

**REGULATIONS FOR THE DEGREE OF  
MASTER OF STATISTICS  
(MStat)**

[This syllabus is applicable to students admitted to the programme in the academic year 2025-26 and thereafter.]

*(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**

**MS1** To be eligible for admission to the courses leading to the degree of Master of Statistics a candidate shall

- (a) comply with the General Regulations and the Regulations for Taught Postgraduate Curricula;
- (b) hold
  - (i) a Bachelor's degree 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) pass a qualifying examination if so required.

---

**Period of study**

**MS2** The curriculum shall normally extend over one academic year of full-time study or two academic years of part-time study. Candidates shall not be permitted to extend their studies beyond the maximum period of registration of two academic years of full-time study or three academic years of part-time study, unless otherwise permitted or required by the Board of Studies.

---

**Exemption**

**MS3** A candidate with appropriate qualification and professional experiences may, on production of appropriate certification of having satisfactorily completed another course or courses equivalent in content to any of the two compulsory courses as specified in the syllabuses, be exempted from the compulsory course(s), subject to approval of the Board of Studies. 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**

**MS4** To be eligible for the award of the degree of Master of Statistics, a candidate shall

- (a) comply with the General Regulations and the Regulations for Taught Postgraduate Curricula;
- (b) successfully complete the curriculum in accordance with the regulations set out below; and
- (c) have achieved a cumulative grade point average of at least 2.0.

Any student who does not meet the exit requirement 4(c) has neither fulfilled the requirements for graduation or discontinuation. Therefore, the students concerned would be allowed to continue their study until the end of the maximum study period as outlined in the regulations.

---

## Completion of curriculum

**MS5** 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

- MS6**
- (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.
- 

**MS7** 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 MS8, 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.
- 

**MS8** 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 seven calendar days of the examination concerned.

---

- MS9** A candidate who
- (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

may be required to discontinue his/her studies.

---

## Grading system

**MS10** Individual courses shall be graded according to one of the following grading systems as determined by the Board of Examiners:

(a) Letter grades, their standards and the grade points for assessment are as follows:

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

or

(b) 'Distinction', 'Pass' or 'Fail'.

Courses which are graded according to (b) above will not be included in the calculation of the GPA.

---

## Assessment results

**MS11** 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 STATISTICS

The Master of Statistics postgraduate programme offered by the School of Computing and Data Science allows students to enroll in either a full-time one-year mode or a part-time two-year mode. The curriculum is designed to provide graduates with training in the principles and practice of statistics. Candidates should have knowledge of matrices and calculus, introductory statistics and linear modelling.

The curriculum offers great flexibilities for students who wish to take a general approach or a specialized theme in Risk Management, Data Analytics, or Financial Statistics. A student may choose to have his/her theme printed on the transcript if he/she has satisfied one of the theme requirements.

### STRUCTURE AND EVALUATION

Each student must complete at least 60 credits of courses. If a student selects an MStat course whose contents are similar to a course (or courses) which he/she has taken in his/her previous study, the School may not approve the selection in question.

### CURRICULUM

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

<b>Two compulsory courses (12 credits)</b>	
STAT7101 Fundamentals of statistical inference (6 credits)	
STAT7102 Advanced statistical modelling (6 credits)	
<b>Theme-specific disciplinary elective courses (24 Credits)</b>	
<b><u>Risk Management theme</u></b> plus 24 credits from	<b><u>Data Analytics theme</u></b> plus 24 credits from
STAT6015 Advanced quantitative risk management (6 credits)	STAT6008 Advanced statistical inference (6 credits)
STAT6017 Operational risk and insurance analytics (6 credits)	STAT6011 Computational statistics and Bayesian learning (6 credits)
STAT7009 Stochastic dependence modelling (6 credits)	STAT6016 Spatial data analysis (6 credits)
STAT8003 Time series forecasting (6 credits)	STAT7005 Multivariate methods (6 credits)
STAT8007 Statistical methods in economics and finance (6 credits)	STAT7007 Categorical data analysis (3 credits)
STAT8015 Actuarial statistics (6 credits)	STAT8003 Time series forecasting (6 credits)
STAT8017 Data mining techniques (6 credits)	STAT8016 Biostatistics (6 credits)
STAT8308 Blockchain data analytics (3 credits)	STAT8017 Data mining techniques (6 credits)
	STAT8019 Marketing analytics (6 credits)
	STAT8302 Structural equation modelling (3 credits)
	STAT8306 Statistical methods for network data (3 credits)
<b><u>Financial Statistics theme</u></b> plus 24 credits from	STAT8307 Natural language processing and text analysis (6 credits)
STAT6013 Financial data analysis (6 credits)	
STAT7009 Stochastic dependence modelling (6 credits)	
STAT8003 Time series forecasting (6 credits)	
STAT8007 Statistical methods in economics and finance (6 credits)	
STAT8015 Actuarial statistics (6 credits)	
STAT8017 Data mining techniques (6 credits)	
STAT8020 Quantitative strategies and algorithmic trading (6 credits)	
STAT8309 Monte Carlo simulation in finance (6 credits)	
FITE7405 Techniques in computational finance (6 credits)	

