UNIVERSITY OF CAPE TOWN
DEPARTMENT OF STATISTICAL SCIENCES
STATISTICS HONOURS MODULE INFORMATION 2017

THEORY OF STATISTICS A

This module is intended for students with undergraduate training in mathematical statistics and comprises of the following four sections (12 lectures each).

1. Advanced Stochastic Processes (Dr Miguel Lacerda)

   This section covers renewal theory. A renewal process is a counting process, $\{N(t), t \geq 0\}$, where the times between renewals are identically and independently distributed according to some distribution. The Poisson process (with exponentially distributed inter-arrival times) is a specific example of a renewal process.

   Content:

   - The distribution of the random variable $N(t)$
   - The renewal function and its attributes (including the elementary renewal theorem)
   - Renewal reward processes
   - Regenerative processes
   - Alternating renewal processes
   - The inspection paradox

   Assessment:

One class assignment	30%
Final examination	70%

2. Probability Theory (Mr Melusi Mavuso)

   This module covers the basic foundations of probability and measure theory.

   Content:

   - Set Theory
   - σ algebras, π and λ systems
   - Measures, probability measures, Lebesgue measure

- Random variables and measurable functions
- Integration: Beppo-Levi Theorem, layer cake representation, Expectation
- Radon-Nikodym Theorem and derivative, Probability distributions
- Product measures and independence

Assessment:

Tutorials	30%
Final examination	70%

3. Likelihood Theory (Dr Birgit Erni)

   - The likelihood: definition, estimation, properties
   - Multiparameter models and model selection

4. Computational Bayesian Inference (Mr Chun-Kai Huang)

   The rise in computing power has allowed for the introduction of several computationally intensive statistical procedures. This has helped to revolutionise the field of statistics over the past few decades. This module will cover a number of standard computational Bayesian techniques used in statistical analysis. This will include Acceptance-Rejection sampling, Gibbs sampling and Markov Chain Monte Carlo methods. The course is intended to be practical in nature with a large practical component.

THEORY OF STATISTICS B

This module is intended for students with undergraduate training in applied statistics and comprises of the following four sections (12 lectures each).

1. Generalised Linear Models (Assoc Prof Francesca Little)

   An introduction to GLMs and the underlying theoretical concepts, including:

   - The exponential family of distributions
   - Estimation
   - Inference

2. Introduction to Stochastic Processes (Mr Dominique Katshunga)

   - Examples of stochastic processes
   - Markov chains
   - Transition possibility matrix
   - Time homogeneity
   - Classification of states
   - Probability models for transition
   - Stationarity

3. Likelihood Theory (Dr Birgit Erni)

   - The likelihood: definition, estimation, properties
   - Multiparameter models and model selection

4. Introduction to Bayesian Inference (Mr Chun-Kai Huang)

   This module gives a basic overview of the philosophy and methods of Bayesian analysis. Key Bayesian concepts are introduced and it is shown how they may be used to conduct statistical inference under a Bayesian framework. The Bayesian analysis is contrasted with the traditional frequentist approach.

   Content:

   - Philosophy
   - Bayesian concepts
   - Hypothesis Testing: the Bayesian approach
   - Bayesian Estimation
   - The Acceptance/Rejection algorithm

   Assessment:

One class assignment	30%
Final examination	70%

OPERATIONS RESEARCH A

This module is intended for students with who have not completed STA3036S or equivalent.

Lecturers: Dr Sheetal Silal, Dr Juwa Nyirenda

Content: Definitions of operations research, linear programming (formulation, solving in Excel, sensitivity, network and project management models), queuing theory, simulation.

Assessment: Three class assignments (1/3) and one exam (2/3).

OPERATIONS RESEARCH B

This module is intended for students who have completed STA3036S or Operations Research A.

Lecturers: Dr Sheetal Silal, Dr Juwa Nyirenda

Content: Linear programming (multi-objective and goal programming, duality, data envelopment analysis, integer programming), deterministic and stochastic dynamic programming, simulation.

