THE HONG KONG STATISTICAL SOCIETY

SYLLABUSES

for

the Examinations of the Society

FOR EXAMINATIONS TO BE HELD IN 2017

and subsequently until further notice

Please note that the examinations will be withdrawn after the May 2017 session - http://www.hkss.org.hk/index.php/prof/exam/discon-exam

CONTENTS

Page

INTRODUCTION TO THE EXAMINATIONS

The professional examinations conducted by the Hong Kong Statistical Society provide a route into the statistical profession at introductory, intermediate and advanced levels. They will be suitable for many people who wish to enter into or advance within the profession. They may be particularly helpful for those who wish to convert to statistics from another discipline, and for those in countries overseas where little or no statistical training and few formal statistical qualifications are available.

The examinations consist of papers at three levels, namely *Ordinary Certificate*, *Higher Certificate* and *Graduate Diploma*. Taken in sequence, these levels provide a comprehensive syllabus of study in applied statistics.

The Ordinary Certificate is offered in a modular form consisting of two modules which may be taken singly or together. Each module is separately certificated. The Ordinary Certificate as a whole is awarded on successful completion of both modules; the "shelf life" of modules leading to this Certificate is unlimited.

The Higher Certificate is offered in a modular form consisting of eight modules which may be taken singly or in any combination. Each module is separately certificated. The Higher Certificate as a whole is awarded on successful completion of six modules (subject to some restrictions on choice of modules); the "shelf life" of modules leading to this Certificate is unlimited.

The Graduate Diploma is offered in a modular form consisting of five modules which may be taken singly or in any combination. Each module is separately certificated. The Graduate Diploma as a whole is awarded on successful completion of all modules; the "shelf life" of modules leading to the Diploma is unlimited.

The Graduate Diploma is widely recognised and respected, locally and internationally, by employers in the public and private sectors. It is also widely recognised by universities as being sufficient for admission to an MSc course in Statistics.

The examinations are held in May each year. Individuals entering for examinations must be Full Members/Student Members of the Society or Examination Associates. Full details of the procedures for entering examinations are contained in document "Guide to Examinations" which is available from the Society.

The Society provides statistical tables in each examination. Copies of these tables may be obtained from the Society so that candidates may familiarise themselves with the layout.

Candidates are expected to make appropriate use of calculators in the examinations. Regulations concerning calculators are contained in document "Guide to Examinations".

Copies of past papers and solutions are available from the Society, as are suggested reading lists for the examinations. Additional specimen papers and solutions for the Higher Certificate and Graduate Diploma are also available.

The address for all examinations correspondence is Hong Kong Statistical Society Examination Office, c/o HKU School of Professional and Continuing Education, Rm 313, 3/F, Admiralty Centre, 18 Harcourt Road, Hong Kong. Email may be sent to exam@hkss.org.hk.

All examinations material is available on the Society website at

http://www.hkss.org.hk/Exam/Exam.htm

ORDINARY CERTIFICATE IN STATISTICS

The Ordinary Certificate of the Hong Kong Statistical Society is offered in a modular form. This gives the opportunity for candidates to proceed at their own pace. The object of this first level certificate is to provide a sound grounding in the principles and practice of statistics with emphasis on practical data collection, presentation and interpretation. It is intended both as an end in itself in terms of being a first qualification in statistics, and as a basis for further work in statistics, as for example in the more advanced syllabuses of the Society. In addition, the individual modules are intended as valuable free-standing elements in their own right; they may for example be useful components of a formal or informal continuing professional development programme. The entry level is a good Hong Kong Diploma of Secondary Education (HKDSE) in mathematics (compulsory part) or a good Hong Kong Certificate of Education Examination (HKCEE) in mathematics or an equivalent qualification, or relevant work experience.

The two modules cover the essential ideas of statistics in practice. Anyone familiar with them should be able to carry out supervised statistical work of a routine kind, or be able to apply statistical methods, at an elementary level, within work of a more general kind. Module 1 includes the key topics of data collection in the field: what data should be collected, who from, and how will the data be captured? It also includes work on data processing, including a brief introduction to statistical computing. Module 2 encompasses the basic techniques of descriptive statistics. Two ideas are of paramount importance: what is the most appropriate way to analyse the data, and how can the message within the data most clearly be communicated?

Candidates may enter for either or both modules in any examination session. Candidates may resit any module any number of times; the highest mark achieved will stand. The pass mark for each individual module is 50%. Module marks are "banked" without time limitation.

Each module is examined by a 3-hour written paper containing from 7 to 9 (inclusive) questions of different lengths. There are no restrictions on the number that may be answered. The marks available for each question are printed on the examination papers. Candidates will be advised of their marks for each module taken. Candidates will receive a separate certificate for each module in which they are successful.

Candidates who are successful in both modules (not necessarily in the same session) will be awarded the Ordinary Certificate in Statistics. The Ordinary Certificate will only be awarded on the first occasion when a candidate becomes eligible for it. If a candidate, having been awarded the Ordinary Certificate, takes either or both modules again, advice of the marks earned will be issued and, if appropriate, a certificate to indicate success in the individual module(s); but the Ordinary Certificate itself will not be re-awarded.

MODULE SYLLABUSES

Syllabuses for the two modules are presented over the ensuing pages.