**Other disciplinary elective courses (18 credits)\*****plus** at least **18** credits\* from

- STAT6009 Research methods in statistics (6 credits)
- STAT6010 Advanced probability (6 credits)
- STAT6019 Current topics in statistics (6 credits)
- STAT7006 Design and analysis of sample surveys (6 credits)
- STAT7008 Programming for data science (6 credits)
- STAT8000 Workshop on spreadsheet modelling and database management (3 credits)
- STAT8300 Career development and communication workshop (Non-credit-bearing)
- Any theme-specific disciplinary elective courses

**Capstone requirement (6 credits)****plus 6** credits from

- STAT8017 Data mining techniques (6 credits)
- STAT8088 Statistical practicum (6 credits)
- STAT8089 Capstone project (6 credits)

Apart from the two compulsory courses and capstone requirement, candidates may choose not to follow any theme and may take 42 credits of disciplinary elective courses in any order, whenever feasible.

\* Students may also in exceptional circumstances select other postgraduate level courses (at most 12 credits in total) offered by the School of Computing and Data Science to partially meet the disciplinary electives courses requirement. All course selection will be subject to approval by the Master Programmes Associate Director/Course Coordinator concerned.

**Compulsory Course Replacement**

<b>Students with prior background may replace each compulsory course with a more advanced course as shown below:</b>	
<b>Replace ...</b>	<b>With</b>
STAT7101 Fundamentals of statistical inference	STAT6008 Advanced statistical inference or STAT6009 Research methods in statistics or STAT7005 Multivariate methods
STAT7102 Advanced statistical modelling	Any other course

## COURSE DESCRIPTION

---

### Compulsory Courses

---

#### **STAT7101 Fundamentals of statistical inference (6 credits)**

Motivated by real problems involving uncertainty and variability, this course introduces the basic concepts and principles of statistical inference and decision-making. Contents include: large-sample theories; estimation theory; likelihood principle; maximum likelihood estimation; hypotheses testing; likelihood ratio tests; computer-intensive methods such as EM algorithm. (Only under exceptional academic circumstances can this compulsory course be replaced by an elective course.)

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

---

#### **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) Model selection and regularization; (iv) Kernel and local polynomial regression; selection of smoothing parameters; (v) Generalized additive models; (vi) Hidden Markov models and Bayesian networks.

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

---

### Elective Courses

---

#### **FITE7405 Techniques in computational finance (6 credits)**

This course introduces the major computation problems in the field of financial derivatives and various computational methods/techniques for solving these problems. The lectures start with a short introduction on various financial derivative products, and then move to the derivation of the mathematical models employed in the valuation of these products, and finally come to the solving techniques for the models.

Pre-requisites: No prior finance knowledge is required. Students are assumed to have basic competence in calculus and probability (up to the level of knowing the concepts of random variables, normal distributions, etc.). Knowledge in at least one programming language is required for the assignments/final project.

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

---

#### **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%)

---

### **STAT6009 Research methods in statistics (6 credits)**

This course introduces some statistical concepts and methods which potential graduate students will find useful in preparing for work on a research degree in statistics. Focus is on applications of state-of-the-art statistical techniques and their underlying theory. Contents may be selected from: (1) Basic asymptotic methods: modes of convergence; stochastic orders; laws of large numbers; central limit theorems; delta method; (2) Parametric and nonparametric likelihood methods: high-order approximations; profile likelihood and its variants; signed likelihood ratio statistics; empirical likelihood; (3) Nonparametric statistical inference: sign and rank tests; Kolmogorov-Smirnov test; nonparametric regression; density estimation; kernel methods; (4) Computationally-intensive methods: cross-validation; bootstrap; permutation methods; (5) Robust methods: measures of robustness; M-estimator; L-estimator; R-estimator; estimating functions; (6) Other topics as determined by the instructor.

