Advanced Higher Course Specification

Advanced Higher Statistics

Course code:	C803 77
Course assessment code:	X803 77
SCQF:	level 7 (32 SCQF credit points)
Valid from:	session 2022–23

This document provides detailed information about the course and course assessment to ensure consistent and transparent assessment year on year. It describes the structure of the course and the course assessment in terms of the skills, knowledge and understanding that are assessed.

This document is for teachers and lecturers and contains all the mandatory information required to deliver the course.

This edition: February 2023 (version 5.0)

Contents

Course overview

This course consists of 32 SCQF credit points, which includes time for preparation for course assessment. The notional length of time for candidates to complete the course is 160 hours.

The course assessment has two components.

Component	Marks	Duration
Component 1: question paper 1	30	1 hour
Component 2: question paper 2	90	2 hours and 45 minutes

Recommended entry	Progression
Entry to this course is at the discretion of the centre. Candidates should have achieved the Higher Mathematics course or equivalent qualifications and/or experience prior to starting this course.	♦ other qualifications in mathematics or related areas ♦ further study, employment and/or training

Conditions of award

The grade awarded is based on the total marks achieved across both course assessment components.

Course rationale

National Courses reflect Curriculum for Excellence values, purposes and principles. They offer flexibility, provide time for learning, focus on skills and applying learning, and provide scope for personalisation and choice.

Every course provides opportunities for candidates to develop breadth, challenge and application. The focus and balance of assessment is tailored to each subject area.

Learning statistics develops logical reasoning, analysis, problem-solving skills, creativity, and the ability to think in abstract ways. It uses a universal language of numbers and symbols, which allows us to communicate ideas in a concise, unambiguous and rigorous way.

The course develops important statistical techniques that allow candidates to make links between statistical models and the real world, facilitating reasoned arguments based on sound logic. The skills, knowledge and understanding in the course also supports learning in technology, health and wellbeing, science, and social studies.

Purpose and aims

Statistics is important in everyday life. It helps us to make sense of inherent natural variation in a wide variety of contexts.

Using statistics enables us to collect, analyse, and interpret data. It equips us with the skills we need to understand the degree of certainty that we can attribute to inferences made and conclusions reached when we interpret and analyse data.

The course aims to:

- motivate and challenge candidates by enabling them to select and apply statistical techniques in a variety of situations
- develop candidates’ understanding of the appropriateness of different methods of data collection, particularly ways of sampling from a population
- enable candidates to select and use appropriate statistical models to analyse data
- allow candidates to consider and evaluate assumptions required for chosen models
- develop candidates’ understanding of the notion of probability
- allow candidates to interpret results in context, evaluating the strength and limitations of their models and conclusions
- develop candidates’ skills in effectively communicating conclusions reached on the basis of statistical analysis

Who is this course for?

This course is particularly suitable for candidates who:

- have demonstrated an aptitude for mathematics
- are interested in developing statistical techniques to use in further study or in the workplace

Course content

The Advanced Higher Statistics course introduces experimental design and instils and nurtures the ability to employ good analytical practice on data sets.

Throughout the course, candidates develop and apply skills in data collection, hypothesis testing, and selecting and using appropriate statistical models.

Candidates develop the ability to make informed judgements on calculated statistics and the ability to communicate appropriate conclusions

Skills, knowledge and understanding

Skills, knowledge and understanding for the course

The following provides a broad overview of the subject skills, knowledge and understanding developed in the course:

- knowledge and understanding of a range of complex statistical concepts
- identifying and using appropriate statistical models
- applying more advanced operational skills in statistical contexts
- using statistical reasoning skills to extract and interpret information, think logically and solve problems
- communicating conclusions, exhibiting appreciation of their limitations
- thinking analytically about the consequences of methodological choices

Skills, knowledge and understanding for the course assessment

The following provides details of skills, knowledge and understanding sampled in the course assessment.

Data analysis and modelling	
Skill	**Explanation**
Interpreting the exploratory data analysis (EDA) of univariate data	♦ presenting and interpreting sample data in an appropriate form using a table, dot plot, stem-and-leaf diagram and box plot ♦ appreciating that there are different methods of data collection and the difference between discrete and continuous data ♦ identifying possible outliers and suggesting possible action to take
Working with theoretical and experimental probabilities	♦ using set notation for probability theory $\mathrm{P}(\mathrm{E} \cup \mathrm{F})$, $\mathrm{P}(\mathrm{E} \cap \mathrm{F})$, $\mathrm{P}(\mathrm{E} \mid \mathrm{F})$ and $\mathrm{P}(\bar{\mathrm{E}})$ and compounds of these, such as $\mathrm{P}(\bar{\mathrm{E}} \cap \mathrm{F})$ ♦ appreciating the necessary conditions for, and use of, the addition and multiplication laws of probability ♦ calculating probabilities for events that are not mutually exclusive ♦ comparing calculated theoretical probabilities with those obtained experimentally, or by simulation, using appropriate technology
Calculating conditional probabilities	♦ calculating simple conditional probabilities ♦ calculating conditional probabilities requiring the use of Bayes' theorem or equivalent methods
Modelling a discrete random variable	♦ constructing the probability distribution of a discrete random variable ♦ generating values of discrete data by simulation or experiment and comparing their distribution to theoretical models ♦ calculating the mean and standard deviation of a discrete random variable
Using the laws of expectation and variance	♦ using the laws of expectation and variance: $\mathrm{E}(aX+b)=a\mathrm{E}(X)+b$ $\mathrm{E}(X \pm Y)=\mathrm{E}(X) \pm \mathrm{E}(Y)$ $\mathrm{E}(aX \pm bY)=a\mathrm{E}(X) \pm b\mathrm{E}(Y)$ $\mathrm{V}(aX+b)=a^2\mathrm{V}(X)$ $\mathrm{V}(X \pm Y)=\mathrm{V}(X)+\mathrm{V}(Y)$, where X and Y are independent ♦ calculating $\mathrm{SD}(aX+bY)$, where X and Y are independent

Data analysis and modelling	
Skill	**Explanation**
Using discrete probability distributions	♦ calculating uniform, binomial and Poisson probabilities ♦ using standard results for the mean and variance of these distributions ♦ simulating these distributions using appropriate technology and comparing them to probability distribution models
Using continuous probability distributions	♦ calculating rectangular (continuous uniform) probabilities and using standard results for the mean and variance of this distribution ♦ calculating and using normal probabilities ♦ calculating probabilities in problems involving the sum or difference of two independent normal random variables
Using the normal approximation to discrete probability distributions	♦ demonstrating an understanding of appropriate conditions for a normal approximation to a binomial or Poisson distribution, together with the parameters of the approximate distribution ♦ demonstrating the use of a continuity correction when applying a normal approximation to the binomial and Poisson distributions

