REGULATIONS FOR THE DEGREE OF
MASTER OF DATA SCIENCE
(MDASC)
For students admitted in 2018-2019

(See also General Regulations and Regulations for Taught Postgraduate Curricula)

Any publication based on work approved for a higher degree should contain a reference to the effect that the work was submitted to the University of Hong Kong for the award of the degree.

Admission requirements

MD1   To be eligible for admission to the courses leading to the degree of Master of Data Science a candidate
- (a) shall comply with the General Regulations and the Regulations for Taught Postgraduate Curricula;
- (b) shall hold
  - (i) a Bachelor's degree with honours of this University, or
  - (ii) another qualification of equivalent standard from this University or another University or comparable institution acceptable for this purpose; and
- (c) shall pass a qualifying examination if so required; and
- (d) shall have taken at least one university or post-secondary certificate course in each of the following three subjects (calculus and algebra, computer programming and introductory statistics) or related areas.

Qualifying examination

MD2   (a) A qualifying examination may be set to test the candidate's formal academic ability
or   his ability to follow the courses of study prescribed. It shall consist of one or more written papers or their equivalent and may include a project report.
            (b) A candidate who is required to satisfy the examiners in a qualifying examination shall not be permitted to register until he has satisfied the examiners in the examination.

Period of study

MD3   (a) The curriculum shall normally extend over one and a half academic years of full-time study or two and a half academic years of part-time study. Candidates shall not be permitted to extend their studies beyond the maximum period of registration of three academic years of full-time study or four academic years of part-time study, unless otherwise permitted or required by the Board of the Faculty.
            (b) Candidates of full-time study may be permitted to complete the curriculum in one academic year, subject to the approval of the Board of the Faculty. The candidate should write formally to apply for shortening the normative period of study via the Department within one month after admission to the curriculum. For such candidates, it is recommended that the capstone course will start in the second semester with expectation to be completed in the summer semester.

Course Exemption and advanced standing

MD4   (a) In recognition of studies completed successfully before admission to the curriculum,

advanced standing of up to 12 credits may be granted to a candidate with appropriate qualification and professional experiences, on production of appropriate certification, subject to the approval of the Board of the Faculty. The candidate should write formally to apply for advanced standing via the Department within two weeks after admission to the curriculum.

(b) For cases of having satisfactorily completed more than 12 credits of another course
or courses equivalent in content to any of the compulsory courses as specified in the syllabuses, candidates may, on production of appropriate certification, be exempted from the compulsory course(s), subject to approval of the Board of the Faculty. Candidates so exempted must replace the number of exempted credits with electives course(s) in the curriculum of the same credit value.

Award of degree

MD5 To be eligible for the award of the degree of Master of Data Science, a candidate shall
(a) comply with the General Regulations and the Regulations for Taught Postgraduate Curricula; and
(b) successfully complete the curriculum in accordance with the regulations set out below.

A candidate who fails to fulfill the requirements within the maximum (i) three academic years for full-time mode of study or (ii) four academic years for part-time mode of study shall be recommended for discontinuation under the provisions of General Regulation G12, except that a candidate is granted permission to extend period of study by the Board of the Faculty in accordance with Regulation MD3.

Completion of curriculum

MD6 To successfully complete the curriculum, a candidate shall satisfy the requirements prescribed in TPG 6 of the Regulations for Taught Postgraduate Curricula; follow courses of instruction; and satisfy the examiners in the prescribed courses and in any prescribed form of examination in accordance with the regulations set out below.

Assessments

MD7 (a) In any course where so prescribed in the syllabus, coursework or a project report may constitute part or whole of the examination for the course.
(b) The written examination for each module shall be held after the completion of the prescribed course of study for that module, and not later than January, May or August immediately following the completion of the course of study for that module.

MD8 If during any academic year a candidate has failed at his/her first attempt in a course or courses, but is not required to discontinue his/her studies by Regulation MD9, the candidate may be permitted to make up for the failed courses in the following manner:
(a) undergoing re-assessment/re-examination in the failed course or courses to be held before the next academic year; or
(b) for repeating the course and re-examination in the failed course or courses in the next academic year; or
(c) for elective courses, taking another course in lieu and satisfying the assessment requirements.