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

---

### **STAT6010 Advanced probability (6 credits)**

This course provides a mathematical introduction to measure theory and probability theory to equip students with the necessary mathematical concepts important for research in probability, statistics and actuarial science. Contents include sigma-algebras, measurable spaces, measures and probability measures, independence and dependence, random variables, integration theory, modes of convergence, characteristic functions, conditional expectations and more.

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

---

### **STAT6011 Computational statistics and Bayesian learning (6 credits)**

This course aims to give students an introduction on modern computationally intensive methods in statistics, with a strong focus on Bayesian methods. The role of computation as a fundamental tool in data analysis and statistical inference will be emphasized. The course will introduce topics including the generation of random variables, optimization techniques, and numerical integration using quadrature and Monte Carlo methods. This course will then cover the fundamental Bayesian framework, including prior elicitation, posterior inference and model selection. For posterior computation, Monte Carlo methods such as importance sampling and Markov chain Monte Carlo will be introduced. Methods for approximate inference such as variational Bayes will also be covered. Advanced Bayesian modeling with nonparametric Bayes will then be explored, with applications in machine learning. This course is particularly suitable for students who intend to pursue further studies or a career in research.

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

---

### **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: classical portfolio theory, portfolio selection in practice, single index market model, robust parameter estimation, copula and high frequency data analysis.

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

---

### **STAT6015 Advanced quantitative risk management (6 credits)**

This course covers statistical methods and models of risk management, especially of Value-at-Risk (VaR). Contents include: Value-at-risk (VaR) and Expected Shortfall (ES); univariate models (normal model, log-normal model and stochastic process model) for VaR and ES; models for portfolio VaR; time series models for VaR; extreme value approach to VaR; back-testing and stress testing.

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

---

### **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%)

---

### **STAT6017 Operational risk and insurance analytics (6 credits)**

This course aims to provide the foundation of operational risk management and insurance. Special emphasis will be put on the analytical and modeling techniques for operational risk and insurance. Contents include fundamentals of operational risk and Basel regulation, loss distribution, estimation of risk models, copula and modeling dependence, insurance and risk transfer for operational risk.

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

---

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

This course may include modules such as:

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.

Functional data analysis, covers topics from: 1) Base functions; 2) Least squares estimation; 3) Constrained functions; 4) Functional PCA; 5) Regularized PCA; 6) Functional linear model; 7) Other topics as determined by the instructor.

Assessment: coursework (100%)

---

### **STAT7005 Multivariate methods (6 credits)**

In many disciplines the basic data on an experimental unit consist of a vector of possibly correlated measurements. Examples include the chemical composition of a rock; the results of clinical observations and tests on a patient; the household expenditures on different commodities. Through

the challenge of problems in a number of fields of application, this course considers appropriate statistical models for explaining the patterns of variability of such multivariate data. Topics include: multiple, partial and canonical correlation; multivariate regression; tests on means for one-sample and two-sample problems; profile analysis; test for covariances structure; multivariate ANOVA; principal components analysis; factor analysis; discriminant analysis and classification.

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

---

### **STAT7006 Design and analysis of sample surveys (6 credits)**

Inferring the characteristics of a population from those observed in a sample from that population is a situation often forced on us for economic, ethical or technological reasons. This course considers the basic principles, practice and design of sampling techniques to produce objective answers free from bias. This course will cover design and implementation of sample surveys and analysis of statistical data thus obtained. Survey design includes overall survey design, design of sampling schemes and questionnaires, etc. Sampling methods include sample size determination, sampling and non-sampling errors and biases, methods of estimation of parameters from survey data, imputation for missing data etc.

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

---

### **STAT7007 Categorical data analysis (3 credits)**

Many social and medical studies, especially those involving questionnaires, contain large amounts of categorical data. Examples of categorical data include presence or absence of disease (yes / no), mode of transportation (bus, taxi, railway), attitude toward an issue (strongly disagree, disagree, agree, strongly agree). This course focuses on analyzing categorical response data with emphasis on hands-on training of analyzing real data using statistical software SAS. Consulting experience may be presented in the form of case studies. Topics include: classical treatments of contingency tables; measures of association; logistic linear models and log-linear models for binary responses; and log-linear models for Poisson means.

