**PROGRAMME SPECIFICATION FOR DIPLOMA AND MSC IN APPLIED STATISTICS**

| | |
|---|---|
| **1. Awarding institution/body** | **University of Oxford** |
| **2. Teaching institution** | **University of Oxford** |
| **3. Programme accredited by** | **University of Oxford** |
| **4. Final award** | **(a) MSc in Applied Statistics** |
| | **(b) Diploma in Applied Statistics** |
| **5. Programme** | **(a) MSc in Applied Statistics** |
| | **(b) Diploma in Applied Statistics** |
| **6. UCAS code** | |
| **7. Relevant subject benchmark statement** | |
| **8. Date of programme specification** | **November 2006** |

**9. Educational aims of the programme**

The aims of the programme are:

- learn a wide range of statistical methods (especially modern, computer intensive methods);
- gain extensive hands-on experience of the analysis of real data from a wide variety of fields
- develop the skills to interpret and communicate their results.

**10. Programme outcomes and means by which they are achieved**:

*Lectures*

- provide information to gain a full understanding of the general theory and practice of statistical analysis at an advanced level appropriate for MSc study;
- provide lectures on core topics which cover some of the fundamentals of statistics that include statistical inference, statistical methods and multivariate analysis; core material also covers modern computational aspects of statistics through courses in computer-intensive statistics and MCMC and Bayesian statistics;
- provide an extensive menu of optional topics on further statistical methodology and applications including for example courses in statistical genetics, disease modelling, combinatorial optimisation and actuarial science;
- provide non-examinable skills support lectures on statistical computing, report writing, and LaTeX document production.

*Case studies*

- enable students to present their review and analysis of a statistical problem or research paper as a seminar to the MSc group and academic staff.

*Guided reading*

- recommended reading is provided for all modules of the course in advance in a student handbook.

*Course assignments and example classes*

- assignments are provided to further understanding and extend knowledge in modules, together with example classes covering problem solving.

*Practical Sessions*

- enable students to undertake practical statistical data analysis that complement lectures on almost all core topics and some optional topics;
- enable students to learn statistical computing skills using modern statistical software such as S-PLUS and R;
- enable students to learn to write a report on the statistical analysis of data.

*Dissertation (MSc, but not Diploma students)*

- enable students to undertake an in-depth study of a statistical problem involving modelling, computing and data analysis; almost all dissertations contain an account of the analysis of some body of real data;
- enable students to learn to undertake directed research, report writing, presentation and communication of research results.

*Seminar series*

- the department has a weekly seminar series with external invited speakers which students are encouraged to attend to learn current research in statistics. In addition there are many specialist seminar series that students may attend, according to their interests.

**11. Programme Structures and Features**

*Lecture courses*

Coursework is divided into core and optional topics. Core topics cover some of the fundamentals of statistics. There is an extensive menu of optional topics on further statistical methodology and applications. There are two examination papers of three hours each with 75 hours or equivalent work per paper, with 1 question per 8-10 hours work:

- Paper (i): Principles of statistical analysis
  This consists of six compulsory questions on core material.

- Paper (ii): Further statistical methodology
  This contains around 12 questions on core and optional material: students are expected to do five including two on the core topics

The distinctions between the MSc and the Diploma courses are:

The Diploma course ends after the examination in June. The MSc course involves in addition a project during the summer and submission of a substantial dissertation by the middle of September.

The choice between these courses rests partly on the student's preference for a 9- or 12-month course, but also on the decision of the Director of Graduate Studies.

*Practical Sessions*

There are compulsory practical sessions each week in Michaelmas and Hilary Terms which students attend. Two practical sessions in each term are assessed. Students write reports on their analysis of a statistical problem which are marked and feedback is communicated to students by the assessor. In week 5 of Trinity Term there is a week-long practical where a major assignment involving the analysis of 3-4 data sets is completed for assessment.

*Dissertation*

MSc students are required to submit a dissertation on some topic with an agreed supervisor and approved by the MSc Supervisory Committee.

The dissertation is expected to include evidence that the candidate is capable of applying statistical methods, operational research methods, or stochastic modelling to realistic problems. Most dissertations will therefore contain an account of the analysis of some body of real data, and this work constitutes the summer project.

**Learning**

## Paper I *Principles of statistical analysis*

**Statistical Methods (18 hrs)**

- Visualization: plots including histograms, box plots, scatterplots, scatterplot matrices, parallel coordinate plots, mosaicplots.
- Summary statistics and goodness-of-fit tests. One- and two-sample examples. t and F distributions. [All revision.] Robustness and robust summaries.
- Concepts of simulation: simple simulation experiments.
- Linear regression, including multiple linear regression. Associated inference problems. Regression diagnostics and resistant regression. Classical applications to ANOVA. Model selection.
- Logistic regression and Poisson regression. Proportional-odds logistic regression.
- Introduction to the design of experiments, observational studies and sampling methods.

