#) – responsible for all regulations, policies and procedures governing your award. The Student Registry is also responsible for managing your official record, including personal details. They can provide information on many aspects of student administration.

[International Student Advisory Service](#) - advice on visa extension, rules on working in the UK, budgeting, general welfare and cultural orientation. They are the designated point of advice for immigration issues.

[Disabilities Service](#) – the University has been developing services for students in this area for over 20 years and aims to help all prospective and current students who have a disability.

[Assessment Centre](#) – Lancaster University has its own assessment Centre to identify study aids and strategies required to provide equal access to the curriculum.

Counselling and Mental Health Service - staff provide confidential and professional support on issues such as personal, family, social or academic matters over the short term, to more complex or difficult longer term problems. The service offers both appointment and drop-in sessions.

Student Funding & Financial Aid - provide information, advice and guidance on student funding and financial aid. This includes student living cost loans/grants, tuition fee loans and living costs/budgeting.

Appendix A: Module Descriptions

Module Mnemonic:	CFAS406
Module Title:	Statistical Inference
Module Convenor:	Dr Steffen Grunewalder
Assessment:	Coursework (100%)
Duration:	16 hours
Credits:	15
Term:	M1, M2

This module aims to provide an in-depth understanding of statistics as a general approach to the problem of making valid inferences about relationships from observational and experimental studies. Examples from social science and environmental science are used to illustrate this approach. The emphasis will be on the principle of Maximum Likelihood as a unifying theory for estimating parameters. The module is delivered as a combination of lectures and practical's over two days.

Topics covered will include:

- Revision of probability theory and parametric statistical models.
- The properties of statistical hypothesis tests, statistical estimation and sampling distributions.
- Maximum Likelihood Estimation of model parameters.
- Asymptotic distributions of the maximum likelihood estimator and associated statistics for use in hypothesis testing.
- Application of likelihood inference to simple statistical analyses including linear regression and contingency tables.

Learning: Students will learn through the application of concepts and techniques covered in the module by application to real data sets. Students will be encouraged to examine issues of substantive interest in these studies. Students will acquire knowledge of:

- Application of likelihood inference to simple statistical analyses including linear regression and contingency tables.
- The basic principles of probability theory.
- The properties of a statistical test.
- Maximum Likelihood as a theory for estimation and inference.
- The application of the methodology to hypothesis testing for model.
- Parameters, the analysis of contingency tables and linear regression.

and develop skills to:

- apply theoretical concepts
- identify and solve problems

Assessment: One assignment covering all aspects of the module material.

Recommended texts and other learning resources:

- Dobson, A. J. (1983). *An Introduction to Statistical Modelling*. Chapman and Hall.
- Eliason, S. R. (1993). *Maximum Likelihood Estimation: Logic and Practice*. Sage Publications.
- Pickles, A. (1984). *An introduction to likelihood analysis*. Geo Books.
- Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.

Module Mnemonic:	CFAS440
Module Title:	Statistical Methods and Modelling
Module Convenor:	Prof Brian Francis, Dr Michalis Kolossiatis, Dr Emma Eastoe
Assessment:	Coursework (100%)
Duration:	45 hours
Credits:	15
Term:	M1, M2

The aim of this module will be to address the fundamentals of statistics for those who do not have a mathematics and statistics background. The module is delivered over three intensive two-day sessions of lectures and practical's. Students will develop an understanding of the theory behind core statistical topics; sampling, hypothesis testing, and modelling. They will also be putting this knowledge into practice, by applying it to real data to address research questions.

The module is an amalgamation of three two-day short courses taught in weeks 1 - 4:

- Mathematics for Statistics; the nature of variables, scientific notation, logarithms, combinations, algebra, matrices and an introduction to differentiation and integration.
- Statistical Methods; commonly used probability distributions, parameter estimation, sampling variability, hypothesis testing, basic measures of bivariate relationships.
- Generalised Linear Models; the general linear model and the least-squares method, logistic regression for binary responses, Poisson regression for count data. More broadly, how to build a flexible linear predictor to capture relationships of interest.

These short courses will be followed by supported tutorial sessions in weeks 3-10.

Learning: On completion of this module a student will be able to:

- Comprehend the mathematical notation used in explaining probability and statistics.
- Demonstrate knowledge of basic principles in probability, statistical distributions, sampling and estimation.
- Make decisions on the appropriate way to test hypothesis, and carry out the test and interpret the results.
- Demonstrate knowledge of the general linear model, the least-squares method of estimation, and the linear predictor. As well the extensions to generalised linear models for discrete data.
- Decide on the appropriate way to statistically address a research question. Carry out all aspects of a statistical analysis, assessing model results and performance.
- Report their findings.

Assessment: There will be three pieces of coursework:

- One assessment for Statistical Methods; assessing understand and application of statistical concepts, and interpretation of results from hypothesis testing.
- Two independently produced reports for Generalized Linear Models; centred on in-depth statistical analyses.

Recommended texts and other learning resources:

- Upton, G., & Cook, I. (1996). *Understanding statistics*. Oxford University Press.
- Rice, J. (2006). *Mathematical statistics and data analysis*. Cengage Learning.
- Dobson, A. J., & Barnett, A. G. (2008). *An Introduction to Generalized Linear Models*. CRC Press.
- Fox, J. (2008). *Applied regression analysis and generalized linear models*. Sage Publications.