MODULE 1: Collection and Compilation of Data

The origin, use and interpretation of published or administrative data.	*Overview of official statistics in a country of the candidate's choice. What statistical series are produced, how are the data collected and to what uses are the data put?*
Elementary ideas of sampling methods. Definitions of population and sampling frame. Methods of selecting samples (including practical problems) and implications of sample size: simple random sampling, systematic sampling, cluster sampling, quota sampling, stratified random sampling and multi-stage sampling.	*No formulae are required.* *Advantages and disadvantages of each method, including considerations of cost, precision and accuracy. Use of random number tables.*
Pilot surveys, censuses, sample surveys, personal interviews, self-completion questionnaires, postal and telephone enquiries. Serial surveys - longitudinal or cross-sectional. Problems arising in the collection of data, dropouts, late returns, 'freak' values and their treatment, 'cleaning' data.	*No formulae are required.*
Non-sampling errors. Identification and inter-pretation of bias error (e.g. from non-response, errors in defining the population, enumerator distortion, etc).	*No formulae are required. Candidates may be asked to provide examples of how different types of bias error may arise.*
Design of simple questionnaires and forms for collection of data. Formulation, classification and coding of questions, including verification. Making questionnaires suitable for data processing and analysis; use of missing value codes.	*Candidates should be able to produce their own simple questionnaire and data form.*
Distinction between observational and experimental studies.	
Use of computers for data storage and retrieval.	*Access to a computer when preparing for the examinations is not necessary.*

MODULE 2: Analysis and Presentation of Data

Use of rough checks for order of magnitude and leading digits in results.	*Candidates are encouraged to check that their answer 'looks about right'.*
Approximation, limits of accuracy, rounding and accuracy of recording. Percentages, ratios, rates and linear interpolation. Distinction between discrete and continuous data.	*Error in the result of a simple expression when the consistent parts are all rounded.*
Construction and uses of frequency tables for one or more variables and contingency tables. Tables for presenting collections of results together with summary tables of frequencies, relevant averages, standard deviations, etc.	*Credit will be given for neatness and clarity of presentation.*
Graphs and diagrams, their use in analysis and presentation. Construction, uses and limitations of scatter diagrams, time charts, stem and leaf diagrams, histograms, bar charts, pie charts, frequency and cumulative frequency curves and boxplots (box and whisker plots).	*Credit will be given for neatness and clarity of presentation.*
Sample measures of location and dispersion. Arithmetic mean, median, mode, percentiles, range, inter-quartile range, variance, standard deviation, coefficient of variation; their uses and limitations as measures; their calculation from frequency tables and raw data; graphical methods of estimation. Distinction between inter- and intra-subject variation.	*Formulae should be known, except that the definition of 'hinges' is not required.* *Candidates should be able to estimate percentiles from a cumulative frequency curve (ogive).*
Identification of outliers in a data set, and appreciation of steps that may be taken to deal with them.	*Outliers should be considered in the context of the data-set as a whole. Candidates need only use informal criteria for classifying observations as outliers (as in the construction of boxplots), but should be able to discuss how they might be dealt with.*
Probability as a measure of uncertainty. Link between probability and relative frequency. Allocation of probabilities in 'equally likely' cases. Mutually exclusive events. Independent events. Addition and multiplication of probabilities with simple applications. Use of Venn diagrams and tree diagrams.	*The long-run concept of probability, e.g. from tossing a coin repeatedly, should be understood.* *Permutations and combinations are not required.*
Calculation of least squares regression line and its interpretation. Correlation as a measure of linear association between two variables. Product-moment correlation coefficient. Spearman's rank correlation coefficient.	*Derivation of the least squares estimates (by calculus) is not required.* *Candidates should know when it is appropriate to use each method.* *Consideration of tied ranks will not be expected.*
Simple moving averages for detecting trends and for smoothing time series. Seasonal data.	*Candidates should know when it is appropriate to use additive or multiplicative models for seasonality, and how each is calculated.*
Knowledge of weighted forms of moving average.	*No calculation is required.*

continued on next page

Simple and weighted averages of price relatives. Construction of aggregate (Paasche, Laspeyres and Fisher) averages. Simple chain-based indices. Limitation and use of index numbers e.g. in assessment of productivity and prices.	*Knowledge of how and why index numbers are used in real life is assumed.*
Interpretation. Translation of written statements into tabular forms; simple fallacies, typical misleading distortion in popular published graphs. Answers to questions about tables and charts.	*Candidates may be asked to explain what the data presentation tells the reader.*
Writing of clear and concise reports on numerical data in different contexts.	*Candidates may be tested on their spelling and grammar as well as their logic.*

HIGHER CERTIFICATE IN STATISTICS

The Higher Certificate of the Hong Kong Statistical Society is offered in a modular form. This gives the opportunity for candidates to proceed at their own pace. The Certificate as a whole is intended both as an end in itself in terms of a qualification in statistics at a more advanced level than that of the Ordinary Certificate, and as a basis for further work in statistics. In addition, the individual modules are intended as valuable free-standing elements in their own right; they may for example be useful components of a formal or informal continuing professional development programme.

Candidates may enter for any number of modules, either singly or in combination, in any examination session. Candidates may resit any module any number of times; the highest mark achieved will stand. The pass mark for each individual module is 50%. Module marks are "banked" without time limitation.

Each module is examined by a 1½ -hour written paper containing four questions of which candidates are asked to answer three. Candidates will be advised of their marks for each module taken. Candidates will receive a separate certificate for each module in which they are successful.

Candidates who are successful in six modules **including modules 1 to 4** (not necessarily all in the same session) will be awarded the Higher Certificate in Statistics. The Higher Certificate will only be awarded on the first occasion when a candidate becomes eligible for it. If a candidate, having been awarded the Higher Certificate, takes any module(s) again or takes any further module(s) not already taken, advice of the marks earned will be issued and, if appropriate, a certificate to indicate success in the individual module(s); but the Higher Certificate itself will not be re-awarded.

Candidates who wish to proceed to the Graduate Diploma in Statistics are advised that it is desirable that they have been successful in **modules 1 to 6** (again, not necessarily all in the same session).

Candidates are advised that entry to any of the modules assumes knowledge of the material in the Ordinary Certificate in Statistics, or equivalent. Candidates should therefore ensure that they are familiar with this material before sitting for any modules. This is particularly important for candidates who wish to attempt to achieve the Higher Certificate in Statistics by taking six modules (including modules 1 to 4) in the same session.

MODULE SYLLABUSES

Syllabuses for the eight modules are presented over the ensuing pages.

The modules are

Module 1	Data collection and interpretation
Module 2	Probability models
Module 3	Basic statistical methods
Module 4	Linear models
Module 5	Further probability and inference
Module 6	Further applications of statistics
Module 7	Time series and index numbers
Module 8	Survey sampling and estimation

Information on advised academic progression through the modules, on the required mathematics background and on interpretation of computer output follows after the last module.