Assessment: Two class assignments (1/3) and one exam (2/3).

MATRIX METHODS

Lecturer: Mr Stefan Britz

Content: Basic matrix operations, determinants, inverses, rank, generalised inverses and linear equations, eigenvalues and eigenvectors

---

STATISTICAL COMPUTING

- The R programming language (Dr Birgit Erni) – 24 lectures
- Visual Basic for Applications (Prof Graham Barr) – 12 lectures

---

BIOSTATISTICS

This course is an introduction to the analysis of data from medical research. Students will be prepared for entry positions as bio-statisticians and for postgraduate research degrees in medical statistics. The course could be called "Applied generalized mixed effect models" and though the applications taught in this course will refer to medical data, the methods have much wider application. The emphasis of the course will be on the application and interpretation of the different statistical methods.

Lecturers: Dr Freedom Gumedze and Dr Greg Distiller

Content:

- Design of Studies
- Introduction to programming in STATA/ R for biostatistics
- Measures of disease frequency and effect
- GLMs, with specific emphasis on Logistic Regression
- Analysis of Longitudinal Data, including GEE models and mixed effect models
- Survival Analysis

Assessment:

Two class assignments	50%
Final examination	50%

A minimum of 40% is required for the class assignment component and for the exam in order to pass the module.

ECONOMETRICS

This is a course in econometrics with an applied orientation and makes intensive utilisation of econometric computer software (Eviews). Forms part of the triad of honours modules (together with Portfolio Theory and Time Series Analysis) intended for those wanting to pursue a financial career.

Lecturer: Prof Graham Barr

Content:

Causality. Passive observation versus controlled experiments. What we can do (and can't do) with statistical analysis. The role of regression and the theory of the regression model and its attendant assumptions. Heteroscedasticity and autocorrelation. Multicollinearity and mis-specification. Time series regression. Econometric modelling. Single-equation and simultaneous equation modelling. Applications to the SA financial and economic sector.

Assessment:

One class assignment	30%
Final examination	70%

PORTFOLIO THEORY

Prescribed text: Investments by Brodie *et al.*

Content:

- Portfolio Theory and Practice: Return and risk from the historical record; Risk aversion and capital allocation to risky assets; optimal risky portfolios, index model
- Equilibrium in Capital Markets
- Fixed-Income Securities
- Applied Portfolio Management

Assessment:

Two class assignments	25%
Final examination	75%

TIME SERIES ANALYSIS

This module may not be taken by students who have completed STA3041F.

Time series data differs from data used when undertaking regression analysis since the data is not collected at one point in time. The data is collected at different time points and thus requires specialised techniques to understand the underlying generating process governing it. In this course, we will specifically focus on the analysis of time series. The techniques learnt are particularly useful for the analysis of financial and economic data. We will cover topics such as model building, estimation, prediction/forecasting, volatility modelling (useful when undertaking risk management and options trading) and cointegration (the technique can be used to uncover long-term economic relationships that may be useful for traders (equity readers specifically) and asset managers). The course is not intended to be overly theoretical and practical applications will be emphasised. Eviews (an econometrics and time series package) will be used during the practical sessions and for assignments.

Lecturer: Prof Graham Barr

Content:

- Stationarity, simple time series models, random walks
- Unit root tests
- The Box Jenkins methodology – building simple time series models
- Forecasting
- Volatility models – ARCH and GARCH models
- Bivariate and Multivariate Cointegration

Assessment:

One class assignment	30%
Final examination	70%

DECISION MODELLING

The aim of this module is to develop an understanding of human preferences and subjective judgements, for purposes of constructing models for decision support, within which we can apply quantitative tools studied elsewhere. The module thus provides a bridge between technical statistical and operational research tools on the one hand, and interaction with our clients on the other hand (especially in the context of decision support). We study both how to represent human preferences in mathematical models, and what cognitive odeli and errors may be exhibited by clients when we elicit subjective estimates of underlying parameters.