Module Mnemonic:	MATH551
Module Title:	Likelihood Inference
Module Convenor:	Dr Juhyun Park
Assessment:	Coursework (35%), module test (15%), and end of year exam (50%)
Duration:	25 hours
Credits:	15
Term:	M1, M2
Prerequisites:	UG Mathematics/Statistics (probability theory; calculus; matrices etc)

This course considers the idea of statistical models, and how the likelihood function, the probability of the observed data viewed as a function of unknown parameters, can be used to make inference about those parameters. This inference includes both estimates of the values of these parameters, and measures of the uncertainty surrounding these estimates. We consider single and multi-parameter models, and models which do not assume the data are independent and identically distributed. We also cover basic computational aspects of likelihood inference that are required in many practical applications.

Topics covered will include:

- Definition of the likelihood function for single and multi-parameter models, and how it is used to calculate point estimates (maximum likelihood estimates).
- Asymptotic distribution of the maximum likelihood estimator, and the profile deviance, and how these are used to quantify uncertainty in estimates.
- Inter-relationships between parameters, and the definition and use of orthogonality.
- Generalised Likelihood Ratio Statistics, and their use for hypothesis tests.
- Calculating likelihood functions for non-IID data.
- Simple use of computational methods to calculate maximum likelihood estimates and confidence intervals and to perform hypothesis tests.
- Model criticism through residual analysis.

On successful completion students will be able to:

- Understand how to construct statistical models for simple applications.
- Appreciate how information about the unknown parameters is obtained and summarized via the likelihood function.
- Calculate the likelihood function for IID data, and for some statistical models which do not assume independent identically distributed data.
- Evaluate point estimates and make statements about the variability of these estimates.
- Understand the inter-relationships between parameters, and the concept of orthogonality.
- Perform hypothesis tests using the generalised likelihood ratio statistic.
- Use computational methods to calculate maximum likelihood estimates.
- Use computational methods to construct confidence intervals, and perform hypothesis tests.
- Look at residuals to judge how appropriate a model is.

Bibliography:

- Azzalini. Statistical Inference: Based on the Likelihood. Chapman and Hall. 1996.
- D R Cox and D V Hinkley. Theoretical Statistics. Chapman and Hall. 1974.
- Y Pawitan. In All Likelihood: Statistical Modeling and Inference Using Likelihood. OUP. 2001.

Module Mnemonic:	MATH552
Module Title:	Generalised Linear Models
Module Convenor:	Dr Kanchan Mukherjee
Assessment:	Coursework (35%), module test (15%), and end of year exam (50%)
Duration:	25 hours
Credits:	15
Term:	M1, M2
Co-requisites:	MATH551

Generalised linear models are now one of the most frequently used statistical tools of the applied statistician. They extend the ideas of regression analysis to a wider class of problems that involves exploring the relationship between a response and one or more explanatory variables. In this course we aim to discuss applications of the generalised linear models to diverse range of practical problems involving data from the area of biology, social sciences and time series to name a few and to explore the theoretical basis of these models.

Topics covered:

- We introduce a large family of models, called the generalised linear models (GLMs), that includes the standard linear regression model as a special case and to discuss the theoretical properties of these models.
- We learn a common algorithm called iteratively reweighted least squares algorithm for the estimation of parameters.
- We fit and check these models with the statistical package R; produce confidence intervals and tests corresponding to questions of interest; and state conclusions in everyday language.

On successful completion students will be able to demonstrate knowledge and practice in:

- Model specification: choosing a suitable GLM; equivalent models; aliasing.
- Fitting models: maximum likelihood estimation using R.
- Effects of covariates: confidence intervals and tests of quantities of interest, interaction.
- Variable selection: backwards stepwise selection of covariates.
- Assessing model fit: deviance and Pearson residuals; leverage; residual deviance test of model fit; over-dispersion.
- State conclusions in everyday language.

Bibliography:

- P. McCullagh and J. Nelder. Generalized Linear Models, Chapman and Hall, 1999.
- A.J. Dobson, An Introduction to Generalised Linear Models, Chapman and Hall, 1990.

Module Mnemonic:	MATH555
Module Title:	Bayesian Inference for Data Science
Module Convenor:	Chris Nemeth
Assessment:	Coursework (50%) and end of year exam (50%)
Duration:	25 hours
Credits:	15
Term:	L1, L2
Prerequisites:	MATH551, MATH552

This module aims to introduce the Bayesian view of statistics, stressing its philosophical contrasts with classical statistics, its facility for including information other than the data into the analysis and its coherent approach towards inference and model selection. The module will also introduce students to MCMC (Markov chain Monte Carlo), a computationally intensive method for efficiently applying Bayesian methods to complex models. By the end of the course the students should be able to formulate an appropriate prior for a variety of problems, calculate, simulate from and interpret the posterior and the predictive distribution, with or without MCMC as appropriate and to carry out Bayesian model selection using the marginal likelihood. Students should be able to carry out all of this using the programming language R.