Statistical inference	
Skill	**Explanation**
Identifying and using appropriate random sampling methods	♦ appreciating that there are different methods of data collection, and being able to generate a simple random sample from a population ♦ describing and distinguishing between random sampling methods such as systematic, stratified and cluster sampling ♦ appreciating that non-random sampling methods, such as quota or convenience sampling, could lead to an unrepresentative sample and biased conclusions
Working with the distribution of sample means and sample proportions	♦ demonstrating an understanding that the sampling distribution of the sample mean from a population that is normal is, itself, normal ♦ demonstrating an understanding that the sampling distribution of the sample mean from a population that is not normal is approximately normal, by invoking the central limit theorem when the sample is large enough ♦ describing the sampling distribution of the sample mean and using the appropriate standard error in calculations involving this distribution

Statistical inference	
Skill	**Explanation**
	♦ describing the sampling distribution of the sample proportion and using the appropriate standard error in calculations involving this distribution ♦ using the sample mean as a best estimate of the population mean ♦ using the sample variance as an estimate of the population variance
Obtaining confidence intervals	♦ calculating a z-interval for the population mean ♦ appreciating the need to use Student's t-distribution when the population variance is unknown ♦ calculating an approximate confidence interval for the population proportion
Using control charts	♦ constructing and interpreting a Shewhart control chart for the sample mean or proportion ♦ using the Western Electric rules to recognise when a process may be out of statistical control and influenced by common or special causes ♦ estimating chart parameter(s) from a sample when population data is unavailable
Fitting a linear model to bivariate data	♦ interpreting a scatter plot, observing whether or not a linear model is appropriate ♦ calculating the least squares regression line of y on x and x on y ♦ appreciating the difference between regressing y on x and x on y and being able to assess the linear association for y on x and x on y
Assessing the linear association between two variables	♦ calculating and interpreting the product moment correlation coefficient ♦ calculating and interpreting the coefficient of determination ♦ calculating a fitted value and its residual ♦ interpreting a residual plot ♦ commenting on given simple transformations to obtain improved models
Estimating with bivariate data	♦ assessing the reliability of prediction based on fitted values, considering the effects of correlation, interpolation and extrapolation ♦ calculating the appropriate statistics required for bivariate intervals ♦ constructing a prediction interval for an individual response ♦ constructing a confidence interval for a mean response

<table>
<tr><th colspan="2">Hypothesis testing</th></tr>
<tr><th>Skill</th><th>Explanation</th></tr>
<tr><td>Identifying and performing an appropriate one-sample test for the population mean and proportion</td><td>♦ performing a specified test for the population mean, for the cases:
▪ σ^2 known (z-test)
▪ σ^2 unknown, but a large sample (z-test)
▪ σ^2 unknown, with a small sample (t-test)
♦ performing a z-test for the population proportion
♦ selecting and justifying the choice of an appropriate test, together with its underlying assumptions</td></tr>
<tr><td>Identifying and performing an appropriate two-sample test (independent or paired data) for comparing population means and proportions</td><td>♦ using a t-test to assess evidence about the population mean difference in a paired data experiment
♦ testing the hypothesis that two populations have the same mean, for cases where population variances are:
▪ known (z-test)
▪ unknown, but samples are large (z-test)
▪ unknown, and samples are small (t-test)
♦ testing the hypothesis that two populations have the same proportion for only the case where both samples are large
♦ selecting and justifying the choice of an appropriate test, together with its underlying assumptions</td></tr>
<tr><td colspan="2">Non-parametric tests make no assumptions about the distributional form of populations, for example normality. As a result, the hypotheses are often framed in terms of medians rather than means.

The use of ranks may help to reduce the influence of outliers in the data.

The use of a continuity correction is expected if a normal approximation is employed. Formulae for the mean and variance of the test statistic are given.</td></tr>
<tr><td>Identifying and performing an appropriate test for population median(s)</td><td>♦ using a Wilcoxon signed-rank test to assess evidence about the population median from a simple random sample and about the population distributions from paired data
♦ using a Mann-Whitney test to assess evidence about the medians of two populations using independent samples
♦ using a normal approximation, when required in any calculation of a test statistic or p value
♦ selecting and justifying the choice of an appropriate test, together with its underlying assumptions</td></tr>
<tr><td>Identifying and performing an appropriate chi-squared test</td><td>♦ performing a chi-squared test for goodness-of-fit to a discrete distribution
♦ performing a chi-squared test for association in a contingency table
♦ dealing with small expected frequencies</td></tr>
</table>

Hypothesis testing	
Skill	**Explanation**
Identifying and performing an appropriate hypothesis test on bivariate data	♦ testing the hypothesis that the slope parameter in a linear model is zero ♦ testing the hypothesis that the population correlation coefficient is zero ♦ communicating appropriate assumptions

Skills, knowledge and understanding included in the course are appropriate to the SCQF level of the course. The SCQF level descriptors give further information on characteristics and expected performance at each SCQF level, and are available on the SCQF website.

Skills for learning, skills for life and skills for work

This course helps candidates to develop broad, generic skills. These skills are based on SQA's Skills Framework: Skills for Learning, Skills for Life and Skills for Work and draw from the following main skills areas:

2 Numeracy
2.1 Number processes
2.2 Money, time and measurement
2.3 Information handling

5 Thinking skills
5.3 Applying
5.4 Analysing and evaluating

You must build these skills into the course at an appropriate level, where there are suitable opportunities.

Course assessment

Course assessment is based on the information in this course specification.

The course assessment meets the purposes and aims of the course by addressing:

- breadth — drawing on knowledge and skills from across the course
- challenge — requiring greater depth or extension of knowledge and/or skills
- application — requiring application of knowledge and/or skills in practical or theoretical contexts as appropriate

This enables candidates to:

- use a range of complex statistical concepts
- identify and use appropriate statistical models and skills
- use statistical reasoning skills to extract and interpret information, think logically, and evaluate evidence
- communicate conclusions, exhibiting appreciation of their limitations
- explain the consequences of choice of method

Course assessment structure: question paper

Question paper 1 — 30 marks

This question paper allows candidates to demonstrate the application of statistical skills, knowledge and understanding from across the course. Candidates can use a calculator.

This question paper gives candidates an opportunity to apply the statistical analysis specified in the ‘Skills, knowledge and understanding for the course assessment’ section.

This question paper has 30 marks out of a total of 120 marks for the course assessment. It consists of short-answer and extended-response questions.