MD9   Failure to undertake the examination of a course as scheduled shall normally result in automatic failure in that course. A candidate who, because of illness, is unable to be present at the written examination of any course may apply for permission to present himself/herself at a supplementary examination of the same course to be held before the beginning of the following academic year. Any such application shall be made on the form prescribed within two weeks of the first day of the candidate's absence from any examination.

MD10   A candidate may be required to discontinue his/her studies if he/she
    (a)   during any academic year has failed in half or more than half the number of credits of all the courses to be examined in that academic year; or
    (b)   has failed at a repeated attempt in any course; or
    (c)   has exceeded the maximum period of registration.

Grading

MD11   Individual courses shall be graded according the letter grading system as determined by the Board of Examiners. The standards and the grade points for assessment are as follows:

Grade	Standard	Grade Point
A+	**Excellent**	4.3
A		4.0
A-		3.7
B+	**Good**	3.3
B		3.0
B-		2.7
C+	**Satisfactory**	2.3
C		2.0
C-		1.7
D+	**Pass**	1.3
D		1.0
F	**Fail**	0

MD12   On successful completion of the curriculum, candidates who have shown exceptional merit at the whole examination may be awarded a mark of distinction, and this mark shall be recorded in the candidates' degree diploma.

**SYLLABUSES FOR THE DEGREE OF
MASTER OF DATA SCIENCE**

The Department of Statistics and Actuarial Science and Department of Computer Science jointly offer a postgraduate curriculum leading to the degree of Master of Data Science, with two study modes: the one and a half academic years' full-time mode and the two and a half academic years' part-time mode. The curriculum is designed to provide graduates with training in the principles and practice of data science. Candidates should have knowledge of calculus and algebra, computer programming and introductory statistics and should have taken at least one university or post-secondary certificate course in each of these three subjects or related areas.

STRUCTURE AND EVALUATION

Each student must complete at least 72 credits of courses. Courses with 6 credits are offered in the first and second semesters while courses with 3 credits are normally offered in the summer semester. If a student selects a course whose contents are similar to a course (or courses) which he/she has taken in his/her previous study, the Department may not approve the selection in question.

CURRICULUM
(applicable for both full-time and part-time modes)

Compulsory Courses (36 credits)	
COMP7305	Cluster and cloud computing
COMP7404	Computational intelligence and machine learning
DASC7011	Statistical inference for data science
DASC7104	Advanced database systems
STAT6014^	Advanced statistical modelling (previously named as Advanced statistical learning)
STAT7102^	Advanced statistical modelling
STAT7008	Programming for data science

^ Students should take either STAT6014 or STAT7102 to fulfil the requirement, but not both.

Disciplinary Electives (24 credits)*
with at least 12 credits from List A and 12 credits from List B

List A

COMP7105	Advanced topics in data science
COMP7409	Machine learning in trading and finance
COMP7503	Multimedia technologies
COMP7506	Smart phone apps development
COMP7507	Visualization and visual analytics
COMP7605	Advanced multimedia data analysis and applications
COMP7906	Introduction to cyber security
DASC7606	Deep learning
FITE7410	Financial fraud analytics
ICOM6044	Data science for business

List B

MATH6502	Topics in applied discrete mathematics
MATH6503	Topics in mathematical programming and optimization
MATH8502	Topics in applied discrete mathematics
MATH8503	Topics in mathematical programming and optimization
STAT6008	Advanced statistical inference
STAT6013	Financial data analysis
STAT6015	Advanced quantitative risk management and finance
STAT6016	Spatial data analysis
STAT6019	Current topics in statistics
STAT8003	Time series forecasting
STAT8017	Data mining techniques
STAT8019	Marketing analytics
STAT8021	Big data analytics
STAT8301	Big data analytics (3 credits)
STAT8306	Statistical methods for network data (3 credits)
STAT8307	Natural language processing and text analytics (3 credits)

	(previously named as Text analytics)
	*Students who have completed the same courses in their previous studies in HKU, e.g. Master of Statistics or Master of Science in Computer Science may, on production of relevant transcripts, be permitted to select up to 24 credits of disciplinary electives from either List A or List B above if they are not able to find any untaken options from either of the lists of disciplinary electives.
Capstone requirement (12 credits)	
DASC7600 Data science project (12 credits)	

All courses should be 6-credit bearing unless otherwise stated.