MODULE 1: Data collection and interpretation

Summarising and interpreting data

Frequency distributions. Numerical and graphical forms of presentation and statistical interpretation. Scatter diagrams, time charts, stem and leaf diagrams, histograms, bar charts, pie charts, frequency and cumulative frequency curves, boxplots (box and whisker plots), dotplots.	*Credit will be given for graphs that are well presented, including neat appearance and inclusion of appropriate title and labels.*
Summary statistics for measures of location, variability and skewness.	*Sample mean, median, mode, quartiles, semi-inter-quartile range, standard deviation, variance, range.*

Surveys

Target and study populations. Sampling frames. Problems arising in the collection of data. Censuses, sample surveys and routine collection of data at intervals of time.	*Principles and practice, with examples from candidates' knowledge and experience.*
Design of questionnaires and forms for collecting data. Personal and telephone interviews, postal enquiries, pilot enquiries. Problems of non-response, bias among interviewers, question bias, non-sampling errors.	*Advantages and disadvantages of these methods.*
Simple random sampling. Uses and limitations. Estimators for means, totals and proportions and the variances of these estimators.	*Candidates will be expected to know and be able to use, but not to derive, formulae for estimators and variances, along with an informal understanding of standard errors. Use of the finite population correction is not required.*
Use of other practical methods of sampling: systematic sampling, cluster sampling, quota sampling, stratified random sampling and multi-stage sampling.	*No formulae are required.*

Exploratory analysis

Candidates should be prepared to examine a set of data, to choose and carry out suitable methods of analysis, to answer questions from non-statistical users and to present the analysis and conclusions in the form of a short report. The techniques required may be of the simplest kind, e.g. plotting, grouping, transforming or calculating from the data. Candidates will be expected to use box-plots and other similar graphical displays.	*Particular questions may be specified, and also particular types of non-statistical user (see below). Otherwise a brief general report on the main points in the data analysis will be appropriate.*

Report writing

Candidates should be prepared to produce a well-ordered, well-reasoned argument in a style suitable for a designated readership. (This readership could be, for example, non-statistical colleagues, managers, or users of official reports.)	*The use of technical words alone will not be enough; explanations must be helpful to the non-specialist.*
Candidates will be expected to make use of graphical methods to summarise data and identify unusual features.	*Graphs and diagrams must be clearly labelled so that the reader understands them properly.*
Interpretation of published data	*Candidates should be able to provide critical comment on tables and graphs.*

MODULE 2: Probability models

Probability

Definitions of probability: equally likely outcomes; relative frequency; degrees of belief. Addition and multiplication of probabilities, conditional probability, statistical independence. Bayes' theorem.

Distributions

Random variables. Discrete and continuous probability distributions. Probability mass function, probability density function, cumulative distribution function. Simple theory of elementary probability distributions, including discrete uniform, Bernoulli, binomial, Poisson, geometric, negative binomial, hypergeometric, Normal, exponential, gamma and continuous uniform.

Derivation of cumulative distribution function from probability density function and vice versa.

Candidates should be able to show that the probability mass function (probability density function) of the named distributions sums (integrates) to unity. Some knowledge of the derivation and application of these distributions is expected. The Poisson approximation to the binomial is included.

Candidates should know how these distributions arise in practice and be able to recognise them from a brief description of a situation.

Properties of distributions

Expectation and variance; their general properties and values for standard distributions.

Derivation of the expected value and variance of random variables with the distributions listed above. Questions may be set for other simple distributions.

Distributions, means and variances of sums of independent and identically distributed random variables and simple functions, such as $aX + b$. Linear combinations of independent Normally distributed variables.

Results for distributions of sums of Poisson, Normal and exponential random variables should be known.

Distribution of $\sum a_i X_i$ *when* X_i *has* $N\left(\mu_i, \sigma_i^2\right)$ *distribution.*

Statement and use of central limit theorem for independent, identically distributed random variables with finite variance.

Proof of CLT not required.

Use of Normal approximations, including those for binomial and Poisson distributions.

Use of appropriate continuity corrections will be expected.

MODULE 3: Basic statistical methods

Inference

Sample and population. Concept of a sampling distribution. Standard error.	*Sampling distribution of the mean.*
Point and interval estimates. Construction and interpretation of confidence limits.	
Hypothesis tests, test statistic, one- and two-sided tests. Significance level. Type I and II errors. Power as 1–*P*(type II error).	*Knowledge of p-values and their interpretation.*
Use of Normal, *t*, χ^2 and *F* distributions in testing and interval estimation. Paired and unpaired two-sample tests.	*Tests and confidence intervals involving means, variances and proportions. Use of tables to obtain percentage points.*
Power curves.	*Restricted to cases of testing for the mean of a Normal distribution with known variance.*
The χ^2 goodness-of-fit test of standard distributions to observed data.	*Including pooling of classes. Uniform (discrete and continuous), binomial, Poisson and Normal distributions; distributions in specified proportions.*
Analysis of two-way contingency tables; χ^2 test for association.	*Use of Yates' correction in 2×2 tables is expected.*
McNemar's test.	*Use of a continuity correction is not expected. Consideration of tied ranks will not be expected.*

Non-parametric methods

Use of non-parametric and distribution-free significance tests for paired and unpaired data: sign test, Wilcoxon rank sum test (Mann-Whitney *U* test), Wilcoxon signed-rank test.	*Candidates will be expected to be able to use tables of percentage points but do not need to know how the tables are obtained.* *Consideration of tied ranks will not be expected.*

MODULE 4: Linear models

Correlation

Product-moment correlation (Pearson). Rank correlation – Spearman's coefficient. Calculation and interpretation.	*Association versus causality. Simple tests using tables, with informal understanding of when each measure of correlation is appropriate. Consideration of tied ranks will not be expected.*

Design of experiments

Reasons for experimentation, causality.	
Principles of replication and randomisation, completely randomised design.	