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

---

### **STAT7008 Programming for data science (6 credits)**

Capturing and utilising essential information from big datasets poses both statistical and programming challenges. This course is designed to equip students with the fundamental computing skills required to use Python for addressing these challenges. The course will cover a range of topics, including programming syntax, files IO, object-oriented programming, scientific data processing and analysis, data visualization, data mining and web scraping, programming techniques for machine learning, deep learning, computer vision, and natural language processing, etc.

Assessment: coursework (100%)

---

### **STAT7009 Stochastic dependence modelling (6 credits)**

This course provides an introduction to the stochastic modelling of dependence. Topics include univariate distribution functions, quantile functions, multivariate distribution functions, copulas, properties of copulas, simulation, measures of association, copula constructions and selected further aspects of copula modelling.

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

---

### **STAT8000 Workshop on spreadsheet modelling and database management (3 credits)**

This course aims to enhance students' IT knowledge and skills which are essential for career development of statistical and risk analysts. The course contains a series of computer hands-on workshops on Excel VBA programming, MS-Access and SQL and C++ basics.

Assessment: coursework (100%), assessment of this course is on a pass or fail or distinction basis

---

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

Discrete time series are integer-indexed sequences of random variables. Such series arise naturally in climatology, economics, finance, environmental research and many other disciplines. The course covers the statistical modelling and forecasting of time series. Topics may include stationary and nonstationary time series, ARMA models, identification based on autocorrelation and partial autocorrelation, GARCH models, goodness-of-fit, forecasting, and nonlinear time series modelling.

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

---

### **STAT8007 Statistical methods in economics and finance (6 credits)**

This course provides a comprehensive introduction to state-of-the-art statistical techniques in economics and finance, with emphasis on their applications to time series and panel data sets in economics and finance. Topics include: regression with heteroscedastic and/or autocorrelated errors; instrumental variables and two stage least squares; panel time series model; unit root tests, co-integration, error correction models; and generalized method of moments.

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

---

### **STAT8015 Actuarial statistics (6 credits)**

The main focus of this module will be on financial mathematics of compound interest with an introduction to life contingencies and statistical theory of risk. Topics include simple and compound interest, annuities certain, yield rates, survival models and life tables, population studies, life annuities, assurances and premiums, reserves, joint life and last survivor statuses, multiple decrement tables, expenses, individual and collective risk theory.

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

---

### **STAT8016 Biostatistics (6 credits)**

Statistical methodologies and applications in fields of medicine, clinical research, epidemiology, public health, biology and biomedical research are considered. The types of statistical problems encountered will be motivated by experimental data sets. Important topics include design and analysis of randomized clinical trials, group sequential designs and crossover trials; survival studies; diagnosis; risks; statistical analysis of the medical process.

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, conjoint analysis and extracting insights from text data. Contents include statistical methods for segmentation, targeting and positioning, statistical methods for new product design, text mining techniques and market response models.

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

---

**STAT8020 Quantitative strategies and algorithmic trading (6 credits)**

Quantitative trading is a systematic investment approach that consists of identification of trading opportunities via statistical data analysis and implementation via computer algorithms. This course introduces various methodologies that are commonly employed in quantitative trading.

The first half of the course focuses at strategies and methodologies derived from the data snapshotted at daily or minute frequency. Some specific topics are: (1) techniques for trading trending and mean-reverting instruments, (2) statistical arbitrage and pairs trading, (3) detection of “time-series” mean reversion or stationarity, (4) cross-sectional momentum and contrarian strategies, (5) back-testing methodologies and corresponding performance measures, and (6) Kelly formula, money and risk management. The second half of the course discusses statistical models of high frequency data and related trading strategies. Topics that planned to be covered are: (7) introduction of market microstructure, (8) stylized features and models of high frequency transaction prices, (9) limit order book models, (10) optimal execution and smart order routing algorithms, and (11) regulation and compliance issues in algorithmic trading.