COURSE DESCRIPTION

Compulsory Courses

COMP7305   Cluster and cloud computing (6 credits)

This course offers an overview of current cloud technologies, and discusses various issues in the design and implementation of cloud systems. Topics include cloud delivery models (SaaS, PaaS, and IaaS) with motivating examples from Google, Amazon, and Microsoft; virtualization techniques implemented in Xen, KVM, VMWare, and Docker; distributed file systems, such as Hadoop file system; MapReduce and Spark programming models for large-scale data analysis, networking techniques in hyper-scale data centers. The students will learn the use of Amazon EC2 to deploy applications on cloud, and implement a novel cloud computing application on a Xen-enabled PC cluster as part of their term project.

Prerequisites: The students are expected to install various open-source cloud software in their Linux cluster, and exercise the systems configuration and administration. Basic understanding of Linux operating system and some programming experiences (C/C++, Java or Python) in a Linux environment are required.

Assessment: coursework (50%) and examination (50%)

COMP7404   Computational intelligence and machine learning (6 credits)

This course will teach a broad set of principles and tools that will provide the mathematical, algorithmic and philosophical framework for tackling problems using Artificial Intelligence (AI) and Machine Learning (ML). AI and ML are highly interdisciplinary fields with impact in different applications, such as, biology, robotics, language, economics, and computer science. AI is the science and engineering of making intelligent machines, especially intelligent computer programs, while ML refers to the changes in systems that perform tasks associated with AI. Ethical issues in advanced AI and how to prevent learning algorithms from acquiring morally undesirable biases will be covered.

Topics may include a subset of the following: problem solving by search, heuristic (informed) search, constraint satisfaction, games, knowledge-based agents, supervised learning, unsupervised learning; learning theory, reinforcement learning and adaptive control and ethical challenges of AI and ML.

Pre-requisites: Nil, but knowledge of data structures and algorithms, probability, linear algebra, and programming would be an advantage.

Assessment: coursework (50%) and examination (50%)

DASC7011   Statistical inference for data science (6 credits)

Computing power has revolutionized the theory and practice of statistical inference. Reciprocally, novel statistical inference procedures are becoming an integral part of data science. By focusing on the interplay between statistical inference and methodologies for data science, this course reviews the main concepts underpinning classical statistical inference, studies computer-intensive methods for conducting statistical inference, and examines important issues concerning statistical inference drawn upon modern learning technologies. Contents include classical frequentist and Bayesian inferences, computer-intensive methods such as the EM algorithm, the bootstrap and the Markov chain Monte Carlo, large-scale hypothesis testing, high-dimensional modeling, and post-model-selection inference.

Assessment: coursework (40%) and examination (60%)

DASC7104   Advanced database systems (6 credits)

The course will study some advanced topics and techniques in database systems, with a focus on the aspects of big data analytics, algorithms, and system design & organisation. It will also survey the recent development and progress in selected areas. Topics include: query optimization, spatial-spatiotemporal data management, multimedia and time-series data management, information retrieval and XML, data mining.

Assessment: coursework (50%) and examination (50%)

STAT6014   Advanced statistical modelling (6 credits)
            (previously named as Advanced statistical learning)

This course introduces modern methods for constructing and evaluating statistical models and their implementation using popular computing software, such as R or Python. It will cover both the underlying principles of each modelling approach and the model estimation procedures. Topics from: (i) Linear regression models; (ii) Generalized linear models; (iii) Mixed models; (iv) Kernel and local polynomial regression; (v) Generalized additive models; (vi) Hidden Markov model and Bayesian network.

Pre-requisites: Students should not be taking or have taken STAT7102 Advanced statistical learning or equivalent

Assessment: coursework (50%) and examination (50%)

STAT7008   Programming for data science (6 credits)

In the big data era, it is very easy to collect huge amounts of data. Capturing and exploiting the important information contained within such datasets poses a number of statistical challenges. This course aims to provide students with a strong foundation in computing skills necessary to use R or Python to tackle some of these challenges. Possible topics to be covered may include exploratory data analysis and visualization, collecting data from a variety of sources (e.g. excel, web-scraping, APIs and others), object-oriented programming concepts and scientific computation tools. Students will learn to create their own R packages or Python libraries.

Assessment: coursework (100%)

STAT7102   Advanced statistical modelling (6 credits)

This course introduces modern methods for constructing and evaluating statistical models and their implementation using popular computing software, such as R or Python. It will cover both the underlying principles of each modelling approach and the model estimation procedures. Topics from: (i) Linear regression models; (ii) Generalized linear models; (iii) Mixed models; (iv) Kernel and local polynomial regression; (v) Generalized additive models; (vi) Hidden Markov models and Bayesian networks.