Setting, conducting and marking the question paper

This question paper is set and marked by SQA, and conducted in centres under conditions specified for external examinations by SQA.

Candidates have 1 hour to complete this question paper.

Question paper 2 — 90 marks

This question paper allows candidates to demonstrate the application of statistical skills, knowledge and understanding from across the course. Candidates can use a calculator.

This question paper gives candidates an opportunity to apply an understanding of the underlying processes involved in hypothesis testing, statistical inference and data analysis specified in the ‘Skills, knowledge and understanding for the course assessment’ section.

This question paper has 90 marks out of a total of 120 marks for the course assessment. It consists of short-answer and extended-response questions.

Setting, conducting and marking the question paper

This question paper is set and marked by SQA, and conducted in centres under conditions specified for external examinations by SQA.

Candidates have 2 hours and 45 minutes to complete this question paper.

Specimen question papers for Advanced Higher courses are published on SQA’s website. These illustrate the standard, structure and requirements of the question papers. The specimen papers also include marking instructions.

Grading

Candidates’ overall grades are determined by their performance across the course assessment. The course assessment is graded A–D on the basis of the total mark for both course assessment components.

Grade description for C

For the award of grade C, candidates will typically have demonstrated successful performance in relation to the skills, knowledge and understanding for the course.

Grade description for A

For the award of grade A, candidates will typically have demonstrated a consistently high level of performance in relation to the skills, knowledge and understanding for the course.

Equality and inclusion

This course is designed to be as fair and as accessible as possible with no unnecessary barriers to learning or assessment.

Guidance on assessment arrangements for disabled candidates and/or those with additional support needs is available on the assessment arrangements web page: www.sqa.org.uk/assessmentarrangements.

Further information

- Advanced Higher Statistics subject page
- Assessment arrangements web page
- Building the Curriculum 3–5
- Guide to Assessment
- Guidance on conditions of assessment for coursework
- SQA Skills Framework: Skills for Learning, Skills for Life and Skills for Work
- Coursework Authenticity: A Guide for Teachers and Lecturers
- Educational Research Reports
- SQA Guidelines on e-assessment for Schools
- SQA e-assessment web page
- SCQF website: framework, level descriptors and SCQF Handbook

Appendix 1: course support notes

Introduction

These support notes are not mandatory. They provide advice and guidance to teachers and lecturers on approaches to delivering the course. Please read these course support notes in conjunction with the course specification and the specimen question papers.

Approaches to learning and teaching

Approaches to learning and teaching should be engaging, with opportunities for personalisation and choice built in where possible. These could include:

- using active and open-ended learning activities, such as research, case studies, project-based tasks and presentation tasks
- encouraging candidates to engage in independent reading from a range of sources, including the internet
- demonstrating how candidates should record the results of their research and independent investigation from different sources
- asking candidates to share the findings and conclusions of their research and investigation activities in a presentation
- using collaborative learning opportunities to develop team working
- a mix of collaborative, co-operative or independent tasks that engage candidates
- solving problems and thinking critically
- explaining thinking and presenting strategies and solutions to others
- using questioning and discussion to encourage candidates to explain their thinking and to check their understanding of fundamental concepts
- making links in themes that cut across the curriculum to encourage transferability of skills, knowledge and understanding — including with technology, geography, sciences, social subjects, and health and wellbeing
- debating and discussing topics and concepts so that candidates can demonstrate skills in constructing and sustaining lines of argument to provide challenge, enjoyment, breadth, and depth in their learning
- drawing conclusions from complex information
- using sophisticated written and/or oral communication and presentation skills to present information
- using technological and media resources, for example BBC's *More or Less: Behind the Stats* podcast, and the Royal Statistical Society publication and website, *Significance*
- using real-life contexts and experiences familiar and relevant to candidates to hone and exemplify skills, knowledge and understanding

You should support candidates by having regular discussions with them and giving them regular feedback. For group activities, candidates could also receive feedback from their peers.

You should, where possible, provide opportunities for candidates to personalise their learning and give them choices about learning and teaching approaches. The flexibility in Advanced Higher courses and the independence with which candidates carry out the work lend themselves to this.

You should use inclusive approaches to learning and teaching. There may be opportunities to contextualise approaches to learning and teaching to Scottish contexts in this course. You could do this through mini-projects or case studies.

Preparing for course assessment

The course assessment focuses on breadth, challenge and application. Candidates draw on and extend the skills they have learned during the course. These are assessed through two question papers; candidates can use a calculator in both.

To help candidates prepare for the course assessment, they should have the opportunity to:

- analyse a range of real-life problems and situations involving data and statistics
- select and adapt appropriate statistical skills
- apply statistical skills with the aid of technology
- determine solutions
- explain solutions and/or relate them to context
- present statistical information appropriately

The question papers assess a selection of knowledge and skills acquired during the course and provide opportunities for candidates to apply skills in a wide range of situations, some of which may be new.

Before the course assessment, candidates may benefit from reviewing and interpreting statistical reports, and responding to short-answer questions and extended-response questions.

Developing skills for learning, skills for life and skills for work

You should identify opportunities throughout the course for candidates to develop skills for learning, skills for life and skills for work.

Candidates should be aware of the skills they are developing and you can provide advice on opportunities to practise and improve them.

SQA does not formally assess skills for learning, skills for life and skills for work.

There may also be opportunities to develop additional skills depending on the approach centres use to deliver the course. This is for individual teachers and lecturers to manage.

Some examples of potential opportunities to practise or improve these skills are provided in the following table.

SQA skills for learning, skills for life and skills for work framework definition	Suggested approaches for learning and teaching
Numeracy is the ability to use numbers to solve problems by counting, doing calculations, measuring, and understanding graphs and charts. It is also the ability to understand the results.	Candidates could: ♦ develop their numerical skills throughout the course, for example by assessing significance levels ♦ use numbers to solve contextualised problems involving other STEM subjects ♦ manage problems, tasks and case studies involving numeracy by analysing the context, carrying out calculations, drawing conclusions, and making deductions and informed decisions
Applying is the ability to use existing information to solve a problem in a different context, and to plan, organise and complete a task.	Candidates could: ♦ apply the skills, knowledge and understanding they have developed to solve statistical problems in a range of real-life contexts ♦ think creatively to adapt strategies to suit the given problem or situation ♦ show and explain their thinking to determine their level of understanding ♦ think about how they are going to tackle problems or situations, decide which skills to use, and then carry out the calculations necessary to complete the task, for example using normal approximations
Analysing and evaluating is the ability to identify and weigh up the features of a situation or issue and to use judgement to come to a conclusion. It includes reviewing and considering any potential solutions.	Candidates could: ♦ identify which real-life tasks or situations require the use of statistics ♦ interpret the results of their calculations and draw conclusions; conclusions drawn could be used to form the basis of making choices or decisions ♦ identify and analyse situations involving statistics that are of personal interest

During the course, candidates have opportunities to develop their literacy skills and employability skills.