Regression

Simple linear regression. Least squares estimation.	*Models involving qualitative regressor variables are not included in this module.*
Multiple linear regression – concepts, interpretation of computer output, inference for regression coefficients using estimates and estimated standard errors from computer output.	*Knowledge of F test for regression and partial t test for regression coefficients. Methods for selecting variables are not included in this module.*
Analysis of variance for regression models.	
Calculation and interpretation of the multiple correlation coefficient (coefficient of determination).	R^2 *as a measure of the proportion of variation explained.* *Calculation of prediction intervals will not be required.*
Simple cases of transforming to linearity.	*For example, problems (such as growth curves) in which taking logarithms or reciprocals leads to a straight-line relationship.*

Analysis of variance

One-way analysis of variance.	*Relationship to completely randomised design.*
Inference for means and for differences in means.	*Multiple comparison procedures will not be required.*

MODULE 5: Further probability and inference

Bivariate distributions

Simple bivariate discrete distributions. Joint, conditional and marginal distributions: probability mass function, expectation and variance. Covariance and correlation.	*Distributions with probabilities presented either by formula or in a two-way table.*
Simple bivariate continuous distributions. Joint, conditional and marginal distributions: probability density function, expectation and variance. Covariance and correlation.	*Simple cases only – rectangular and triangular spaces, with one or more sides parallel to a coordinate axis.*
The bivariate Normal distribution.	*Familiarity with the bivariate Normal distribution as a model (knowledge of joint pdf is not required). Proofs of results are not required.*

Generating functions

Probability and moment generating functions. Use to find expectations and variances. Use to establish the distribution of sums of random variables.	*No limiting results.*

Inference

The likelihood function.	*Plots of likelihood against the parameter.*
Estimation of a single parameter of a distribution using the method of moments and the method of maximum likelihood.	*Maximisation of likelihood using graphical methods and calculus. Questions may be set involving standard distributions listed in Module 2 and other simple cases.*
Properties of point estimators. Range, unbiasedness, consistency. Efficiency and relative efficiency.	*Introductory treatment of these concepts in simple cases.*
Calculation of approximate variance of a maximum likelihood estimator using second derivative of log likelihood.	*Use in constructing a confidence interval. Regularity conditions may be assumed.*

MODULE 6: Further applications of statistics

Design and analysis of experiments

Principles of design including randomisation, blinding, pairing and blocking.	*Reasons for using these.*
Randomised block designs. Latin squares.	*Analysis of variance, inference for means and for differences in means.*
Factorial treatment structure with two factors. Advantages of factorial experimentation. Diagrammatic explanation of interaction. Two-way analysis of variance.	*Diagrams of means of treatment combinations and their use for explaining interactions of two factors. Analysis of variance, inference for means and for differences in means.*
Residuals and their use in checking assumptions.	

Multiple regression

Least squares estimation for multiple regression.	*Extension of material in Module 4. Derivation of normal equations in simple cases. Matrix notation will not be required. Solution of simultaneous equations in simple cases only.*
Regression through the origin.	*Including multiple regression with zero intercept.*
Use of backwards elimination in multiple regression.	*Use of F tests.*
Polynomial regression.	*Simple cases only.*
Use of indicator variables to model factors or qualitative variables.	*Simple cases only.*
Residuals and their use in checking assumptions.	

Quality control and acceptance sampling

Charts for mean and range for Normal data. Charts for proportions.	*Construction and use of Shewhart charts, including use of warning and action lines.*
Cusum charts.	*Construction and use.*
Attribute sampling. Single and double sampling schemes.	

MODULE 7: Time Series and Index Numbers

Time series

Decomposition of time series into trend, cycles, seasonal variation and residual (irregular) variation.	*Additive, multiplicative and pseudo-additive decompositions.*
Estimation of trend using regression or moving averages.	*Equally and unequally weighted moving averages.*
Examination of seasonal terms in time series decompositions. Seasonal adjustment using regression or moving averages.	*Prior adjustment. Detection and control of special events in seasonal adjustment, for example religious holidays, changes in taxation levels, … .*
Elementary forecasting methods: exponential smoothing and Holt-Winters.	
Introduction to ARIMA models.	*Definition, interpretation, forecasting. Details of techniques for fitting ARIMA models are not required.*
Examination and interpretation of residuals from fitted models.	*Including the correlogram and the portmanteau lack-of-fit test.*
Interpretation of computer output.	*Analysis and commentary on tables and graphs produced by seasonal adjustment programs. Confirmation of some results through hand calculation.*

Index numbers

Introduction to index numbers

Index numbers and their uses. Simple price relatives, Laspeyres and Paasche. Relationship between Laspeyres and Paasche; and the relative merits of each. Further index numbers – Törnqvist, Walsh, Fisher and Geometric Laspeyres. Index aggregation.	*Consideration of price and volume indices, with examples from candidate's knowledge and experience.* *Calculation of index numbers by aggregating others.*

Deflation

Why deflation is used and how it works; what makes a good deflator; how deflation is carried out.	*With reference to specific examples from candidate's knowledge and experience.*

Re-basing

Why, when and how re-basing is done.	*With reference to specific examples from candidate's knowledge and experience. Benefits and pitfalls.*

Chain linking

Chain linking of simple price relatives, and chain linking using Laspeyres.	*With reference to specific examples from candidate's knowledge and experience. Benefits and pitfalls.*
Use of index numbers	*With reference to specific examples from candidate's knowledge and experience, for example within National Accounts.*

MODULE 8: Survey Sampling and Estimation

Populations and frames

Target and study populations. Types of frames available, uses and sources of error.

Sampling methods

Non-probability methods, haphazard sampling, quota sampling.

Advantages and drawbacks of these methods, sources of bias, unknown precision.

Simple random sampling, stratified random sampling (with equal, proportional and optimal allocation), cluster and multi-stage sampling, systematic sampling.

Uses, benefits and limitations of each method. Discussion of practical and theoretical utility of methods in the context of specific examples.

Simple and stratified random sampling

Uses, limitations, applications to different data types, practical examples.

Estimates of totals, means and proportions, construction of confidence intervals.

Neyman and optimal allocation, use to reduce variance.

Discussion and comparison of these methods.

Calibration techniques for estimation

Ratio and regression methods. Use of supplementary information.

Discussion of how and why other survey information may be useful in estimation. Formulae will not be required.