Topics covered:

- Inference by updating belief.
- The ingredients of Bayesian inference: the prior, the likelihood, the posterior, the predictive and the marginal distribution.
- Methods for formulating the prior.
- Conjugate priors for single parameter models.
- Normal distribution, known and unknown variance, regression.
- Sampling for the posterior and predictive distributions.
- Model checking and model selection.
- Gibbs sampling, data augmentation, hierarchical models.
- The Metropolis-Hastings algorithm, random walk Metropolis, independence sampler.

Bibliography:

- Hoff, P. (2008) A first course in Bayesian statistics. Springer
- Gamerman, D. and Lopez, H. (2006) MCMC statistical simulation for Bayesian inference. Chapman and Hall 2nd Edition.
- Gilks, W.R., Richardson, S. and Spiegelhalter, D. (1996) Markov chain Monte Carlo in Practice. Chapman and Hall.

Module Mnemonic:	MATH562
Module Title:	Extreme Value Theory
Module Convenor:	Dr Jennifer Wadsworth
Assessment:	Coursework (50%) and end of year exam (50%)
Duration:	20 hours (intensive teaching mode in week 11)
Credits:	10
Term:	L1 (weeks 11/12)
Prerequisites:	MATH551, MATH552

This module aims to develop the asymptotic theory, and associated techniques for modelling and inference, associated with the analysis of extreme values of random processes. The course will focus on the mathematical basis of the models, the statistical principles for implementation and the computational aspects of data modelling. Students are expected to acquire the following: an appreciation of, and facility in, the various asymptotic arguments and models; an ability to fit appropriate models to data using specially developed R software; the ability to understand and interpret fitted models.

For many physical processes, especially environmental processes, it is extremes of the process that are of greatest concern; the highest sea-levels cause floods; the fastest wind-speeds destroy buildings, etc. Most of the statistical theory is concerned with modelling typical behaviour; in contrast, the analysis of extremes requires us to model the unusual. This means that we have very little data with which we can either develop or estimate models. In the absence of alternatives, asymptotic theory is used as the basis for model development, but the issue of data scarcity leads to interesting challenges for creating models that optimise such data as are available.

Topics covered:

- Asymptotic theory for maxima of univariate independent and identically distributed (iid) random variables: limit distributions, GEV distribution, and domains of attraction.
- Extension of asymptotic theory for univariate iid variables to cover top order statistics and threshold exceedances: GP distribution.
- Statistical modelling and inference using maxima and threshold methods.
- Statistical modelling of extremes of non-identically distributed random variables.
- Asymptotic theory and statistical methods for extreme values of stationary sequences: clustering, extremal index.

Bibliography:

- S G Coles, An Introduction to the Statistical Modelling of Extreme Values, Springer-Verlag, London, 2001.

Module Mnemonic:	MATH563
Module Title:	Clinical Trails
Module Convenor:	Dr Deborah Costain
Assessment:	Course work (50%) and written exam (50%)
Duration:	20 hours (intensive teaching in week 11)
Credits:	10
Term:	L1 (weeks 11/12)

This course aims to introduce students to aspects of statistics, which are important in the design and analysis of clinical trials. On completion of the module students should understand the basic elements of clinical trials, be able to recognise and use principles of good study design, and be able to analyse and interpret study results to make correct scientific inferences.

Clinical trials are planned experiments on human beings designed to assess the relative benefits of one or more forms of treatment. For instance, we might be interested in studying whether aspirin reduces the incidence of pregnancy-induced hypertension; or we may wish to assess whether a new immunosuppressive drug improves the survival rate of transplant recipients.

Topics covered:

- Clinical trials fundamentals: design issues, ethics and defining and estimating treatment.
- Cross-over trials.
- Sample size determination.
- Equivalence and non-inferiority trials.
- Meta-analysis.
- Discussion of more general methodological and ethical issues.

Bibliography:

- D.G. Altman, Practical Statistics for Medical Research, Chapman and Hall, 1991.
- S. Senn, Cross-over trials in clinical research, Wiley, 1993.
- S. Piantadosi, Clinical Trials: A Methodologic Perspective, John Wiley & Sons, 1997.
- ICH Harmonised Tripartite Guidelines.
- J.N.S. Matthews, Introduction to Randomised Controlled Clinical Trials, Arnold, 2000.

Module Mnemonic:	MATH564
Module Title:	Principles of Epidemiology
Module Convenor:	Dr Tom Palmer
Assessment:	Coursework (50%) and written Exam (50%)
Duration:	20 hours (intensive teaching in week 13)
Credits:	10
Term:	L1 (weeks 13/14)

Epidemiology is the study of the distribution and determinants of disease in human populations. This course provides an introduction to the principles and statistical methods of epidemiology. Various concepts and strategies used in epidemiological studies are examined. Most inference will be likelihood based, although the emphasis is on conceptual considerations.

Topics covered:

- The history of epidemiology and the role of statistics
- Measures of health and disease: incidence, prevalence and cumulative incidence risk
- Types of epidemiological studies: Randomized controlled trials, cohort studies, case-control studies, cross-sectional and ecological studies
- Causation in epidemiology
- Potential errors in epidemiological studies: selection bias, confounding
- Remedies for confounding: Standardized rates, stratification and matching
- Diagnostic test studies

Bibliography:

- R. Beaglehole, R. Bonita and T. Kjellstroem (1993) Basic epidemiology. Geneva: World Health Organization.
- D. Clayton and M. Hills (1993) Statistical models in epidemiology. Oxford: Oxford University Press.
- M. Woodward (1999) Epidemiology: Study design and data analysis. Chapman & Hall, Boca Raton
- K.J. Rothman, S. Greenland and T.L. Lash. Modern Epidemiology. Lippincott Williams & Wilkins, US, 2008.