Pre-requisites: Pass in STAT6013 Financial data analysis or equivalent

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

---

**STAT8088 Statistical practicum (6 credits)**

It provides students with first-hand experience in applying academic knowledge in a real-life work environment. To be eligible, students should be undertaking a statistics-related or risk-management-related practicum with no less than 160 hours in at least 20 working days spent in a paid or unpaid position. It is allowed for part-time students to complete their practicum within their current place of employment. The practicum will normally take place in the second semester or summer semester for full-time students or during the second year for part-time students.

Note: Students should not be taking or have taken STAT8089 Capstone project

Assessment: Upon completion of the practicum, each student is required to submit a written report (60%) and to give an oral presentation (40%) on his/her practicum experience. Supervisors will assess the students based on their performance during the practicum period. Assessment of this course is on a Pass, Fail or Distinction basis with three criteria: (1) supervisor's evaluation, (2) written report, (3) oral presentation. Failing in fulfilling any of the three criteria satisfactorily leads to a "Fail" grade in the course.

---

**STAT8089 Capstone project (6 credits)**

This project-based course aims to provide students with capstone experience to work on a real-world problem and carry out a substantial data analysis project which requires integration of the knowledge they have learnt in the curriculum. Students will work in small groups under the guidance of their supervisor(s). The project topic is not limited to academic context, but can also be extended to a community or corporate outreach project. Students will need to find an interesting topic of their own, conduct literature search regarding the most recent research related to the problem, make suggestions to improve the current situations or even solve the problem identified in their project. A substantial written report is required.

Note: Students should not be taking or have taken STAT8088 Statistical practicum

Assessment: project proposal (15%); written report (55%) and oral presentation (30%)

---

### **STAT8300 Career development and communication workshop (Non-credit-bearing)**

The course is specially designed for students who wish to sharpen their communication and career preparation skills through a variety of activities including but not limited to lectures, skill-based workshops, small group discussion, mock interviews, role plays. All of which aim to (i) facilitate students in making informed career choices; (ii) provide students with practical training to enrich communication, presentation, time management and advanced interview skills. The course is designed to enhance students' overall competitiveness in the employment markets and the readiness in job hunting.

Assessment: coursework (100%), assessment of this course is on a pass or fail or distinction basis

---

### **STAT8302 Structural equation modelling (3 credits)**

Structural Equation Modelling (SEM) is a general statistical modelling technique to establish relationships among variables. A key feature of SEM is that observed variables are understood to represent a small number of "latent constructs" that cannot be directly measured, only inferred from the observed measured variables. This course covers the theories of structural equation models and their applications. Topics may include path models, confirmatory factor analysis, structural equation models with latent variables, Sub-models including multiple group analysis, MIMIC model, second order factor analysis, two-wave model, and simplex model, model fitness, model identification, and Comparison with competing models.

Pre-requisites: Pass in STAT7005 Multivariate methods or equivalent

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

---

### **STAT8306 Statistical methods for network data (3 credits)**

The six degrees 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 (100%)

---

**STAT8307 Natural language processing and text analytics (6 credits)**

Natural Language Processing (NLP) is a core area of artificial intelligence and data science that focuses on enabling machines to understand, process, and generate human language. This course introduces fundamental NLP techniques and text analytics methods, covering topics such as information retrieval, text classification, word embeddings, neural networks, sequence models, encoder-decoder architectures, transformers, contextualized word representations, and modern language models. The course emphasizes practical applications and hands-on experience in analyzing textual data. Proficiency in Python programming is required.

Pre-requisites: Pass in STAT7008 Programming for data science or equivalent [for MStat and MDASC students]

Pass in ARIN7102 Applied data mining and text analytics or equivalent [for MSc(AI) students]

Assessment: coursework (100%)

---

**STAT8308 Blockchain data analytics (3 credits)**

In this course, we start by studying the basic architecture of a blockchain. Then we move on to several major applications including (but not limited to) cryptocurrencies, fintech and smart contracts. We conclude by examining the cybersecurity issues facing the blockchain ecosystems.

Assessment: coursework (100%)

---

**STAT8309 Monte Carlo simulation in finance (6 credits)**

This course covers basic knowledge of Monte Carlo simulation and its application to finance. Course contents include random variate generation, Monte Carlo methods, simulation stock prices under different models, risk neutral probabilities, evaluation of European options, evaluation of American options, evaluation of Asian options, variance reduction techniques and efficiency improvement.

Assessment: coursework (100%)