Practical problems in planning and conducting surveys

Choice of sampling method and estimators to be used in a survey, trade-off between bias and variance. Discussion of sampling problems in an actual survey, recommendations for improvement using practical examples.

Cross-sectional, longitudinal and panel surveys, drug trials, etc.

ADVISED ACADEMIC PROGRESSION THROUGH THE MODULES

The Society does not operate any system of formal prerequisites for the modules. Any candidate may enter for any module at any time. However, the Society strongly encourages candidates to ensure that they are properly prepared for the modules they intend to take. The advice stated above is repeated here, that entry to any of the modules assumes knowledge of the material in the Ordinary Certificate in Statistics, or equivalent. Candidates should therefore ensure that they are familiar with this material before sitting for any modules. This is particularly important for candidates who wish to attempt to achieve the Higher Certificate in Statistics by taking six modules (including modules 1 to 4) in the same session. Such familiarity with the material need not have been acquired by formally passing the Ordinary Certificate. It could have been achieved in other ways, including by experiential learning through a candidate's work.

Candidates are ***very strongly*** advised to study the published specimen papers, past papers and solutions. These give a good indication of the sorts of questions that will be asked, their general level, and the breadth and depth of answers that will be expected.

Although there are no formal prerequisites, it is an inherent feature of the subject that some modules require knowledge of some topics in other modules. Candidates may of course have already acquired such knowledge through other routes, or may be willing to acquire it as part of their studies for the module(s) at which they are aiming. As guidance, the Society offers the following advice, but this should not be seen as exhaustive:

Module	Some dependence on topics in Module or Modules
1	–
2	–
3	2
4	2, 3
5	2, 3, 4
6	1, 2, 3, 4
7	1, 2, 4
8	1, 2, 3, 4

The form and extent of any dependence varies considerably from case to case. For some modules, detailed knowledge of some of the topics contained in other module(s) might be needed. In other cases, candidates will just need to be familiar with or have some knowledge of the concepts and techniques contained in the other module(s).

MATHEMATICS BACKGROUND

The mathematics required to support the modules in the Higher Certificate in Statistics is naturally at a higher level than that required for the Society's Ordinary Certificate. A summary of the required mathematics is set out here, but prospective candidates should study the published specimen and past papers and solutions to get a detailed understanding of what is required in each module.

The examination papers will not concentrate on mathematics for its own sake. Rather, it is the ability to *apply* mathematics within the statistical contexts defined by the module syllabuses that will be examined.

Candidates should be aware that the general level of the Higher Certificate is, broadly, that of the first year of a university undergraduate degree in Statistics, also including some topics that might typically be found in the second year. The level of mathematics required is commensurate with this.

The detailed mathematical requirements naturally vary from module to module. As a general guide, a candidate who has knowledge of the topics contained in the extended part (i.e. module 1 on Calculus and Statistics or module 2 on Algebra and Calculus) of Mathematics of the Hong Kong Diploma of Secondary Education (HKDSE), or the A-level Mathematics in the schools examination system in Hong Kong or the equivalent in other systems, will *certainly* be able to start work on *any* of the modules. This does *not* imply that the extended part of HKDSE or A-level Mathematics (or equivalent) must necessarily have been formally taken and passed (or be studied concurrently). It is intended as a guide to the breadth and depth of mathematical topics with which candidates should be familiar. Some of the modules do not in fact require as much mathematics as this. On the other hand, completion of the work in other modules will require knowledge of additional mathematical topics beyond the extended part of HKDSE or A-level Mathematics (or equivalent).

The table on the next page sets out the various mathematical topics in broad headings and indicates which are required for which modules.

INTERPRETATION OF COMPUTER OUTPUT

In any of the modules, candidates will be expected to be able to interpret computer output from statistical packages. Detailed knowledge of specific packages is **not** required.

Table showing mathematics requirements for each module (see notes on previous page)

Topic	Module							
	1	2	3	4	5	6	7	8
Algebra								
Use of Σ notation	✓	✓	✓	✓	✓	✓	✓	✓
Permutations and combinations		✓	✓	✓	✓	✓		✓
Solutions of linear and quadratic equations	✓	✓	✓	✓	✓	✓	✓	✓
Manipulation and solution of simple inequalities	✓		✓	✓	✓	✓		✓
Arithmetic and geometric series	✓	✓		✓	✓	✓	✓	✓
Summation of series		✓		✓	✓	✓	✓	✓
Limits of sequences and functions		✓			✓	✓		
Positive numbers raised to any real power. Exponential and logarithmic functions, including their expansions in series		✓	✓	✓	✓		✓	
Use of the following results: $\lim(1+n^{-1})^n = e$ and $\lim(1+(x/n))^n = e^x$		✓	✓	✓	✓		✓	
Use of binomial theorem with any integer index		✓			✓	✓		✓
Solution of simple sets of linear equations having unique solutions		✓		✓		✓	✓	✓
Differential calculus								
Derivatives of polynomial, logarithmic and exponential functions, and of sums, products, quotients or functions of these functions		✓		✓	✓	✓		✓
Maxima and minima. Simple examples of asymptotes		✓		✓	✓	✓		✓
Graphical representation of functions and simple examples of curve sketching		✓		✓	✓	✓		
Simple examples of partial differentiation				✓	✓	✓		
Integral calculus								
Indefinite and definite integrals, including those with infinite limits		✓			✓			
Integrals of algebraic, exponential and logarithmic functions		✓			✓			
Simple examples of integration by substitution and by parts, including reduction formulae		✓			✓			
Double integration – elementary aspects only					✓			

GRADUATE DIPLOMA IN STATISTICS

The Graduate Diploma of the Hong Kong Statistical Society is offered in a modular form. This gives the opportunity for candidates to proceed at their own pace. The Graduate Diploma as a whole is a qualification in applied statistics at a level equivalent to that of a good Honours Degree in Statistics. In addition, the individual modules are intended as valuable free-standing elements in their own right; they may for example be useful components of a formal or informal continuing professional development programme.

Candidates may enter for any number of modules, either singly or in combination, in any examination session. Candidates may resit any module any number of times; the highest mark achieved will stand. The pass mark for each individual module is 50%. Module marks are "banked" without time limitation.