Module Mnemonic:	CHIC565
Module Title:	Environmental Epidemiology
Module Convenor:	Dr Ben Taylor
Assessment:	Coursework (50%) and written Exam (50%)
Duration:	20 hours (intensively in week 19)
Credits:	10
Term:	L2 (weeks 19/20)
Prerequisites:	MATH551; MATH552; MATH564

This course aims to introduce students to the kinds of statistical methods commonly used by epidemiologists and statisticians to investigate the relationship between risk of disease and environmental factors. Specifically the course concentrates on studies with a spatial component. A number of published studies will be used to illustrate the methods described, and students will learn how to perform similar analyses using the statistical R package. By the end of the course students should have an awareness of the kinds of methods used in environmental epidemiology, including an appreciation of their limitations. They should also be capable of conducting a number of these methods themselves.

Topics covered:

- Introduction: Motivating examples for methods in course.
- Spatial Point Processes: Theory and methods for the analysis of distributions of points in two-dimensional space; the Poisson process; univariate and bivariate K-functions.
- Spatial variation in risk: Case-control point-based methods and methods based on counts; spatial clustering.
- Disease mapping: Graphical investigation of spatial variation in risk; constructing smooth maps of disease risk from area-level count data.
- Geographical correlation studies: Poisson regression; the ecological fallacy and its relation with disease mapping.
- Point source methods: Investigation of risk associated with distance from a point or line source, for point and count data; practical implementation in epidemiological studies.

Bibliography:

- P.J. Diggle. Statistical Analysis of Spatial Point Patterns (2nd edition). London: Edward Arnold. 2003.
- P. Elliott, J. Cuzick, D. English and R. Stern (eds). Geographical and environmental epidemiology: methods for small-area studies. Oxford: Oxford University Press, 1992.
- O. Schabenberger and C.A. Gotway. Statistical Methods in Spatial Data Analysis. Boca Raton: Chapman & Hall/CRC, 2005.
- L. Waller and C.A. Gotway. Applied Spatial Statistics for Public Health Data. New York: Wiley, 2004.

Module Mnemonic:	MATH566
Module Title:	Longitudinal Data Analysis
Module Convenor:	Dr Juhyun Park
Assessment:	Coursework (50%) and written Exam (50%)
Duration:	20 hours (intensive teaching in week 15)
Credits:	10
Term:	L2 (weeks 15/16)
Prerequisites	MATH551; MATH552

Longitudinal data arise when a time-sequence of measurements is made on a response variable for each of a number of subjects in an experiment or observational study. For example, a patient's blood pressure may be measured daily following administration of one of several medical treatments for hypertension. The practical objective of many longitudinal studies is to find out how the average value of the response varies over time, and how this average response profile is affected by different experimental treatments. This module presents an approach to the analysis of longitudinal data, based on statistical modelling and likelihood methods of parameter estimation and hypothesis testing.

The specific aim of this module is to teach students a modern approach to the analysis of longitudinal data. Upon completion of this course the students should have acquired, from lectures and practical classes, the ability to build statistical models for longitudinal data, and to draw valid conclusions from their models.

Syllabus:

- What are longitudinal data?
- Exploratory and simple analysis strategies
- The independence working assumption
- Normal linear model with correlated errors
- Linear mixed effects models
- Generalised estimating equations
- Dealing with dropout

Bibliography:

- H. Brown and R. Prescott, Applied Mixed Models in Medicine, Wiley, 1999.
- P.J. Diggle, P. Heagerty, K.Y. Liang and S.L. Zeger, Analysis of Longitudinal Data (second edition), Oxford University Press, 2002.
- G. Verbeke and G. Molenberghs, Linear Mixed Models for Longitudinal Data, Springer, 2000.
- R. E. Weiss, Modelling longitudinal data, Springer, 2005.

Module Mnemonic:	MATH572
Module Title:	Genomics: technologies and data analyses
Module Convenor:	Dr Thomas Jaki
Assessment:	Coursework (100%)
Duration:	Distance learning with workshops in Lent term (weeks 11 – 20)
Credits:	10
Term:	L1,L2

To describe several modern genomics technologies in their biological context, to describe the types of data that are obtained from these technologies, to describe the statistical methodologies used to analyse such data and to have the students use statistical packages to perform such analyses on example data sets and interpret the results.

Genomics is a large field dealing with everything from DNA to metabolites and from evolution to microbiology. This course focuses on several genomics technologies by first of all putting them into their biological and genetic context and secondly describing the types of biological questions that can be answered with the data from these technologies. As an example of technologies that are to be discussed during this course are (i) DNA sequencing (ii) SNP (iii) microarrays and (iv) blotting and other proteomics methods. The most commonly use statistical analysis tools for each technology described will be discussed. The types of biological questions that are going to be addressed relate to issues such as genomic variation, constant genome and changing expression, human evolution and migration, disease and normality.