Literacy skills are particularly important as these skills allow candidates to access, engage in and understand their learning, and to communicate their thoughts, ideas and opinions. The course provides candidates with the opportunity to develop their literacy skills by analysing real-life contexts and communicating their thinking by presenting statistical information in a variety of ways. This could include the use of numbers, formulae, diagrams, graphs, symbols and words.

Employability skills are the personal qualities, skills, knowledge, understanding and attitudes required in changing economic environments. Candidates can apply the statistical operational and reasoning skills developed in this course in the workplace. The course provides them with the opportunity to analyse a situation, decide which statistical strategies to apply, work through those strategies effectively, and make informed decisions based on the results.

Appendix 2: skills, knowledge and understanding with suggested learning and teaching contexts

The first two columns are identical to the tables of 'Skills, knowledge and understanding for the course assessment' in the course specification.

The third column gives suggested learning and teaching contexts. These provide examples of where the skills could be used in individual activities or pieces of work.

Candidates could benefit from using a statistical calculator during this course.

Data analysis and modelling		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Interpreting the exploratory data analysis (EDA) of univariate data	♦ presenting and interpreting sample data in an appropriate form using a table, dot plot, stem-and-leaf diagram and box plot ♦ appreciating that there are different methods of data collection and the difference between discrete and continuous data	♦ categorising data as discrete or continuous ♦ applying EDA techniques to all data to establish reasonable assumptions about the underlying population
	♦ identifying possible outliers and suggesting possible action to take	♦ communicating reasons for judgements and suggestions ♦ using fences to identify outliers within a data set: values that lie beyond fences are considered to be possible outliers ($Q_1 - 1{\cdot}5 \times IQR$ is a commonly used definition of a lower fence. $Q_3 + 1{\cdot}5 \times IQR$ is a commonly used definition of an upper fence. In each definition, IQR represents the interquartile range.)

Data analysis and modelling		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Working with theoretical and experimental probabilities	♦ using set notation for probability theory $P(E \cup F)$, $P(E \cap F)$, $P(E \mid F)$ and $P(\bar{E})$ and compounds of these, such as $P(\bar{E} \cap F)$	
	♦ appreciating the necessary conditions for, and use of, the addition and multiplication laws of probability	♦ using well-constructed experiments to demonstrate the concepts of probability and emphasise uncertainty as a central tenet of statistics
	♦ calculating probabilities for events that are not mutually exclusive	♦ exploring simple, practical examples of mutually exclusive events and non-mutually exclusive events (This illustrates that calculating probabilities in each case is different and can give very different results depending on the exclusivity, or not, of two or more events.) ♦ using Venn diagrams as a visual aid
	♦ comparing calculated theoretical probabilities with those obtained experimentally, or by simulation, using appropriate technology	♦ exploring the accuracy and limitations of simulations in predicting further events ♦ exploring how closely theoretical probabilities resemble actual events
Calculating conditional probabilities	♦ calculating simple conditional probabilities	♦ encouraging candidates to establish conditional probabilities intuitively and to formalise calculations using appropriate notation
	♦ calculating conditional probabilities requiring the use of Bayes' theorem or equivalent methods	♦ using Bayes' theorem to reverse the condition, finding $P(E \mid F)$, given $P(F \mid E)$ ♦ using equivalent methods to Bayes' theorem: ▪ tree diagrams ▪ tabulated probabilities ▪ set notation

Data analysis and modelling		
Skill	**Explanation**	**Suggested learning and teaching contexts**
		Candidates do not need to produce a full algebraic treatment of Bayes' theorem. Candidates should understand the concept of combining several events into two, E and $\bar{\text{E}}$, and practise doing this.
Modelling a discrete random variable	♦ constructing the probability distribution of a discrete random variable ♦ generating values of discrete data by simulation or experiment and comparing their distribution to theoretical models ♦ calculating the mean and standard deviation of a discrete random variable	♦ repeating an experiment or simulation to compare the results with a mathematical model ♦ investigating a model's strengths in accurately predicting long-term results, and its limitations in terms of immediate prediction
Using the laws of expectation and variance	♦ Using the laws of expectation and variance: $\text{E}(aX+b)=a\text{E}(X)+b$ $\text{E}(X\pm Y)=\text{E}(X)\pm\text{E}(Y)$ $\text{E}(aX\pm bY)=a\text{E}(X)\pm b\text{E}(Y)$ $\text{V}(aX+b)=a^2\text{V}(X)$ $\text{V}(X\pm Y)=\text{V}(X)+\text{V}(Y)$, where X and Y are independent ♦ calculating $\text{SD}(aX+bY)$, where X and Y are independent	♦ using simple examples to demonstrate the general results ♦ emphasising the difference between variance and standard deviation ♦ demonstrating how candidates should apply the laws to variance and convert to or from standard deviation, where necessary ♦ discussing why statisticians use standard deviation more ♦ using graphs to highlight $\text{V}(-X)=\text{V}(X)$ and $\text{V}(X+b)=\text{V}(X)$

Data analysis and modelling		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Using discrete probability distributions	♦ calculating uniform, binomial and Poisson probabilities ♦ using standard results for the mean and variance of these distributions ♦ simulating these distributions using appropriate technology and comparing them to probability distribution models	♦ selecting an appropriate distribution to model data from a given context ♦ using EDA and/or other techniques, such as chi-squared goodness-of-fit test, to confirm or establish a possible distribution of a data set ♦ introducing candidates to ${}^{n}C_{r}$ in a variety of practical contexts — useful links include Pascal's triangle, the binomial theorem, and how combinations are relevant in some algebraic expansions ♦ introducing and discussing assumptions required for a particular model ♦ using cumulative probability tables or functions on calculators and spreadsheets ♦ introducing the idea of hypothesis-testing in an informal way, using a binomial distribution; candidates only need to decide if a result is probable or not, given the assumed binomial model ♦ ensuring candidates know how to use the standard results from the statistical formulae and tables ♦ using software packages or websites to generate simulated data and modelled results for chosen parameters, and then displaying a comparison of the two in an EDA; candidates should be able to select a model, explain their decisions relating to the model, and make suggestions for improving the quality of the model