Each module is examined by a 3-hour written paper containing eight questions of which candidates are asked to answer five. Candidates will be advised of their marks for each module taken. Candidates will receive a separate certificate for each module in which they are successful. Candidates who are successful in all five modules (not necessarily all in the same session) will be awarded the Graduate Diploma in Statistics. The Graduate Diploma will only be awarded on the first occasion when a candidate becomes eligible for it. If a candidate, having been awarded the Graduate Diploma, takes any module(s) again, advice of the marks earned will be issued and, if appropriate, a certificate to indicate success in the individual module(s); but the Graduate Diploma itself will not be re-awarded.

Candidates are advised that entry to any of the modules assumes knowledge of the material in modules 1 to 6 (inclusive) of the Higher Certificate in Statistics, or equivalent (note: the former "traditional", non-modular, form of the Higher Certificate in Statistics, offered for the last time in 2008, is equivalent). Questions may be set using this material whether or not a topic is specifically mentioned in the individual Graduate Diploma module syllabuses below.

Candidates must not regard the Graduate Diploma modules as being fully academically separate from each other, because there are many interactions between branches of the subject at this level. In all cases, study of one module may require some study of work in other modules. This is especially true in respect of modules 5 and 4, but it is reiterated that it may apply to *all* modules. For this reason, and because candidates are likely to have particular interests and particular strengths in particular areas, the Society does not suggest an advised academic progression through the modules in the manner that it does for the Higher Certificate. *To some extent* modules 1 and 2 can be considered as a pair and modules 4 and 5 as another pair, with these pairs being independent of each other and with module 3 being more closely related to the pair (1, 2) than it is to the pair (4, 5). But this *must not be viewed as a strict separation*; there are potential connections between *all* of the modules.

A summary of the mathematics required for the Graduate Diploma is set out below, but

prospective candidates should study the published specimen papers, past papers and solutions to get a detailed understanding of what is required in each module. The mathematics background is naturally at a higher level than that required for the Society's Higher Certificate. It is commensurate with what might typically be required for the final year of an undergraduate degree in Statistics in a university. The Society does not separate the required mathematics background into distinct sections for each module. Candidates must assume that they might need any of the required mathematics in any of the modules.

The examination papers will not concentrate on mathematics for its own sake. Rather, it is the ability to *apply* mathematics within the statistical contexts defined by the module syllabuses that will be examined.

In any of the modules, candidates will be expected to be able to interpret computer output from statistical packages. Detailed knowledge of specific packages is **not** required.

Candidates are ***very strongly*** advised to study the published specimen papers, past papers and solutions. These give a good indication of the sorts of questions that will be asked, their general level, and the breadth and depth of answers that will be expected.

MODULE SYLLABUSES

Syllabuses for the five modules are presented over the ensuing pages, immediately following the information on the required mathematics background.

The modules are

Module 1	Probability distributions
Module 2	Statistical inference
Module 3	Stochastic processes and time series
Module 4	Modelling experimental data
Module 5	Topics in applied statistics

Summary of mathematics background

Knowledge of the mathematical topics listed below is required. These include those topics previously specified for the Higher Certificate. The ability to apply this mathematics will be examined within the statistical contexts defined by the module syllabuses.

Algebra

Permutations and combinations. Partial fractions, including quadratic factors. Solution of linear and quadratic equations. Manipulations and solution of simple inequalities. Trigonometric functions and their inverses. Summation of series with Σ notation. Limits of sequences and functions. Geometric series. Exponential and logarithmic functions, including their expansions in series and the results $\lim_{n\to\infty}\left(1+n^{-1}\right)^n = e$ and $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$. Taylor series expansion. Double summation ($\Sigma_i\Sigma_j$).

Positive numbers raised to any real power. Use of the binomial theorem with any real index.

Differential calculus

Differentiation. Derivatives of polynomial, trigonometric, inverse trigonometric, logarithmic and exponential functions, and of sums, products, quotients or functions of these functions. Maxima and minima; asymptotes; points of inflexion (defined as those points at which a curve crosses its tangent). Graphical representation of functions and simple examples of curve sketching. Partial differentiation.

Integral calculus

Integration. Indefinite and definite integrals, including those with infinite limits. Integrals of algebraic, exponential, logarithmic and trigonometric functions. Simple examples of integration by substitution and by parts, including reduction formulae.

Double integration over rectangular and triangular regions having one or more sides parallel to a co-ordinate axis. Functions of a single variable and of two variables. Interchange of order of integration.

Jacobians of transformations.

Matrix algebra

Vectors. Square matrices: symmetry, singularity and non-singularity, determinants, inverses, relation with sets of linear equations. Solution of simple sets of linear equations.

Rectangular matrices: addition, subtraction and multiplication. Rank of matrices via linear dependence and independence of vectors. Quadratic forms, including their expression in matrix notation.

Numerical methods

Iterative solution of equations, including Newton-Raphson method.

MODULE 1: Probability distributions

This module concentrates on probability and distribution theory, extending the coverage of these topics in the Higher Certificate. Simulation is also included.

Probability

Sampling with and without replacement. Elementary problems involving urn models.

Joint probability, marginal and conditional probability, independence.

Law of total probability. Bayes' Theorem.

Distribution theory

Random variables. Discrete and continuous random variables. The probability mass function and probability density function. Cumulative distribution function.

Expectation as a linear operator. Expectation of functions of a random variable. Mean and variance.

Approximate mean and variance of a function of a random variable. Variance-stabilising transformations.

Derivation of E[f(X)] by Taylor series approximation. Use in finding appropriate transformations.

Standard distributions and their use in modelling, including Bernoulli, binomial, Poisson, geometric, negative binomial, hypergeometric, discrete uniform, Normal, exponential, gamma, contin-uous uniform, beta, Weibull, Cauchy, lognormal.

Ability to recognise the appropriate distribution from a model description.

Relationships between distributions (e.g. exponential and Weibull).

Joint, marginal and conditional distributions. Independence. Covariance and correlation.

The multivariate Normal distribution. Regression; multiple correlation, partial correlation.