Syllabus:

- DNA sequencing, Single Nucleotide Polymorphisms (SNPs), transcriptional analysis via microarrays and Proteomics.
- Visualisation methods, hypothesis testing, multiple testing problems, multivariate methods, regression analyses for high dimensional data.

Bibliography:

- Malcolm Campbell and Laurie J. Heyer. Genomics, Proteomics and Bioinformatics. CSHL Press, 2003.
- Lange, K. Mathematical and Statistical Methods for Genetic Analysis, Springer, 2nd ed. 2002.
- Wit, E. C. and McClure, J. D. Statistics for Microarrays: Design, Analysis and Inference, John Wiley & Sons, 2004.

Module Mnemonic:	MATH573
Module Title:	Survival and Event History Analysis
Module Convenor:	Dr Andrew Titman
Assessment:	Coursework (50%) and written Exam (50%)
Duration:	20 hours (intensive teaching in week 17)
Credits:	10
Term:	L2 (weeks 17/18)
Pre-requisites:	MATH551; MATH552

To describe the theory and to develop the practical skills required for the design and analysis of medical studies leading to the observation of survival times or multiple failure times. By the end of the course students should be able to develop study designs and to carry out sophisticated analyses of this type, should be aware of the variety of statistical models and methods now available, and understand the nature and importance of the underlying model assumptions.

In many medical applications interest lies in times to or between events. Examples include time from diagnosis of cancer to death, or times between epileptic seizures. This advanced course begins with a review of standard approaches to the analysis of possibly censored survival data. Survival models and estimation procedures are reviewed, and emphasis is placed on the underlying assumptions, how these might be evaluated through diagnostic methods and how robust the primary conclusions might be to their violation. Study design is considered, in particular how to define failure and censoring and how to determine a suitable sample size and duration of follow-up.

The course closes with a description of models and methods for the treatment of multivariate survival data, such as repeated failures, the lifetimes of family members or competing risks. Stratified models, marginal models and frailty models are discussed.

Syllabus:

- Survival data. Censoring. Survival, hazard and cumulative hazard functions. Kaplan-Meier plots. Parametric models and likelihood construction. Cox's proportional hazards model, partial likelihood
- Time-dependent covariates. Diagnostic methods. Residual analysis. Testing the proportional hazards assumption
- Competing risks data, cause-specific hazard and cumulative incidence functions
- Stratified models, marginal models, frailty models

Bibliography:

- Collett, D. Modelling Survival Data in Medical Research. Chapman and Hall/CRC, 2003.
- Hougaard, P. Analysis of Multivariate Survival Data. Springer, 2000.
- Therneau, TM and Grambsch, PM. Modelling Survival Data: Extending the Cox Model. Springer, 2000.
- Pintilie, M. Competing Risks: A Practical Perspective. Wiley, 2006.

Module Mnemonic:	MSCI523
Module Title:	Forecasting
Module Convenor:	Dr Sven Crone TBC
Assessment:	Coursework (100%)
Credits:	10
Term:	L1, L2
Prerequisites:	(MATH551 and MATH552) OR (CFAS406 and CFAS440)

The module introduces time series and causal forecasting methods so that passing students will be able to prepare methodologically competent, understandable and concisely presented reports for clients. By the end of the course, students should be able to model causal and time series models, assess their accuracy and robustness and apply them in a real world problem domain.

Syllabus:

- Introduction to Forecasting in Organisations: Extrapolative vs. Causal Forecasting; History & academic research in Forecasting; Forecasting case studies.
- Data Exploration: Time Series Patterns; Univariate & Multivariate Visualisation; Naïve Forecasting Methods & Averages.
- Exponential Smoothing Methods: Single, Seasonal & Trended Exponential Smoothing; Model Selection; Parameter Selection.
- ARIMA Methods: AR-, MA-, ARMA and ARIMA Models; ARIMA Model specification & estimation; Automatic selection.
- Time Series Regression : Simple & multiple regression on time series; Hypothesis testing; Model evaluation; Diagnostics
- Time Series Regression: Model specification and constraints; Dummy Variables, Lag, Non-linearities; Stationarity; Building regression models.
- Applications in operations and marketing.
- Judgmental Forecasting: Judgmental methods for forecasting; Biases and heuristics.

Bibliography:

- Ord K. & Fildes R. (2013), Principles of Business Forecasting, South-Western Cengage Learning.

Module Mnemonic:	MSCI526
Module Title:	Data Mining for Marketing, Sales and Finance
Module Convenor:	Dr Nicos Pavlidis TBC
Assessment:	Coursework (100%)
Credits:	10
Term:	L1, L2
Prerequisites:	(MATH551 and MATH552) OR (CFAS406 and CFAS440)

The course extends the concepts of statistical model building and the models from the Introductory Statistics module towards methods from machine learning and artificial intelligence.