**Statistical Theory (16 hrs)**

- Introduction. Likelihood and sufficiency; ancillarity and conditioning.
- Point estimation: bias and variance, information and efficiency. Minimum variance unbiased estimation. Maximum likelihood estimation.
- Hypothesis testing: pure tests, significance level. Simple hypotheses, Neyman-Pearson Lemma. Tests for composite hypotheses.
- Interval estimation: confidence regions and prediction regions.
- Asymptotic properties: maximum likelihood estimates; generalized likelihood ratio tests; likelihood confidence regions.
- Bayesian Statistics: Interpretations of probability; the Bayesian paradigm: models, exchangeability and sufficient statistics: prior distributions, conjugate and noninformative priors: posterior summaries.
- Decision theoretic framework: point estimation, loss function, admissibility, minimaxity: model comparisons

**Survival Analysis (6 hrs)**

- Survivor and hazard functions; censoring. Nonparametric analysis; life tables; product-limit estimator; Greenwood's formula; actuarial estimator. Parametric models, medical and industrial applications.
- Parametric analysis for a single sample.
- Regression models for data in continuous time; accelerated life; proportional hazards; model fitting and checking. Partial likelihood. Log-rank tests.

**Time Series (8 hrs)**

- Nature of time series. Autocorrelation. Linear models.
- Box-Jenkins methods.
- Frequency domain methods. Filters. State-space models. Non-linear models.

## Paper II *Further statistical methodology*

### Core Topics

**Multivariate Analysis (3 hrs)**

- Graphical methods. Brush and Spin, Projection pursuit.
- Principal components and factor analysis.
- Discrete methods, including correspondence analysis.

**Computer-intensive Statistics (6 hrs)**

- Density estimation: kernel density estimation: bandwidth selection.
- Non-parametric regression: smoothing/splines; Generalized Additive Models, other methods; software.
- Parametric Frequentist Monte Carlo Inference; Monte Carlo tests and confidence intervals; parametric bootstrap methods.
- The non-parametric bootstrap: sampling from the e.d.f.; estimating the mean and standard errors; bias reduction: the jack-knife.

**Further Statistical Methods (10 hrs)**

- Graphical models for categorical data. Special models for square contingency tables.
- Models for hierarchical ('multi-level') and longitudinal data.
- Latent variable models (including factor analysis)
- Methods for dealing with missing data.

**MCMC and applied Bayesian statistics (6 hrs)**

- Principles of Bayesian modelling.
- Non-conjugate models and inference via MCMC.
- Techniques for MCMC sampling: the Gibbs sampler and Metropolis-Hastings algorithm. Efficiency considerations when using MCMC methods. Output analysis and tests for convergence.
- Examples of MCMC for hierarchical linear and random effects models using WinBUGS.

## Optional Topics

**Mathematical/statistical genetics (8 hrs)**

- Background: DNA, genes, chromosomes; Mendelian segregation; population frequencies, Hardy-Weinberg law.
- Genetic counselling, estimating gene frequencies. Mapping genes: recombination, linkage, likelihoods in Mendelian pedigrees and for complex diseases.
- DNA and protein sequence alignment. Background, scoring functions, gap penalties, global alignment and local alignments – dynamic programming algorithms.

**Infectious Diseases (6 hrs)**

- Models for the spread of infectious diseases in large populations and small groups.
- Bovine spongiform encephalopathy (BSE) and new variant Creutzfeldt-Jakob disease (vCJD) will be examined as case studies.

**Combinatorial Optimization (12hrs)**

- Combinatorial optimisation is mainly concerned with methods for finding extreme values of functions defined on discrete sets. Typically the set may be represented as a graph or network, consisting of nodes joined by edges, and the solution is by means of an algorithm.
- Given a network which is to be loaded in some way there will be corresponding CO problems. The network might be of airports, or oil-pipe-lines, or classes, class-rooms and teachers, or telecommunications links, or simply represent the possible sequences in which a set of tasks may be carried out. The task then is to devise schedules, time-tables, or priority rules which in some sense optimise the system. Manufacturing industry presents many such problems, where operational researchers are often responsible for their solution. Computers have also been a strong stimulus to developments in this field, both by presenting design problems for hardware and software, and as the means whereby CO algorithms may be implemented.
- The main mathematical ideas in this course are some basic graph theory and duality and complementary slackness in linear programming. None of these are prerequisites.

**Actuarial Science (16hrs)**

- Term structure of interest rates, deterministic and stochastic interest rate models
- Effect of changes in interest rates: effective duration, convexity, immunisation
- No arbitrage models: valuation and hedging
- Life insurance: simple life insurance products via life-time distributions and lifetables, premiums and reserves

**Statistical Data Mining (12 hrs)**

- Fundamentals of pattern recognition, machine learning and data mining.
- Exploratory methods: principal components analysis, biplots, independent component analysis, multidimensional scaling.
- Cluster Analysis: K-means, hierarchical methods, vector quantisation, self-organising maps.
- Linear discriminant analysis, logistic discrimination, linear separation and perceptrons.
- Classification trees. Splitting criteria, existence of pruning sequences. V-fold cross-validation.
- Feed-forward neural networks. Universal approximation properties, back-propagation, training algorithms, assessment of fit.