$E(Y \mid X=x)$ and $E(X \mid Y=y)$.

Knowledge and use of $E(Y) = E(E(Y \mid X))$. Use of $\mathrm{Var}(Y) = E(\mathrm{Var}(Y \mid X)) + \mathrm{Var}(E(Y \mid X))$.

Proofs not required. Recall of the result for Var(Y) will not be expected.

Probability generating function. Moment generating function. Applications of generating functions. Distribution of sums of random variables, and of sample mean.

Sums of fixed and random numbers of random variables.

Central limit theorem.

Proof for independent and identically distributed random variables only.

Distributions of functions of several random variables. Transformations, including the probability integral transform. Joint distribution of mean and variance from a Normal random sample.

Univariate and bivariate transformations.

The t, χ^2 and F distributions, and their use as sampling distributions.

Including definitions in terms of N(0,1) random variables.

Joint distribution of order statistics. Distribution of sample range.

Simulation

Generation of uniform pseudo-random numbers; testing for uniformity.

Methods of generating random numbers from common distributions, including inversion, rejection and table look-up techniques. Monte Carlo methods. Use of variance reduction techniques. Applications of simulation.

Applications of methods to standard distributions including binomial, Poisson, exponential, Normal, logistic.

MODULE 2: Statistical inference

This module extends the coverage of the material on statistical inference in the Higher Certificate. It also introduces Bayesian inference, decision theory and non-parametric inference.

Estimation

Unbiasedness, mean square error, consistency, relative efficiency, sufficiency, minimum variance. Fisher information for a function of a parameter, Cramér-Rao lower bound, efficiency. Fitting standard distributions to discrete and continuous data. Method of moments. Maximum likelihood estimation: finding estimators analyt-ically and numerically, invariance.

Hypothesis testing

Simple and composite hypotheses, types of error, power, operating characteristic curves, *p*-value. Neyman-Pearson method. Generalised likelihood ratio test.

Use of asymptotic results to construct tests. Central limit theorem, asymptotic distributions of maximum likelihood estimator and generalised likelihood ratio test statistic.

Proof of the asymptotic distributions of the maximum likelihood estimator and the generalised likelihood ratio test statistic are not required.

Confidence intervals and sets

Random intervals and sets. Use of pivotal quantities. Relationship between tests and confidence intervals. Use of asymptotic results.

Data resampling

Reduction of bias. Estimation of precision. Confidence interval estimation.

Jack-knifing and bootstrapping.

Non-parametric inference

Permutation and randomisation tests. Use of ranks and randomisation; robustness.

Sign, Wilcoxon rank sum (Mann-Whitney U), Wilcoxon signed-rank, Kolmogorov-Smirnov (one and two samples), goodness-of-fit and rank correlation tests.

Bayesian inference

Prior and posterior distributions. Choice of prior: bets, conjugate families of distributions, vague and improper priors. Predictive distributions. Bayesian estimates and intervals for parameters and predictions. Bayes factors and implications for hypothesis tests. Use of Monte Carlo simulation of the posterior distribution to draw inferences.

Knowledge of computationally intensive methods for simulating posterior distributions (e.g. Markov Chain Monte Carlo) is not required.

Decision theory

Loss, risk, admissible and inadmissible decisions, randomised decisions. Minimax decisions and Bayes' solutions, including simple results.

Comparative inference

Different criteria for choosing good estimators, tests and confidence intervals. Different approaches to inference, including classical, Bayesian and non-parametric.

MODULE 3: Stochastic processes and time series

This module provides an extended coverage of stochastic processes, including Markov chains and various forms of Poisson processes, and of time series, including ARIMA modelling.

The syllabus concentrates on underlying theory, but applications in various substantive areas are also important and will be represented in examination questions.

Stochastic processes

General stochastic process models.

Random walks.
Reflecting and absorbing barriers.
Mean recurrence time, mean time to absorption.

Difference equations.	*Use of generating functions.*
Branching processes.	*Recurrence relations for size of n th generation; probability of extinction.*

Markov chain models for discrete-state processes.	*Not restricted to finitely many states.*

Transition matrices: 1-step and *n*-step.
Classification of states.
Equilibrium distributions for time-homogeneous chains.

Poisson processes.
Differential-difference equations.
Birth and death processes.

Queues.

The M/M/1 queue. Differential-difference equa-tions. Conditions for equilibrium. Equilibrium distributions of queue size and waiting time for first-come-first-served queues. Extensions to M/M/*k* and M/M/ ∞ queues.	*Equilibrium behaviour for queues with transition rates dependent on queue size.*
The M/G/1 queue, imbedded Markov chain analysis. The Pollaczek-Khintchine formula. Mean queue length and waiting time.	*Equilibrium treatment only.*
	In examination questions, the word "queue" will refer to all units in a system, i.e. those being served as well as those still waiting to be served.

Time series

Time series models; trend and seasonality.	*Additive and multiplicative models.*

Stationarity.
Autocovariance, autocorrelation and partial autocorrelation functions. Correlograms.

Autoregressive (AR) processes.	*Yule-Walker equations.*
Moving average (MA) processes.	*Invertibility conditions.*

ARMA processes.

ARIMA processes and Box-Jenkins methods.	*Identification, estimation, checking, forecasting. Box-Pierce and Ljung-Box statistics.*
Forecasting and minimising expected prediction variance.	*Exponential smoothing, Holt-Winters.*
Introduction to frequency domain analysis. Spectral density function. Periodograms.	*Candidates will be expected to have some familiarity with the fast Fourier transform.*

MODULE 4: Modelling experimental data

This module covers the general and generalised linear model. It also covers the design of experiments and analysis of experimental results using analysis of variance and multiple regression, extending the coverage of these topics in the Higher Certificate. Candidates are expected to have knowledge of the use of these methods and the interpretation of results from them, including computer output: critical evaluation of computer output is required throughout this module.

General linear model

Least squares. Properties of LS estimators in the linear model, Gauss-Markov theorem. Models for simple and multiple regression and for analysis of variance and covariance. Hat matrix. Estimation of variance. Interval estimates of parameters. Weighted least squares. Importance of assumptions. Analysis of residuals. Transformation of variables. Linearising other models, e.g. multiplicative models and growth curves.