A substantial portion of the course relates to multiple criteria decision analysis (MCDA), in which we specifically aim to capture preferences and value tradeoffs between conflicting objectives and goals (e.g., social, environmental and economic goals).

In the course, we seek to develop an understanding of the following processes:

1. How can we assist decision makers in defining and structuring decision problems in such a way that we are able to apply our quantitative analytical models to them?
2. How should we model the preferences and value judgements which are expressed by decision makers or their advisors, in order to generate maximum insight into which courses of action best satisfy these preferences?
3. How do people form and express subjective value judgements, for example regarding the relative importance of different goals or of the likelihood of certain events, which we may need to use in our models?
4. What computer-aided "decision support systems" (DSS) are appropriate for different decision making contexts?

Decision Modeling is a multi-disciplinary field and in addressing the above questions, we touch on issues in psychology, economics, information systems, and operational research. The course consists of three main parts. In the first part, we consider problems and biases that people commonly experience when they are asked to think about information or make a judgement. In the second part, we look at how we can take a general decision problem and structure it in a way that is useful for later modeling. In the Model part, we look at three broad means of Modeling the way in which people can and should make multicriteria decisions.

Lecturers: Assoc Prof Leanne Scott and Dr Ian Durbach

Assessment:

Three class assignments	40%
Final examination	60%

MULTIVARIATE STATISTICS

Multivariate analysis is broadly concerned with the analysis of data comprising observations on large numbers of variates for each of a sample of individuals. Such data are generated everywhere, industry, biology, engineering and so on, and the attendant methods of analysis have become a vital part of the statisticians tool-kit. The aim of the course is to introduce the theory and practice of multivariate analysis. The R programming language will be used in this course.

Lecturer: Assoc Prof Francesca Little and Mr Stefan Britz

Content:

- Basic Notions
- Graphics
- Tests of Significance
- The Linear Model
- Principal Components
- Factor Analysis
- Discriminant Analysis

Assessment:

Twelve small assignments	40%
Final examination	60%

ANALYTICS

This course will cover computationally-intensive statistical methods for analysing predominantly large datasets. The course will cover three broad sections: (1) High performance computing in R, (2) Supervised Learning and (3) Unsupervised Learning. In the first section, students will learn how to use R to analyse large datasets on multiple computer processors. The second section will expose students to machine learning techniques that are used to infer a classification rule based on labelled training data. The last section will cover statistical methods for classifying observations into groups where the group memberships of the training data are not known in advance.

Lecturers: Dr Miguel Lacerda, Dr Şebnem Er, Dr Juwa Nyirenda

Content:

1. High Performance Computing
    - Parallel computing in R
    - Cluster computing
    - Dealing with "big data" in R
2. Supervised Learning
    - Regression and classification trees
    - Bagging and random forests
    - Boosting
    - Neural networks
3. Unsupervised Learning
    - Self-organising maps
    - Association rule mining
    - Cluster analysis

Assessment:

Assignments	40%
Final examination	60%

STOCHASTIC CALCULUS FOR FINANCE

The aim of this course is to introduce students with a background in probability to stochastic processes, stochastic calculus and derivative pricing. The course begins with some topics in probability theory. Students are expected to be familiar with material from the Probability Theory module taught in Theory of Statistics A.

Lecturer: Mr Melusi Mavuso

Content:

- Conditional Expectation
- Convergence of sequences of random variables, Strong Law of Large numbers, CLT
- Riemann-Stieltjes integration
- Stochastic processes, Filtrations
- Martingales and stopping times, Doob decomposition, Martingale Convergence
- Brownian Motion
- Itô integrals and diffusion processes
- Stochastic differential equations
- Martingales and gambling
- Derivative pricing, the Binomial Model
- The Martingale Representation Theorem, Cameron-Martin-Girsanov Theorem
- Fundamental Theorem of Asset Pricing I and II
- The Feynman-Kac Theorem, the Black-Scholes Equation
- Pricing some exotic options
- An introduction to stochastic optimal control

Assessment:

Tutorials (30%) and exam (70%).