**12. Support for Students and their learning**

- The Department has an excellent library of over 3000 items covering statistics and statistical journals.
- The Department has a network of PCs with extensive software available including S-Plus and R, two statistical packages.
- 
- Courses are supported with example classes and tutorial classes in specialist topics.
- Each student meets with a primary supervisor (usually a postdoc or research student) for an hour every other week during full term. Students are allocated supervisors in Michaelmas term. Groups of students (and their supervisors) are then assigned to an experienced member of academic staff as a secondary advisor. The Department's Course Coordinator, Chairman of Supervisory Committee, Director of Graduate studies and Academic Administrator are all available to provide further guidance and support
- A project supervisor is appointed for each student in Trinity term who will typically meet eight times with the student to supervise their dissertation, and read through and comment on thesis drafts.
- There is a Course Coordinator who makes the everyday arrangements for the course .
- Within College there will also be a Tutor for Graduates or a Senior Tutor and college advisors who will provide support, including welfare, pastoral and financial matters.

**13. Criteria for Admission**

The normal entry requirement is a first or upper second class degree (or equivalent) in an appropriate subject. For the MSc this usually means a degree with a large component of Mathematics or Statistics. There is a minimum requirement on a standard language test for applicants whose first language is not English. Admissions criteria can be found at http://www.stats.ox.ac.uk/~gradstud/admissions_criteria.html

Admissions information is available on the website www.stats.ox.ac.uk/grad/.
Posters are sent to relevant departments in the UK and overseas.

The application form requests:

- a statement of the reasons for wanting to take the course;
- applicants to arrange statements of support from three referees;
- information about financial support.

Prospective colleges check whether the applicant has adequate financial support.

Applicants are chosen on academic merit, their referees' reports, and reasons for wanting to enrol in the course.

**14. Methods for evaluating and improving the quality and standards of learning**

Methods for evaluating and improving quality and standards include:

- The external examiner system (including moderation of students' assessed work).
- Review of examiners reports, and student performance.
- MSc supervisory committee.
- Student feedback (including questionnaires) and Graduate Liaison Committee

***B. Skills and other attributes***

***I. Intellectual skills***

At the end of the course students will be expected to:

- understand a wide range of mathematical theory, statistical theory, and statistical techniques, relating to Applied Statistics;
- be able to apply appropriate Statistical techniques to data;
- be able to undertake a piece of directed research in Statistical analysis and data analysis;
- be able to critically evaluate reports and papers using Statistical analysis.

II. ***Practical skill***s

At the end of the course students will be expected to:

- be able to carry out Statistical computing techniques in data analysis;
- be able to write Statistical reports on data analysis.

***III. Transferable skills***

At the end of the course students will be expected to:

- be able to critically evaluate Statistical reports;
- be able to achieve computing skills, with emphasis on Statistical computing;
- be able to communicate research by seminar presentation and scientific report writing.

**15. Regulation of assessment**

Assessment regulations are provided to students at the beginning of the course in the Student Handbook.

For MSc candidates the overall assessment is based on:
(1) Paper (i) Principles of Statistical Analysis (a 3 hour paper)
(2) Paper (ii) Further Statistical Methodology (a 3 hour paper)`
(3) Assessed Practical Work
(4) Dissertation.
Each of (1)-(4) has equal weight, i.e. each contributes 25% to the overall MSc assessment.

The guidelines are 75 hours of lectures or equivalent work for each of (1) and (2), with 1 question per 8-10 hours work. In recent years (3) has been made up of 2 practical assignments in Michaelmas Term, 2 practical assignments in Hilary Term, and a Practical Assessment in week 1 of Trinity Term, with half of the weight of (3) being given to the Trinity Term Practical Assessment, the other half to the 4 practical assignments from Michaelmas/Hilary Terms. Candidates are told if there are any changes to these guidelines.

Candidates for the Diploma take (1)-(3) only, and the corresponding weights for (1), (2) and (3) are 37.5%, 25% and 37.5% respectively. The overall pass mark is 40%. At least 70% is required for a distinction.

For both MSc and Diploma, candidates can pass, pass with distinction, or fail. Distinction candidates will show excellence over a wide range of topics. Passing candidates will at least show satisfactory work over a reasonable range of topics. These descriptions are of overall performance: weaker performance in part of the overall assessment can be compensated for, if the overall performance merits it. Candidates who just fail the MSc can be allocated a pass on the Diploma if they show understanding and competence equivalent to passing the Diploma.

**16. Indicators of quality and standards**

- The external examiner system (including moderation of students' assessed work). Assessed performance in practical classes, report writing, examinations, minor thesis, and regular supervisory meetings.
- Student feedback (including questionnaires).
- QAA review.