Pre-requisites: Students should not be taking or have taken STAT6014 Advanced statistical learning or equivalent

Assessment: coursework (50%) and examination (50%)

Disciplinary Electives

COMP7105   Advanced topics in data science (6 credits)

This course will introduce selected advanced computational methods and apply them to problems in data analysis and relevant applications.

Assessment: coursework (50%) and examination (50%)

COMP7409   Machine learning in trading and finance (6 credits)

The course introduces our students to the field of Machine Learning, and help them develop skills of applying Machine Learning, or more precisely, applying supervised learning, unsupervised learning and reinforcement learning to solve problems in Trading and Finance.

This course will cover the following topics. (1) Overview of Machine Learning and Artificial Intelligence, (2) Supervised Learning, Unsupervised Learning and Reinforcement Learning, (3) Major algorithms for Supervised Learning and Unsupervised Learning with applications to Trading and Finance, (4) Basic algorithms for Reinforcement Learning with applications to optimal trading, asset management, and portfolio optimization, (5) Advanced methods of Reinforcement Learning with applications to high-frequency trading, cryptocurrency trading and peer-to-peer lending.

Assessment: coursework (65%) and examination (35%)

COMP7503   Multimedia technologies (6 credits)

This course presents fundamental concepts and emerging technologies for multimedia computing. Students are expected to learn how to develop various kinds of media communication, presentation, and manipulation techniques. At the end of course, students should acquire proper skill set to utilize, integrate and synchronize different information and data from media sources for building specific multimedia applications. Topics include media data acquisition methods and techniques; nature of perceptually encoded information; processing and manipulation of media data; multimedia content organization and analysis; trending technologies for future multimedia computing.

Assessment: coursework (50%) and examination (50%)

COMP7506   Smart phone apps development (6 credits)

Smart phones have become very popular in recent years. According to a study, by 2020, 70% of the world's population is projected to own a smart phone, an estimated total of almost 6.1 billion smartphone users in the world.

Smart phones play an important role in mobile communication and applications.

Smart phones are powerful as they support a wide range of applications (called apps). Most of the time, smart phone users just purchase their favorite apps wirelessly from the vendors. There is a great potential for software developer to reach worldwide users.

This course aims at introducing the design issues of smart phone apps. For examples, the smart phone screen is usually much smaller than the computer monitor. We have to pay special attention to this aspect in order to develop attractive and successful apps. Various modern smart phone apps development environments and programming techniques (such as Java for Android phones, and Swift for iPhones) will be introduced to facilitate students to develop their own apps.

Prerequisites: Students should have basic programming knowledge

Assessment: coursework (50%) and examination (50%)

COMP7507   Visualization and visual analytics (6 credits)

This course introduces the basic principles and techniques in visualization and visual analytics, and their applications. Topics include human visual perception; color; visualization techniques for spatial, geospatial and multivariate data, graphs and networks; text and document visualization; scientific visualization; interaction and visual analysis.

Assessment: coursework (50%) and examination (50%)

COMP7605   Advanced multimedia data analysis and applications (6 credits)

This course's objective is to introduce advanced multimedia data analysis techniques, and the design and implementation of signal processing algorithms. It covers topics on Digital Filter Realization, Recursive and Non-Recursive filters, Frequency Domain Processing, Two-Dimensional Signal Processing, and application of multimedia signal processing to speech production and analysis, image and video processing.

Assessment: coursework (50%) and examination (50%)

COMP7906   Introduction to cyber security (6 credits)

The aim of the course is to introduce different methods of protecting information and data in the cyber world, including the privacy issue. Topics include introduction to security; cyber attacks and threats; cryptographic algorithms and applications; network security and infrastructure.

Pre-requisites: Students should not have taken ICOM6045 Fundamentals of e-commerce security or equivalent

Assessment: coursework (50%) and examination (50%)

DASC7606   Deep learning (6 credits)

Machine learning is a fast growing field in computer science and deep learning is the cutting edge technology that enables machines to learn from large-scale and complex datasets. This course will first introduce fundamental machine learning techniques and will then focus on artificial neural networks and how to train and optimize them to solve challenging problems using deep learning. Topics covered include linear and logistic regression, neural networks, convolutional neural networks, deep reinforcement learning and unsupervised feature learning. Popular deep learning software, such as Caffe, Torch and TensorFlow, will also be introduced.