Data analysis and modelling		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Using continuous probability distributions	♦ calculating rectangular (continuous uniform) probabilities and using standard results for the mean and variance of this distribution ♦ calculating and using normal probabilities ♦ calculating probabilities in problems involving the sum or difference of two independent normal random variables	♦ ensuring candidates know how to use the standard results from the statistical formulae and tables ♦ ensuring candidates are familiar with working from probabilities back to parameter values ♦ ensuring candidates are familiar with the laws of expectation and variance and their application in the context of combining independent normal distributions
Using the normal approximation to discrete probability distributions	♦ demonstrating an understanding of appropriate conditions for a normal approximation to a binomial or Poisson distribution, together with the parameters of the approximate distribution	♦ considering the reason (ease of calculation without significant loss of accuracy) for approximating a binomial or Poisson distribution ♦ understanding the limitations of approximations and the necessary conditions for approximations to be sufficiently accurate For this course, candidates should: ♦ use the normal approximation to a binomial distribution, when $np > 5$ and $nq > 5$ ♦ use the normal approximation to a Poisson distribution, when $\lambda > 10$ Candidates could use calculators or software to investigate the validity of these rules.

Data analysis and modelling		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ demonstrating the use of a continuity correction when applying a normal approximation to the binomial and Poisson distributions	Where the distribution of the discrete variable is known, candidates can calculate exact theoretical probabilities without resorting to the normal approximation. Candidates need to know that normal approximations can be used even when the underlying discrete distribution is known, for example comparing probabilities calculated under a normal approximation with those using the binomial or Poisson distributions directly to establish the accuracy of approximations.

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Identifying and using appropriate random sampling methods	♦ appreciating that there are different methods of data collection, and being able to generate a simple random sample from a population	♦ understanding the difference between random and non-random sampling
	♦ describing and distinguishing between random sampling methods such as systematic, stratified and cluster sampling	♦ discussing the difference between a sample and a census ♦ discussing the advantages and disadvantages of each type of sampling ♦ ensuring candidates know that (in this course) stratified sampling is proportional stratified sampling ♦ ensuring candidates know the difference between one-stage cluster sampling and two-stage cluster sampling
	♦ appreciating that non-random sampling methods, such as quota or convenience sampling, could lead to an unrepresentative sample and biased conclusions	♦ ensuring candidates understand that statistical inference can only reliably be done on representative, random samples from populations, and should not be done on a population if it is known
Working with the distribution of sample means and sample proportions	♦ demonstrating an understanding that the sampling distribution of the sample mean from a population that is normal is, itself, normal	♦ using online simulations to clearly illustrate the distribution of sample means from a variety of populations
	♦ demonstrating an understanding that the sampling distribution of the sample mean from a population that is not normal is approximately normal, by invoking the central limit theorem when the sample is large enough	♦ ensuring candidates understand that for sufficiently large n (in this course $n \geq 20$), the distribution of the sample mean is approximately normal, irrespective of the distribution of the population ♦ ensuring candidates can quote and use: $\overline{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ describing the sampling distribution of the sample mean and using the appropriate standard error in calculations involving this distribution	♦ introducing the quantity $\frac{\sigma}{\sqrt{n}}$ as the standard error of the sample mean
	♦ describing the sampling distribution of the sample proportion and using the appropriate standard error in calculations involving this distribution	♦ ensuring candidates understand that for $np > 5$, $nq > 5$, the distribution of a sample proportion is approximately normal ♦ ensuring candidates understand and use: $\frac{X}{n} \approx N\left(p, \frac{pq}{n}\right)$, where p is estimated from $\hat{p}$ ♦ explaining that the distribution is approximate, not the parameters; deriving the formulae for these parameters might be helpful ♦ explaining that the quantity $\sqrt{\frac{pq}{n}}$ is often referred to as the standard error of the sample proportion
	♦ using the sample mean as a best estimate of the population mean	♦ ensuring that candidates understand that in elementary sampling theory the population mean is estimated by the sample mean, the population proportion is estimated by the sample proportion, and the population variance by the sample variance, given by: $s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}$

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ using the sample variance as an estimate of the population variance	♦ demonstrating the reasoning behind the $n-1$ denominator by: ▪ candidates generating their own sample mean distributions ▪ investigating denominators of n and $n-1$, or different sample sizes ▪ using software to run simulations ♦ working with the distribution of sample means or proportions, and clarifying precisely which statistics are calculated from the sample and which parameters are assumed for the population
Obtaining confidence intervals	♦ calculating a z-interval for the population mean	♦ undertaking some practical sampling ♦ discussing what can be inferred about the population, particularly with reference to whether the population variance is known ♦ discussing common statistical contexts where population variance is assumed If σ is known, a 95% confidence interval for the population mean is given by $\bar{x} \pm 1{\cdot}96\frac{\sigma}{\sqrt{n}}$. Other percentage levels of confidence intervals can be calculated using different decimals.

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ appreciating the need to use Student's *t*-distribution when the population variance is unknown	♦ online simulations that illustrate when the *t*-distribution should be used in place of the normal distribution (The shape of the *t*-distribution depends on the number of degrees of freedom, $\upsilon = n - 1$. As $\upsilon \to \infty$ the *t*-distribution tends to the normal distribution.) ♦ ensuring that candidates use the *t*-distribution table in the statistical formulae and tables booklet for $\upsilon \leq 40$ (For $\upsilon > 40$, candidates can use the normal distribution. Note that this approximation is not driven by sample sizes, especially when doing a paired test, but by the number of degrees of freedom involved.) If σ is unknown, an approximate 95% confidence interval for the population mean is given by $\bar{x} \pm t_{n-1,0{\cdot}975} \frac{s}{\sqrt{n}}$. This can be interpreted as saying that the best estimate of the population mean μ is the sample mean $\bar{x}$, but it is recognised that this is only an estimate of the population mean. The true value of the population mean lies within a range of possible values, dependent on the size of the sample. A useful definition of a 95% confidence interval is that if a large number of samples is taken and a confidence interval is computed for each, then 95% of these intervals would be expected to contain the population mean.