By the end of the course you should be able to:

- Understand general modelling concepts in relation to complex models
- Use a wide range of data mining methods to handle data of different types & applications
- Understand how these methods may be applied in practical management contexts
- Use & apply SAS Enterprise Miner to deal with complexity and large datasets

Syllabus:

- Introduction to Data Mining
- Data Mining Process: Methods for data exploration & manipulation; Methods for data reduction & feature selection; Evaluating Classification Accuracy.
- Data Mining Methods for Classification: Logistic Regression; Decision Trees; Nearest neighbour classification; Artificial Neural Networks.
- Data Mining applications in Credit Scoring

Bibliography:

- Tan, P. N., M. Steinbach, et al. (2005). Introduction to data mining. Boston, Pearson Addison Wesley.
- Berry, M. J. A. and G. Linoff (2000). Mastering data mining: the art and science of customer relationship management. New York, NY [u.a.], Wiley Computer Publ.
- Berry, M. J. A. and G. Linoff (2004). Data mining techniques: for marketing, sales, and customer relationship management. Indianapolis, Ind., Wiley Pub.
- Linoff, G. and M. J. A. Berry (2001). Mining the Web: transforming customer data into customer value. New York, John Wiley & Sons.
- Weiss, S. M. and N. Indurkha (1998). Predictive data mining: a practical guide. San Francisco, Morgan Kaufmann Publishers.

Module Mnemonic:	MSCI534
Module Title:	Optimisation and Heuristics
Module Convenor:	Dr Adam Letchford TBC
Assessment:	Coursework (30%), Exam (70%)
Credits:	10
Term:	L1, L2
Prerequisites:	(MATH551 and MATH552) OR (CFAS406 and CFAS440)

Optimisation, sometimes called *mathematical programming*, has applications in many fields, including Operational Research, Computer Science, Statistics, Finance, Engineering and the Physical Sciences. Commercial optimisation software is now capable of solving many industrial-scale problems to proven optimality. On the other hand, there are still many practical applications where finding a provably-optimal solution is not computationally viable. In such cases, *heuristic* methods can allow good solutions to be found within a reasonable computation time.

The course is designed to enable students to apply optimisation techniques to business problems. Building on the introduction to optimisation in MSCI502 and/or MSCI519, students will be introduced to different problem formulations and algorithmic methods to guide decision making in business and other organisations.

Syllabus:

- Linear Programming.
- Non-Linear Programming.
- Integer and Mixed-Integer Programming.
- Dynamic Programming.
- Heuristics.

By the end of the course you should be able to:

- know how to formulate problems as mathematical programs and solve them.
- be aware of the power, and the limitations, of optimisation methods.
- be able to carry out sensitivity analysis to see how robust the recommendation is.
- be familiar with commercial software such as MPL, LINDO and EXCEL SOLVER.
- be aware of major heuristic techniques and know when and how to apply them.

Bibliography:

- HP Williams (2013) *Model Building in Mathematical Programming* (5th edition). Wiley. ISBN: 978-1-118-44333-0 (pbk).
- J Kallrath & JM Wilson (1997) *Business Optimisation Using Mathematical Programming*. Macmillan. ISBN: 0-333-67623-8.
- WL Winston (2004) *Operations Research - Applications and Algorithms* (4th edition). Thompson. ISBN: 978-0534380588.
- DR Anderson, DJ Sweeney, TA Williams & M. Wisniewski (2008) *An Introduction to Management Science*. Cengage Learning. ISBN: 978-1844805952.
- E.K. Burke & G. Kendall (eds.) (2005) *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*. Springer.

Module Mnemonic:	SCC401
Module Title:	Elements of Distributed Systems
Module Convenor:	Dr. Amit Chopra and Dr. Barry Porter TBC
Assessment:	100% Coursework

In this module we explore the fundamental principles, techniques and technologies that underpin today's global IT infrastructure. It is one of two complementary modules that comprise the Systems stream of the Data Science MSc, which together enable students to assess new systems technologies, to know where technologies fit in a comprehensive schema, and to know what to read to go deeper.

The principal ethos of the module is to focus on the properties of system components, with the aim of encouraging a principled understanding of the strengths, weaknesses, scalability and bottlenecks of systems components. This will enable graduates of the MSc to be able to make intelligent and well-reasoned trade-offs between fundamental building blocks of distributed systems in today's IT infrastructure.

Further, the course will review state of the art thinking regarding algorithms, and technologies behind such architectures, placing these within the framework of the current research agenda.

Topics to be covered will include:

The module will cover two key areas, fundamental techniques/ principles and fundamental technologies/ paradigms. Fundamental techniques/ principles will include coverage of the following:

- Caching
- Tiering
- Replication
- Synchronization
- Failure
- Reliability

Fundamental technologies/ paradigms will include coverage of the following:

- Interaction paradigms in distributed systems
- Peer-to-peer architecture
- Scalable and high-performance networking
- Scalable and enterprise storage
- Data acquisition (e.g. sensor networks)
- Enterprise computing and scalable processing
- Large scale distributed information systems (e.g. high-performance web architectures)
- High performance computer clusters, grid architectures
- Enterprise security
- Organisational impacts (e.g. data protection, security)

On successful completion of this module students will:

- Demonstrate a deep understanding of the fundamental elements that underpin state of the art enterprise distributed systems;
- Describe and critically evaluate core techniques and paradigms used within enterprise IT systems;
- Understand and appreciate the trade-offs, strengths and limitations of systems elements in principle and practice in modern IT systems.