Assessment: coursework (50%) and examination (50%)

FITE7410   Financial fraud analytics (6 credits)

This course aims at introducing various analytics techniques to fight against financial fraud. These analytics techniques include, descriptive analytics, predictive analytics, and social network learning. Various data set will also be introduced, including labeled or unlabeled data sets, and social network data set. Students learn the fraud patterns through applying the analytics techniques in financial frauds, such as, insurance fraud, credit card fraud, etc.

Key topics include: Handling of raw data sets for fraud detection; Applications of descriptive analytics, predictive analytics and social network analytics to construct fraud detection models; Financial Fraud Analytics challenges and issues when applied in business context.

Pre-requisites: Students should have basic knowledge about statistics concepts

Assessment: coursework (60%) and examination (40%)

ICOM6044   Data science for business (6 credits)

The emerging discipline of data science combines statistical methods with computer science to solve problems in applied areas. In this case we focus on how data science can be used to solve business problems especially those in electronic commerce. By its very nature e-commerce is able to generate large amounts of data and data mining methods are quite helpful for managers in turning this data into knowledge which in turn can be used to make better decisions. These data sets and their accompanying quantitative methods have the potential to dramatically change decision making in many areas of business. For example, ideas like interactive marketing, customer relationship management, and database marketing are pushing companies to utilize the information they collect about their customers in order to make better marketing decisions.

This course focuses on how data science methods can be applied to solve managerial problems in marketing and electronic commerce. Our emphasis is developing a core set of principles that embody data science: empirical reasoning, exploratory and visual analysis, and predictive modeling. We use these core principles to understand many methods used in data mining and machine learning. Our strategy in this course is to survey several popular techniques and understand how they map into these core principles. These techniques are illustrated with case studies. However, the emphasis is not on the software for implementing these techniques but on understanding the inputs and outputs of these techniques and how they are used to solve business problems.

Assessment: coursework (65%) and examination (35%)

MATH6502   Topics in applied discrete mathematics (6 credits)

This course aims to provide students with the opportunity to study some further topics in applied discrete mathematics. A selection of topics in discrete mathematics applied in combinatorics and optimization (such as algebraic coding theory, cryptography, discrete optimization, etc.) The selected topics may vary from year to year.

Pre-requisites: Knowledge in introductory discrete mathematics. Students may be asked to present appropriate evidence of having met the pre-requisites for enrolling in this course. Students should not be taking or have taken MATH8502 Topics in applied discrete mathematics or equivalent.

Assessment: coursework (50%) and examination (50%)

MATH6503   Topics in mathematical programming and optimization (6 credits)

A study in greater depth of some special topics in mathematical programming or optimization. It is mainly intended for students in Operations Research or related subject areas. This course covers a selection of topics which may include convex programming, nonconvex programming, saddle point problems, variational inequalities, optimization theory and algorithms suitable for applications in various areas such as machine learning, artificial intelligence, imaging and computer vision. The selected topics may vary from year to year.

Pre-requisites: Knowledge in introductory mathematical programming and optimization. Students may be asked to present appropriate evidence of having met the pre-requisites for enrolling in this course. Students should not be taking or have taken MATH8503 Topics in mathematical programming and optimization or equivalent.

Assessment: coursework (100%)

MATH8502   Topics in applied discrete mathematics (6 credits)

This course aims to provide students with the opportunity to study some further topics in applied discrete mathematics. A selection of topics in discrete mathematics applied in combinatorics and optimization (such as algebraic coding theory, cryptography, discrete optimization, etc.) The selected topics may vary from year to year.

Pre-requisites: Knowledge in introductory discrete mathematics. Students may be asked to present appropriate evidence of having met the pre-requisites for enrolling in this course. Students should not be taking or have taken MATH6502 Topics in applied discrete mathematics or equivalent.

Assessment: coursework (50%) and examination (50%)

MATH8503   Topics in mathematical programming and optimization (6 credits)

A study in greater depth of some special topics in mathematical programming or optimization. It is mainly intended for students in Operations Research or related subject areas. This course covers a selection of topics which may include convex programming, nonconvex programming, saddle point problems, variational inequalities, optimization theory and algorithms suitable for applications in various areas such as machine learning, artificial intelligence, imaging and computer vision. The selected topics may vary from year to year.