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ calculating an approximate confidence interval for the population proportion	Many candidates find this concept challenging. Working with real data can be helpful. An approximate 95% confidence interval for the population proportion p is given by $\hat{p} \pm 1{\cdot}96\sqrt{\frac{\hat{p}\hat{q}}{n}}$ for $n\hat{p} > 5$ and $n\hat{q} > 5$, where $\hat{q} = 1 - \hat{p}$ and $\hat{p}$ is the sample proportion. While a confidence interval for the population mean is often calculated with an assumed population variance, this is not the case with the population proportion, and the $\sqrt{\frac{\hat{p}\hat{q}}{n}}$ term is calculated using the estimate of the population proportion p .
Using control charts	♦ constructing and interpreting a Shewhart control chart for the sample mean or proportion	♦ explaining how Shewhart charts attempt to clearly illustrate and quantify the variation inherent in natural processes and allow some statistical determination of whether or not such processes are proceeding in an expected manner ♦ using a 3-sigma Shewhart chart with control limits drawn three standard deviations either side of the expected value: ▪ the control limits for the sample mean are $\mu \pm 3\frac{\sigma}{\sqrt{n}}$ ▪ the control limits for the sample proportion are $p \pm 3\sqrt{\frac{pq}{n}}$ ▪ note the similarity between control limits for the sample proportion and confidence interval for the population proportion

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
		Candidates are not expected to work with a control chart when $np<5$ or $nq<5$ because the normal approximation to the binomial distribution is poor when $np<5$ or $nq<5$.
	♦ using the Western Electric rules to recognise when a process may be out of statistical control and influenced by common or special causes	♦ using the Western Electric rules to determine when a process may be out of statistical control; these are: ▪ Any single data point falls outside the 3σ limits. ▪ Two out of three consecutive points fall beyond the same 2σ limit. ▪ Four out of five consecutive points fall beyond the same 1σ limit. ▪ Eight consecutive points fall on the same side of the centre line. Candidates should know that each of these rules has a probability of less than 0·01, and be able to calculate the probabilities of each of these four events.
	♦ estimating chart parameter(s) from a sample when population data is unavailable	♦ using the sample mean as the best estimate of the population mean when a population mean is unavailable; estimating the population variance is beyond the scope of this course ♦ estimating the population proportion from the sample proportion and, in this case, estimating the population variance

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Fitting a linear model to bivariate data	♦ interpreting a scatter plot, observing whether or not a linear model is appropriate ♦ calculating the least squares regression line of y on x and x on y	♦ using a scatter plot to give a quick impression of the degree of correlation (for a linear relationship) or association (for a non-linear relationship) between the variables ♦ using the linear model $Y_i = \alpha + \beta x_i + \varepsilon_i$, where Y_i is the expected value for a given x_i, α and β are the population y-intercept and gradient respectively, and ε_i is the error term, which may arise from an error in measurement or from natural variation. The model assumes that: ▪ ε_i are independent ▪ $\mathrm{E}(\varepsilon_i) = 0$ ▪ $\mathrm{V}(\varepsilon_i) = \sigma^2$ (a constant for all x_i) ♦ estimates for α and β, a and b, which give the fitted line $y = a + bx$; these are calculated using: 1 $b = \frac{S_{xy}}{S_{xx}}$ and $a = \bar{y} - b\bar{x}$, where 2 $S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$ $S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$ (used in next section) 3 $S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
		Some candidates might find it interesting to derive these formulae, although this is not assessed in this course. There is an opportunity to combine several aspects of mathematics to derive the formulae used for b and a; for example, the minimum value of $\sum \varepsilon_i^2$ can be found by completing the square or using differential calculus.
	♦ appreciating the difference between regressing y on x and x on y and being able to assess the linear association for y on x and x on y	Candidates should understand that: ♦ the regression of y on x gives a formula for y in terms of x, and is used to predict a y value for a given x ♦ the regression of x on y gives a formula for x in terms of y, and is used to predict an x value for a given y
Assessing the linear association between two variables	♦ calculating and interpreting the product moment correlation coefficient	Candidates should understand that: ♦ Pearson's product moment correlation coefficient r is calculated using $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ and measures the strength of linear association between two variables ♦ correlation is not the same as causation

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ calculating and interpreting the coefficient of determination	♦ the coefficient of determination R^2 is given by $R^2 = \frac{S^2{}_{xy}}{S_{xx}\,S_{yy}}$ and is the square of Pearson's product moment correlation ♦ R^2 gives the proportion of the total variation in the response variable that is explained by the linear model, and in some spreadsheets is the value given for lines of best fit; a small value can indicate that the line is of little use for prediction
	♦ calculating a fitted value and its residual	♦ the hat notation is used to specify a fitted value (one obtained from the line of best fit), as opposed to a data value ♦ the calculated (or given) line of best fit $\hat{Y}_i = a + bx_i$ is used to obtain a fitted value for a given x_i ♦ for data point (x_i, y_i) the fitted value at x_i is $\hat{Y}_i$ and the residual e_i is given by $y_i - \hat{Y}_i$
	♦ interpreting a residual plot	♦ a residual plot is used to check the model assumptions that: ▪ $\mathrm{E}(\varepsilon_i) = 0$ ▪ $\mathrm{V}(\varepsilon_i) = \sigma^2$ (a constant for all x_i) Ideally, the plot of residuals against fitted values should show a random scatter centred on zero. If this is not the case (systematic pattern or variance of residuals is not constant), then the model may be inappropriate (perhaps non-linear), or the data may need to be transformed to restore constant variance.

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ commenting on given simple transformations to obtain improved models	♦ if data transformation is appropriate and successful, any assumptions previously not satisfied should now be clear in a second residual plot ♦ the simple transformations given would come from the Tukey ladder of powers, which include squaring, square rooting, and logarithms
Estimating with bivariate data	♦ assessing the reliability of prediction based on fitted values, considering the effects of correlation, interpolation and extrapolation ♦ calculating the appropriate statistics required for bivariate intervals	Even with fairly strong correlation, referring to a fitted value as a definite outcome is not appropriate. (See also prediction intervals.) ♦ using data about gold medal-winning times in the Olympic 100 metres, published in *Nature*, to give an interesting focus for discussion about interpolation and extrapolation (Tatem, A., Guerra, C., Atkinson, P. & Hay, S. (2004) Momentous sprint at the 2156 Olympics?. *Nature*, 431(7008), 525-525. Available from: doi: 10.1038/431525a. [Published online 29 September 2004]. ♦ using the sum of the squared residuals (SSR) given by $\sum\left(y_i - \hat{Y}_i\right)^2$ to show that SSR = $S_{yy} - \frac{S^2{}_{xy}}{S_{xx}}$ (Some candidates may be interested in deriving this alternative formula, although it is not assessed in this course.) ♦ explaining that the variance of the residuals σ^2 is estimated by $s^2 = \frac{\text{SSR}}{n-2}$

Statistical inference		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ constructing a prediction interval for an individual response ♦ constructing a confidence interval for a mean response	♦ explaining a further assumption that $\varepsilon_i \sim N(0, \sigma^2)$ permits the construction of a $100(1-\alpha)\%$ prediction interval for an individual response $Y_i \mid x_i$, which is given by $\hat{Y}_i \pm t_{n-2,1-\alpha/2}\, s\sqrt{1+\frac{1}{n}+\frac{(x_i-\overline{x})^2}{S_{xx}}}$ ♦ explaining that a $100(1-\alpha)\%$ confidence interval for a mean response $\mathrm{E}(Y_i \mid x_i)$ is given by $\hat{Y}_i \pm t_{n-2,1-\alpha/2}\, s\sqrt{\frac{1}{n}+\frac{(x_i-\overline{x})^2}{S_{xx}}}$ Candidates should understand that in a prediction or confidence interval, the reliability of the estimate depends on the sample size, the variability in the sample and the value of x_i .