Module Mnemonic:	SCC403
Module Title:	Data Mining
Module Convenor:	Dr. Plamen Angelov
Assessment:	100% Coursework

This module will provide a comprehensive coverage of the problems related to Data representation, storage, manipulation, retrieval and processing in terms of extracting information from the data. It has been designed to prepare the fundamental theoretical level of knowledge and skills (at the related laboratory sessions) of the students to this specific aspect of Data Science which plays an important role in any system and application. In this way it prepares them for the second module on the topic of Data as well as for their projects.

Topics to be covered will include:

- Data Primer: Setting the scene: Big Data, Cloud Computing; The time, storage and computing power compromise: off-line versus on-line;
- Data Representations:
- Storage Paradigms
- Vector-space models
- Hierchical (agglomerative/diversive)
- k means
- SQL and Relational Data Structures (short refresher)
- NoSQL: Document stores, graph databases
- Inference and reasoning
- Associative and Fuzzy Rules
- Inference mechanisms
- Data Processing:
- Clustering
- Density-based, on-line, evolving
- Classification
- Randomness and determinism, frequentistic and belief based approaches, probability density, recursive density estimation, averages and moments, important random signals, response of linear systems to random signals, random signal models.
- Discriminative (Linear Discriminant Analysis, Single Perceptron, Multi-layer Perceptron, Learning Vector Classifier, Support Vector Machines), Generative (Naive Bayes)
- Supervised and unsupervised learning, online and offline systems, adaptive and evolving systems, evolving versus evolutionary systems, normalisation and standardisation
- Fuzzy Rule-based Classifiers, Regression or Lable based classifiers
- Self-learning Classifiers, evolving Classifiers, dynamic data space partitioning using evolving clustering and data clouds, monitoring the quality of the self-learning system online, evolving multi-model predictive systems.
- Semi-supervised Learning (Self-learning, evolving, Bootstrapping, Expectation-Maximisation, ensemble classifiers)
- Information Extraction vs Retrieval

On successful completion of this module students will:

- Demonstrate understanding of the concepts and specific methodologies for data representation and processing and their applications to practical problems
- Analyse and synthesise effective methods and algorithms for data representation and processing
- Develop software scripts that implement advanced data representation and processing and demonstrate their impact on the performance
- List, explain and generalise the trade-offs of performance and complexity in designing practical solutions for problems of data representation and processing in terms of storage, time and computing power

Module Mnemonic:	SCC411
Module Title:	Systems Architecture and Integration
Module Convenor:	Dr. Barry Porter and Dr Amit Chopra TBC
Assessment:	100% Coursework

Module Details:

In this module we explore the architectural approaches, techniques and technologies that underpin today's global IT infrastructure and particularly large-scale enterprise IT systems. It is one of two complementary modules that comprise the Systems stream of the Data Science MSc, which together provide a broad knowledge and context of systems architecture enabling students to assess new systems technologies, to know where technologies fit in the larger scheme of enterprise systems and state of the art research thinking, and to know what to read to go deeper.

The principal ethos of the module is to focus on the principles, emergent properties and application of systems elements as used in large-scale and high performance systems. Detailed case studies and invited industrial speakers will be used to provide supporting real-world context and a basis for interactive seminar discussions.

Topics to be covered will include:

- Systems of systems composition
- Scalability concerns
- Systems integration/interoperability
- Software and Infrastructure as a Service (i.e. Cloud computing principles)

Supported by a consideration of emerging issues and implications arising from these new technologies:

- Commercial considerations
- Legal and ethical considerations
- New development and support paradigms, including open sourcing

In addition to the discussion and seminar led aspects of the course, we envisage 'hands-on' measurement-based coursework that looks empirically at the scalability of a significant technology, e.g. a cloud system such as Amazon EC2.

On successful completion of this module students will:

- Demonstrate a deep understanding of the architectures and approaches for large-scale systems implementation
- Describe and critically evaluate techniques and paradigms used within enterprise-scale IT systems
- Understand and appreciate the trade-offs, strengths and limitations of systems architectures in principle and practice in modern IT systems.

Module Mnemonic:	SCC413
Module Title:	Applied Data Mining
Module Convenor:	Dr. Matthew Rowe
Assessment:	100% Coursework

Module Details:

This module will provide students with up-to-date information on current applications of data in both industry and research. The module will build on SCC.403 – Fundamentals of Data by explaining how data is processed and applied at large-scale across a variety of different areas.

Topics to be covered will include:

- The Semantic Web: primer, crawling and spidering Linked Data, open-track large-scale problems (e.g. Billion Triples Challenge), distributed and federated querying, distributed reasoning, ontology alignment.
- The Social Web: primer, user-generated content and crowd-sourced data, social networks (theories, analysis), recommendation (collaborative filtering, content recommendation challenges, and friend recommendation/link prediction).
- The Scientific Web: from big data to big science, open data, citizen science, and case studies (virtual environmental observatories, collaboration networks).
- Scalable data processing: primer, scaling the semantic web (scaling distributed reasoning using MapReduce), scaling the social web (collaborative filtering, link prediction), and scalable network analysis for the scientific web.

On successful completion of this module students will be able to:

- Create scalable solutions to problems involving data from the semantic, social and scientific web
- Process networks and perform network analysis to identify key actors in information flow;
- Understand the current trends of research in the semantic, social and scientific web and what challenges still remain to be solved
- Demonstrate working knowledge of distributing work loads for scalable applications.