VISUAL THINKING AND VISUALISATION

In the age of Big Data, it is increasingly important to pay careful attention to the design of data visualisations and the tools to interact with them. This fact is highlighted by the huge success of companies such as Apple and Google, who have prioritised intuitiveness and ease-of-use in their software interface designs and interactions. In this module, we cover the field of visual thinking, outlining current understanding of how humans think visually from a neurological perspective and demonstrating how we can use this knowledge to design for more effective data displays and interaction. This knowledge will be applied in the assignment to the design of multidimensional interactive data graphics.

Lecturer: Assoc Prof Michelle Kuttel (Computer Science)

Prerequisites:

There are no specific prerequisites for this module, other than a background in computing. However, some interest in graphical displays, data and graphics/ visual art/ aesthetics/ design is required to appreciate the course content.

Content:

- Visual queries and how the mind works to process visual information
- Structuring two dimensional space
- Colour
- Visual space and time: depth perception and motion
- Visual objects: how to design visual objects that are easy to identify
- Theory and best practice in the design of multidimensonal data graphics, interfaces and visualisations.

Recommended Reading:

- Visual Thinking for Design by Colin Ware
- The Visual Display of Quantitative Information by Edward R. Tufte (2nd edition)

These recommended books are highly regarded internationally and will make wonderful additions to your library, but you do not have to buy them for this course.

Assessment:

One major assignment (50%) and an exam (50%).

The single major practical will involve multi-stage design and testing of a graphical display of multidimensional data. Topics will be listed in the first week of the course and design stages will be presented to class for discussion and critique. The practical is expected to involve about 16 hours of work.

There will be 8 lectures plus 2 long presentation sessions (total of 14 hours).

ANALYSIS OF HOUSEHOLD SURVEY DATA (ECO4027S)

This course examines a range of statistical techniques for modelling survey data and presents methods to compensate for design features of complex sample survey data. These techniques are then applied to a selection of policy issues through the analysis of South African household surveys. Students are expected to be familiar with basic econometric/statistical methods through multivariate linear regression.

Content:

1. Analysis of complex sample surveys

   Standard courses on statistical analysis assume that survey data are from a simple random sample of the target population. Little attention is given to characteristics often associated with survey data, including missing data, unequal probabilities of observation, stratified multistage sample designs, and measurement errors. Failure to take most of these properties of survey data into account can have an important impact on the results of all types of analysis, ranging from simple descriptive statistics to estimates of parameters of multivariate models. The course will cover sampling design and weights, variance estimation for complex sample surveys and consequences of non-response and missing data and methods for dealing with missing data (ignorable and non-ignorable missing data mechanisms; unit non-response adjustments through weighting and post-stratification; multiple imputation for item non-response).

2. Social policy issues and the analysis of household survey data

   Household surveys provide the most valuable information about household economic circumstances and socio-economic behaviour. The course investigates a range of interesting policy issues through the analysis of South African household surveys. In 2013 the topics will include: the measurement of poverty and inequality and social mobility; the evaluation of social policy with a particular focus on the state old age pensions impact on poverty and the effect of the pension on labour supply, and analysis of the South African labour market using panel data.

The course includes a computer lab component built around statistical analysis of household survey data using the statistical package STATA. The emphasis is on the application of a range of statistical analysis techniques to socio-economic analysis rather than on the theoretical underpinnings of such techniques.

Lecturers:

Dr Andrew Kerr, Room 3.46 New Economics Building, andrew.kerr@uct.ac.za

Mr Arden Finn, Room 3.11 New Economics Building, aj.finn@uct.ac.za

Assessment:

Two problem sets	25%
Two lab practicals	25%
Final examination	50%