Pre-requisites: Knowledge in introductory mathematical programming and optimization. Students may be asked to present appropriate evidence of having met the pre-requisites for enrolling in this course. Students should not be taking or have taken MATH6503 Topics in mathematical programming and optimization or equivalent.

Assessment: coursework (100%)

STAT6008   Advanced statistical inference (6 credits)

This course covers the advanced theory of point estimation, interval estimation and hypothesis testing. Using a mathematically-oriented approach, the course provides a formal treatment of inferential problems, statistical methodologies and their underlying theory. It is suitable in particular for students intending to further their studies or to develop a career in statistical research. Contents include: (1)Decision problem – frequentist approach: loss function; risk; decision rule; admissibility; minimaxity; unbiasedness; Bayes' rule; (2)Decision problem – Bayesian approach: prior and posterior distributions, Bayesian inference; (3) Estimation theory: exponential families; likelihood; sufficiency; minimal sufficiency; completeness; UMVU estimators; information inequality; large-sample theory of maximum likelihood estimation; (4) Hypothesis testing: uniformly most powerful (UMP) test; monotone likelihood ratio; UMP unbiased test; conditional test; large-sample theory of likelihood ratio; confidence set; (5) Nonparametric inference; bootstrap methods.

Assessment: coursework (40%) and examination (60%)

STAT6013   Financial data analysis (6 credits)

This course aims at introducing statistical methodologies in analyzing financial data. Financial applications and statistical methodologies are intertwined in all lectures. Contents include: recent advances in modern portfolio theory, Copula, market microstructure and high frequency data analysis, FinTech applications with various computational tools such as artificial neural networks, Kalman filters and blockchain data analysis.

Assessment: coursework (40%) and examination (60%)

STAT6015   Advanced quantitative risk management and finance (6 credits)

This course covers statistical methods and models of importance to risk management and finance and links finance theory to market practice via statistical modelling and decision making. Emphases will be put on empirical analyses to address the discrepancy between finance theory and market data. Contents include: Elementary Stochastic Calculus; Basic Monte Carlo and Quasi-Monte Carlo Methods; Variance Reduction Techniques; Simulating the value of options and the value-at-risk for risk management; Review of univariate volatility models; multivariate volatility models; Value-at-risk and expected shortfall; estimation, back-testing and stress testing; Extreme value theory for risk management.

Assessment: coursework (25%) and examination (75%)

STAT6016   **Spatial data analysis (6 credits)**

This course covers statistical concepts and tools involved in modelling data which are correlated in space. Applications can be found in many fields including epidemiology and public health, environmental sciences and ecology, economics and others. Covered topics include: (1) *Outline* of three types of spatial data: point-level (geostatistical), areal (lattice), and spatial point process. (2) *Model-based geostatistics*: covariance functions and the variogram; spatial trends and directional effects; intrinsic models; estimation by curve fitting or by maximum likelihood; spatial prediction by least squares, by simple and ordinary kriging, by trans-Gaussian kriging. (3) *Areal data models*: introduction to Markov random fields; conditional, intrinsic, and simultaneous autoregressive (CAR, IAR, and SAR) models. (4) *Hierarchical modelling* for univariate spatial response data, including Bayesian kriging and lattice modelling. (5) *Introduction* to simple spatial point processes and spatio-temporal models. Real data analysis examples will be provided with dedicated R packages such as geoR.

Assessment: coursework (50%) and examination (50%)

STAT6019   **Current topics in statistics (6 credits)**

This course includes two modules.

The first module, *Causal Inference*, is an introduction to key concepts and methods for causal inference. Contents include 1) the counterfactual outcome, randomized experiment, observational study; 2) Effect modification, mediation and interaction; 3) Causal graphs; 4) Confounding, selection bias, measurement error and random variability; 5) Inverse probability weighting and the marginal structural models; 6) Outcome regression and the propensity score; 7) The standardization and the parametric g-formula; 8) G-estimation and the structural nested model; 9) Instrumental variable method; 10) Machine learning methods for causal inference; 11) Other topics as determined by the instructor.

The second module, *Posterior Inference and Simulation*, cover topics from: 1) Large-sample properties of posterior distribution; 2) Langevin dynamics and Hamiltonian MCMC; 3) Sequential Monte Carlo methods; 4) Approximation Bayesian computation; 5) Variational Bayesian methods; 6) Other topics as determined by the instructor.