Module Mnemonic:	SCC460
Module Title:	Data Science Fundamentals
Module Convenor:	El Khatib, Yehia (elkhatib); Gellersen, Hans, Prof. Hans Gellersen
Assessment:	100% Coursework
Duration:	20 hours
Credits:	15
Term:	M2, L1, L2

This module teaches students about how data science is performed within academic and industry (via invited talks), research methods and how different research strategies are applied across different disciplines, and data science techniques for processing and analysing data. Students will engage in group project work, based on project briefs provided by industrial speakers, within multi-skilled teams (e.g. computing students, statistics students, environmental science students) in order to apply their data science skills to researching and solving an industrial data science problem.

Topics covered will include:

- The role of the data scientist and the evolving epistemology of data science.
- The language of research, how to form research questions, writing literature reviews, and variance of research strategies across disciplines.
- Ethics surrounding data collection and re-sharing, and unwanted inferences.
- Identifying potential data sources and the data acquisition processes.
- Defining and quantifying biases, and data preparation (e.g. cleaning, standardisation, etc.).
- Choosing a potential model for data, understanding model requirements and constraints, specifying model properties a priori, and fitting models.
- Inspection of data and results using plots, and hypothesis and significance tests.
- Writing up and presenting findings.

Learning: Students will learn through a series of group exercises around research studies and projects related to data science topics. Invited talks from industry tackling data science problems will be given to teach the students about the application of data science skills in industry and academia. Students will gain knowledge of:

- Defining a research question and a hypothesis to be tested, and choosing an appropriate research strategy to test that hypothesis.
- Analysing datasets provided in heterogeneous forms using a range of statistical techniques
- How to relate potential data sources to a given research question, acquire such data and integrate it together.
- Designing and performing appropriate experiments given a research question.
- Implementing appropriate models for experiments and ensuring that the model is tested in the correct manner.
- Analysing experimental findings and relating these findings back to the original research goal.

Assessment: Assessment is comprised of 50% group work project and 50% individual project proposal. The group project will involve working on a given data science problem provided by industry.

Recommended texts and other learning resources:

- O'Neil. C., and Schutt. R. (2013) *Doing Data Science: Straight Talk from the Frontline*. O'Reilly.
- Trochim. W. (2006) *The Research Methods Knowledge Base*. Cengage Learning

Module Mnemonic:	SCC461
Module Title:	Programming for Data Science
Module Convenor:	Dr Matthew Rowe, Dr Pete Neal
Assessment:	Coursework (50%), End of module report (50%)
Duration:	25 hours
Credits:	15
Term:	M1, M2, L1

This module aims to provide students with the necessary programming skills to statistically process and explore disparate datasets using R, to become confident in using this language to create and analyse variables in order to discover patterns and relationships through the use of visualisation, testing and modelling. It also aims to provide students with experience in using object-oriented programming concepts and principles to read in data from both local files and databases so that it can be merged together, using record-reconciliation techniques, and then output this into a single file for processing; this will be taught using the object-oriented programming language Java. The teaching of both Java and R is essential here as the former is well-suited to handling data, via the creation of bespoke data objects, while the latter is good for statistically assessing data.

Students will gain experience of working through exercise tasks and discussing their work with their peers; thereby fostering interpersonal communications skills - in particular in how students approach the tasks and their proposed solutions. Students will also gain an understanding of how to conceptualise a problem using available programming semantics, and determine the appropriate level of abstraction required in their solutions.

Topics covered will include:

- Principles of object-oriented programming (e.g. variables, abstraction, inheritance, polymorphism);
- Using R to manipulate data sets and to create features from data, while becoming familiar with at a more technical level with its data objects;
- Summarise and visualise data using custom functions and third-party libraries;
- Using Java to read in data from disparate sources and write data to local files;
- Performing record-reconciliation to merge heterogeneous datasets together;
- Producing clean, friendly, reusable code.

On successful completion of this module students will be able to:

- Use R to manipulate clean data sets and to create features from data;
- Have a understanding of the different data objects that exist in R and how to interact with them;
- Create and maintain custom functions and scripts which facilitate methodical data analysis;
- Produce data visualisations for both exploration purposes and for inclusion in reports;
- Read in data from local files and databases using Java;
- Identify the correct level of abstraction to model a system or process in Java and construct the necessary classes to represent that;
- Choose and apply an appropriate record-reconciliation approach for integrating data together.

Bibliography:

- Introductory statistics with R. Dalgaard, Peter. Springer, 2008. ISBN-13: 978-0387954752
- R Cookbook. Paul Teetor. O'Reilly Media; 1 edition. 2011. ISBN-13: 978-0596809157.
- Java, A Beginner's Guide, 5th Edition. Herbert Schildt. McGraw-Hill Osborne. 2011. ISBN-13: 978-0071606325.

Appendix B: Term Dates

Academic Year 2015-2016

- **Michaelmas Term:**
 - 2 October 2015 to 11 December 2015
- **Lent Term:**
 - 8 January 2016 to 18 March 2016
- **Summer Term:**
 - 15 April 2016 to 24 June 2016