Hypothesis testing		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Identifying and performing an appropriate one-sample test for the population mean and proportion	♦ performing a specified test for the population mean, for the cases: ▪ σ^2 known (z-test) ▪ σ^2 unknown, but a large sample (z-test) ▪ σ^2 unknown, with a small sample (t-test)	Candidates should understand the following terms: ♦ null hypothesis H_0 ♦ alternative hypothesis H_1 ♦ level of significance ♦ one-tailed test ♦ two-tailed test ♦ distribution under H_0 ♦ test statistic ♦ critical value ♦ critical region ♦ p value ♦ reject H_0 Learning and teaching contexts could include: ♦ a formal approach to hypothesis testing: ▪ State the hypotheses, the level of significance and whether it is a one- or two-tailed test. ▪ Compute, under H_0, the test statistic and/or p value. ▪ Reject or do not reject H_0, with reference to evidence. ▪ Communicate the conclusion in context. ♦ explaining that both the z-test and t-test are concerned with making inferences about population means and have the underlying assumption that populations are distributed normally, with known or

Hypothesis testing		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ performing a z-test for the population proportion ♦ selecting and justifying the choice of an appropriate test, together with its underlying assumptions	unknown variance respectively, and that sample values are independent ♦ introducing candidates to a continuity correction of $\pm\frac{1}{2n}$ to make the test comparable to that for using the normal approximation to the binomial distribution, although this is not a requirement Candidates should be aware of which formulae are given in the statistical formulae and tables booklet.
Identifying and performing an appropriate two-sample test (independent or paired data) for comparing population means and proportions	♦ using a t-test to assess evidence about the population mean difference in a paired data experiment ♦ testing the hypothesis that two populations have the same mean, for cases where population variances are: ▪ known (z-test) ▪ unknown, but samples are large (z-test) ▪ unknown, and samples are small (t-test)	♦ with paired data, working with the difference between pair values and hence a single distribution paired sample t-test, even for small n, where $t_{n-1}=\dfrac{\bar{X}_d-\mu_d}{\frac{s_d}{\sqrt{n}}}$ where $\mu_d=0$ ♦ comparing two samples of non-paired independent data, each with their own mean and variance, from two populations; this is a more complex undertaking than with a single sample from a given population If the two populations can be assumed to be normal and independent, with either known variances or both sample sizes exceed 20, then, in this course, a two-sample z-test may be used where $Z=\dfrac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$.

Hypothesis testing		
Skill	**Explanation**	**Suggested learning and teaching contexts**
		♦ using the laws of expectation and variance ♦ assuming that the population variances are equal, using a two-sample t-test, where $t_{n_1+n_2-2} = \dfrac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ where $s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
	♦ testing the hypothesis that two populations have the same proportion for only the case where both samples are large	♦ using the population proportion test, where $(n_i p_i > 5$ and $n_i q_i > 5)$, $Z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ ~N(0,1), where p is the pooled proportion $\dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ that is an estimate for the population proportion and $q = 1 - p$
	♦ selecting and justifying the choice of an appropriate test, together with its underlying assumptions	Candidates should be aware of which formulae are given in the statistical formulae and tables booklet.
Non-parametric tests make no assumptions about the distributional form of populations, for example normality. As a result, the hypotheses are often framed in terms of medians rather than means. The use of ranks may help to reduce the influence of outliers in the data. The use of a continuity correction is expected if a normal approximation is employed. Formulae for the mean and variance of the test statistic are given.		

Hypothesis testing		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Identifying and performing an appropriate test for population median(s)	♦ using a Wilcoxon signed-rank test to assess evidence about the population median from a simple random sample and about the population distributions from paired data	Candidates should understand that the Wilcoxon tests assume that the populations are symmetrical (not necessarily normal), and thus the means and medians are equal, so that the null hypothesis can refer to either of these. Learning and teaching contexts could include: ♦ using a table of critical values (appreciating that the smaller sum of ranks is the test statistic) ♦ employing a normal approximation for sample sizes greater than 20
	♦ using a Mann-Whitney test to assess evidence about the medians of two populations using independent samples	Candidates should understand that the Mann-Whitney test assumes that the two populations have the same shape and variability. The table of critical values is provided in the data booklet for sample sizes up to 20. Learning and teaching contexts could include: ♦ using a normal approximation for sample sizes greater than 20 ♦ finding a p value from first principles, leading to an understanding of a table of combinations
	♦ using a normal approximation, when required in any calculation of a test statistic or p value	♦ candidates should be aware of which formulae are given in the statistical formulae and tables booklet
	♦ selecting and justifying the choice of an appropriate test, together with its underlying assumptions	♦ candidates should know to recognise whether they are dealing with either paired or non-paired data

Hypothesis testing		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Identifying and performing an appropriate chi-squared test	♦ performing a chi-squared test for goodness-of-fit to a discrete distribution ♦ performing a chi-squared test for association in a contingency table ♦ dealing with small expected frequencies	Candidates should know: ♦ the appropriate number of degrees of freedom, $k-1-m$, for a 'goodness-of-fit test', where m is the number of parameters that needs to be estimated in order to find the expected frequencies (for example 1 for a Poisson distribution if the mean is not stated) ♦ $(r-1)(c-1)$ for a test of association, r is the number of rows in a contingency table and c the number of columns Candidates should understand that: ♦ the chi-squared test is a non-parametric test ♦ when using a chi-squared statistic, approximating a discrete distribution with a continuous one, the approximation is not reliable if expected frequencies are too small ♦ when working with small expected frequencies, for a reliable test: ▪ at least 80% of expected frequencies should be ≥ 5 ▪ none should be < 1 To meet these criteria, candidates may have to combine categories or frequencies; this could result in a loss in the number of degrees of freedom. ♦ they do not need to use Yates' correction for a 2×2 contingency table, but all expected frequencies should be ≥ 5 Candidates should appreciate that when rejecting the null hypothesis for the chi-squared test of association, the test itself does not reveal the type of relationship between the data in the rows and columns. Candidates could compare the expected frequencies to the observed frequencies in order to conjecture a possible relationship.