Assessment: coursework (25%) and examination (75%)

STAT8003   **Time series forecasting (6 credits)**

A time series consists of a set of observations on a random variable taken over time. Such series arise naturally in climatology, economics, finance, environmental research and many other disciplines. In additional to statistical modelling, the course deals with the prediction of future behaviour of these time series. This course distinguishes different types of time series, investigates various representations for them and studies the relative merits of different forecasting procedures.

Assessment: coursework (40%) and examination (60%)

STAT8017   **Data mining techniques (6 credits)**

With the rapid developments in computer and data storage technologies, the fundamental paradigms of classical data analysis are mature for change. Data mining techniques aim at helping people to work smarter by revealing underlying structure and relationships in large amounts of data. This course takes a practical approach to introduce the new generation of data mining techniques and show how to use them to make better decisions. Topics include data preparation, feature selection, association rules, decision trees, bagging, random forests and gradient boosting, cluster analysis, neural networks, introduction to text mining.

Assessment: coursework (100%)

STAT8019　　Marketing analytics (6 credits)

This course aims to introduce various statistical models and methodology used in marketing research. Special emphasis will be put on marketing analytics and statistical techniques for marketing decision making including market segmentation, market response models, consumer preference analysis and conjoint analysis. Contents include market response models, statistical methods for segmentation, targeting and positioning, statistical methods for new product design.

Assessment:　　coursework (40%) and examination (60%)

STAT8021　　Big data analytics (6 credits)

The recent explosion of social media and the computerization of every aspect of life resulted in the creation of volumes of mostly unstructured data (big data): web logs, e-mails, video, speech recordings, photographs, tweets and others. This course aims to provide students with knowledge and skills of some advanced analytics and statistical modelling for solving big data problems. Students are required to possess basic understanding of R language.

Pre-requisites: Pass in STAT8017 Data mining techniques or equivalent; students should not be taking or have taken STAT8301 Big data analytics or STAT8307 Natural language processing and text analytics or equivalent

Assessment:　　coursework (100%)

STAT8301　　Big data analytics (3 credits)

The recent explosion of social media and the computerization of every aspect of life resulted in the creation of volumes of mostly unstructured data (big data): web logs, e-mails, Tweets, and others. This course aims to provide students with knowledge and skills of some advanced analytics and statistical modelling for solving big data problems. Topics may be selected from the following areas: recommendation systems, collaborative filtering, Non-negative matrix factorization, text analytics, natural language processing, topic modeling, and sentiment analytics.

Pre-requisites: Pass in STAT8017 Data mining techniques or equivalent; students should not be taking or have taken STAT8021 Big data analytics or equivalent

Assessment:　　coursework (50%) and examination (50%)

STAT8306　　Statistical methods for network data (3 credits)

The six degree of separation theorizes that human interactions could be easily represented in the form of a network. Examples of networks include router networks, the World Wide Web, social networks (e.g. Facebook or Twitter), genetic interaction networks and various collaboration networks (e.g. movie actor coloration network and scientific paper collaboration network). Despite the diversity in the nature of sources, the networks exhibit some common properties. For example, both the spread of disease in a population and the spread of rumors in a social network are in sub-logarithmic time. This course aims at discussing the common properties of real networks and the recent development of statistical network models. Topics may include common network measures, community detection in graphs, preferential attachment random network models, exponential random graph models, models based on random point processes and the hidden network discovery on a set of dependent random variables.

Assessment:　　coursework (50%) and examination (50%)

STAT8307   Natural language processing and text analytics (3 credits)

The textual data constitutes an enormous proportion of unstructured data which is characterized as one of 'V's in Big Data. The logical and computational reasonings are applied to transform large collection of written resources to structured data for use in further analysis, visualization, integration with structured data in database or warehouse, and further refinement using machine learning systems. This course introduces the methodology of text analytics. Topics include natural language processing, word representation, text categorization and clustering, topic modelling and sentiment analysis. Students are required to possess basic understanding of Python language.

Pre-requisites: Pass in STAT8017 Data mining techniques and DASC7606 Deep learning or equivalent; students should not be taking or have taken STAT8021 Big data analytics or STAT8301 Big data analytics or equivalent

Assessment: coursework (100%)

Capstone Requirement

DASC7600   Data science project (12 credits)

Candidate will be required to carry out independent work on a major project under the supervision of individual staff member. A written report is required.

Assessment: written report (75%) and oral presentation (25%)