Proof of standard results for multiple regression (matrix notation; matrix differentiation will not be required). Application of these to other situations as listed.

Design of experiments

Randomisation, replication, blocking.

Completely randomised designs, randomised blocks, Latin squares, balanced incomplete blocks. Factorial treatment structure.

2^n designs, including confounding and fractional replication.

Reasons for using these designs; bias and precision. How to construct a valid randomised layout for each design. Knowledge of what balanced incomplete block designs exist for reasonably small block sizes and numbers of treatments.

Analysis of variance

Analysis of variance for the designs listed above, including for cross-classifications with replication, and for nested or hierarchical designs. Fixed and random effects; variance components. Application to data collected in experiments or by sampling.

Estimation of variance components, and use in planning sampling schemes.

General linear contrasts among treatments. Inference for individual treatment means and for contrasts.

Comparisons between and within groups of treatments. Linear and quadratic components for factors at equally spaced levels.

Analysis of residuals. Use of plotting techniques to detect non-Normality of errors.

Multiple regression

Regression with more than one explanatory variable. Use of indicator variables to represent factors. Analysis of variance of regression, including test for lack of fit. Analysis of residuals, regression diagnostics, detection of influential observations, multicollinearity, serial correlation. Model selection, using all-subset and stepwise methods.

The Extra Sum of Squares principle.

Pure error.

Concept of leverage. R^2 and adjusted R^2. Use of the Durbin-Watson statistic.

Including use of C_p plots.

Extension to non-linear modelling. Fitting standard growth curves (including logistic and Gompertz). Estimation of parameters, approximate confidence intervals and tests.

Candidates will be expected to have some familiarity with the Newton-Raphson procedure for fitting non-linear models.

Generalised Linear Model

Theory and examples of use. Link functions.

Logistic regression for analysis of binary data.

The multinomial distribution. Log-linear models for analysis of cross-tabulated categorical data.

Deviance; analysis of deviance. Use of F test for comparing models. Diagnostics.

Poisson regression. Components of a generalised linear model including the exponential family. Adapting to various problems by choosing suitable link functions etc. Candidates will be expected to have some familiarity with the method of iterative reweighted least squares used in fitting models.

MODULE 5: Topics in applied statistics

This module covers four application areas, denoted by bold sub-headings below. It introduces some general techniques for data analysis and some that are more specialised to particular areas. The paper for this module will not be formally divided into sections but it will always consist of two questions on each of the four application areas covered by the syllabus. Candidates are particularly advised that some questions on the paper are likely to require knowledge of material covered in the module on Modelling Experimental Data (module 4).

Critical evaluation of computer output is required throughout this module. An appreciation of the problems raised by missing values in data is expected throughout this module.

Note that this syllabus is continued on the next page.

Multivariate methods

Vectors of expected values. Covariance and correlation matrices.	
Discriminant analysis, choice between two populations, calculation of discriminant function, and probability of misclassification, test and training samples, leave-one-out and *k*-fold cross-validation, idea of extension to several populations.	*Linear and quadratic discrimination.*
Principal components; definition, interpretation of calculated components, use in regression.	*Based on covariance and correlation matrices.*
Cluster analysis, similarity measures, single-link and other hierarchical methods, *k*-means.	
Informal approaches to checking for multivariate Normality. Tests and confidence regions for multivariate means.	*Hotelling's T^2.*

Censored data: survival and reliability

Problems involving censored data, for example in clinical and engineering contexts.	
Reliability and life testing.	
Hazard and survivor functions.	*Confidence intervals for survivor function using Greenwood's formula.*
Kaplan-Meier estimate of survivor function.	
Weibull and hazard plots.	
Logrank test.	
Parametric survival distributions - exponential, Weibull.	*Use of log cumulative hazard plot to check Weibull and proportional hazards assumptions. Use of these distributions in modelling survival data. Fitting methods not required.*
Proportional hazards and Cox regression.	*An understanding of the assumptions and interpretation of the fitted model. Details of partial likelihood and numerical methods for fitting the model not required. Calculation and interpretation of hazard ratios and confidence intervals.*
Checking for non-proportionality of hazards.	*Checking for non-proportional hazards using a log cumulative hazard plot and plots of hazard functions.*

Demography and epidemiology

Population pyramids.	*Construction and use of life tables; derived quantities, including calculation of life expectancy. Direct and indirect standardisation.*
Life tables.	
Standardised rates (e.g. mortality).	

Incidence and prevalence.	*Distinction between these concepts.*
Design and analysis of cohort (prospective) studies.	
Design and analysis of case-control (retro-spective) studies.	
Confounding and interaction.	
Matched case control design and analyses, using McNemar's test.	*Reasons for matching; advantages and dis-advantages relative to unmatched studies.*
Causation.	*Inferring causality from observational studies.*
Relative risk. Odds ratio. Estimation and confidence intervals for 2×2 tables.	
Mantel-Haenszel procedure.	*Use in adjusting for confounding variables.*
Sensitivity, specificity, ROC curves, positive predictive value, negative predictive value.	*Uses in screening and diagnosis.*
Sampling	
Census and sample survey design. Target and study populations, uses and limitations of non-probability sampling methods, sampling frames, sampling fraction.	*Revision and extension of basic concepts from Higher Certificate.*
Simple random sampling. Estimators of totals, means and proportions; bias. Estimated standard errors, confidence intervals and precision. Sampling fraction and finite population correction.	*Examples of practical use in various contexts.*
Ratio and regression estimators.	*Use of supplementary information.*
Stratified random sampling. Estimators of totals, means and proportions; bias. Estimated standard errors, confidence intervals and precision. Cost functions. Proportional and optimal allocations. Limitations of stratified sampling.	*Examples of practical use in various contexts.* *Minimisation of cost × variance.*
One-stage cluster sampling. Estimators for totals, means and proportions with equal cluster sizes and with different cluster sizes. Estimated standard errors, confidence intervals and precision. Link with systematic sampling. Description of two-stage sampling and of multi-stage sampling. Limitations.	*Examples of practical use in various contexts.*