Hypothesis testing		
Skill	**Explanation**	**Suggested learning and teaching contexts**
Identifying and performing an appropriate hypothesis test on bivariate data	♦ testing the hypothesis that the slope parameter in a linear model is zero	Candidates should understand that: ♦ the β -test (for the slope parameter being zero) makes the assumption that ε_i are independent and identically distributed $N(0,\sigma^2)$, and uses the *t*-statistic $t = \frac{b\sqrt{S_{xx}}}{s}$ ♦ they should comment on the evidence for the model being useful for prediction in their conclusion
	♦ testing the hypothesis that the population correlation coefficient is zero	Candidates should understand that: ♦ the ρ–test (for the Pearson product moment correlation coefficient being zero) makes the assumptions that the variables are independent, that they follow approximately a bivariate normal distribution, and they use the *t*-statistic $t = \frac{r\sqrt{(n-2)}}{\sqrt{1-r^2}}$ ♦ they should comment on the evidence for a linear association between the variables in their conclusion Candidates should appreciate that while tests on β and ρ are both two-tailed tests, they should be able to conjecture a one-tailed conclusion where the context allows. Learning and teaching contexts could involve candidates proving that the *t*-statistics for the slope parameter and the correlation coefficient are equivalent.

Hypothesis testing		
Skill	**Explanation**	**Suggested learning and teaching contexts**
	♦ communicating appropriate assumptions	

Appendix 3: question paper brief

The course assessment consists of two question papers, which assess the:

- knowledge and understanding of a range of complex statistical concepts
- ability to identify and use appropriate statistical models
- ability to apply more advanced operational skills in statistical contexts
- ability to use statistical reasoning skills to extract and interpret information, think logically and solve problems
- ability to communicate conclusions, exhibiting appreciation of their limitations
- ability to think analytically about the consequences of methodological choices

The question papers sample the ‘Skills, knowledge and understanding’ section of the course specification.

This sample draws on all of the skills, knowledge and understanding from each of the following areas:

- data handling and interpretation
- probability theory
- discrete random variables
- particular probability distributions
- sampling and the central limit theorem
- intervals and estimation
- bivariate analysis
- parametric tests
- non-parametric tests
- bivariate tests

Command words are the verbs or verbal phrases used in questions and tasks to ask candidates to demonstrate specific skills, knowledge or understanding. For examples of some of the command words used in this assessment please refer to the past papers and specimen question paper on SQA’s website.

The course assessment consists of two question papers:

	Paper 1	Paper 2
Time	60 minutes	2 hours and 45 minutes
Marks	30	90
Skills	This question paper gives candidates an opportunity to analyse a statistical report and to interpret summary statistics, including those generated by statistical software packages.	This question paper gives candidates an opportunity to apply statistical techniques and skills to: ♦ data handling and interpretation ♦ probability theory ♦ discrete random variables ♦ particular probability distributions ♦ sampling and the central limit theorem ♦ intervals and estimation ♦ bivariate analysis ♦ parametric tests ♦ non-parametric tests ♦ bivariate tests
Percentage of marks across the paper	Approximately 25–45% of the overall marks relate to data analysis and modelling. Approximately 25–45% of the overall marks relate to statistical inference. Approximately 25–45% of the overall marks relate to hypothesis testing.	
Type of question	The question papers contain short-answer questions, extended-response questions, and short case studies, set in contexts.	
Type of question paper	Semi-structured question papers: separate question papers and answer booklets. The answer booklets are structured with space for answers.	
Proportion of level 'C' questions	Some questions use a stepped approach to ensure that there are opportunities for candidates to demonstrate their abilities beyond level 'C'. Approximately 65% of marks are available for level 'C' responses.	
Balance of skills	Operational and reasoning skills are assessed in the question papers. Some questions assess only operational skills (approximately 65% of the marks), but other questions assess operational and reasoning skills (approximately 35% of the marks).	

Administrative information

Published: February 2023 (version 5.0)

History of changes

Version	Description of change	Date
2.0	Course support notes; skills, knowledge and understanding with suggested learning and teaching contexts; and question paper brief added as appendices.	May 2019
3.0	Appendix 2: skills, knowledge and understanding with suggested learning and teaching contexts, ‘Suggested learning and teaching contexts’ column: ♦ page 29 — wording about the probability of the Western Electric rules corrected ♦ page 36 — sample size in the example changed from ‘at least 20’ to ‘exceed 20’	August 2019
4.0	Skills, knowledge and understanding for the course assessment’ section, ‘Statistical inference’ table: page 6 — references to ‘parent population’ changed to ‘population’. Appendix 2: skills, knowledge and understanding with suggested learning and teaching contexts: ♦ ‘Statistical inference’ table: ▪ pages 24 and 26 — references to ‘parent population’ changed to ‘population’ ▪ page 24 — supplementary information added to the suggested learning and teaching contexts for the 'Identifying and using appropriate random sampling methods’ skill ▪ page 25 — equations and descriptions corrected in the suggested learning and teaching contexts for the ‘Working with the distribution of sample means and sample proportions’ skill ▪ pages 27 and 28 — equations corrected and note added to the suggested learning and teaching contexts for the ‘Obtaining confidence intervals’ skill	August 2022

<table>
<tr><th>Version</th><th>Description of change</th><th>Date</th></tr>
<tr><td></td><td>♦ ‘Hypothesis testing’ table:
▪ pages 36 and 37 — equations and descriptions corrected in the suggested learning and teaching contexts for the ‘Identifying and performing an appropriate two-sample test (independent or paired data) for comparing population means and proportions’ skill
▪ page 39 — supplementary information added to the suggested learning and teaching contexts for the ‘Identifying and performing an appropriate chi-squared test’ skill
▪ page 40 — supplementary information added to the suggested learning and teaching contexts for the ‘Identifying and performing an appropriate hypothesis test on bivariate data’ skill

Appendix 3: question paper brief: page 43 — information about the question papers amended to reflect the introduction of semi-structured answer booklets.</td><td></td></tr>
<tr><td>5.0</td><td>Skills, knowledge and understanding for the course assessment’ section (page 5) and Appendix 2: skills, knowledge and understanding with suggested learning and teaching contexts, (page 20), ‘Data analysis and modelling’ table:
$\mathrm{E}(aX \pm bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$ changed to
$\mathrm{E}(aX \pm bY) = a\mathrm{E}(X) \pm b\mathrm{E}(Y)$.</td><td>February 2023</td></tr>
</table>

Note: please check SQA’s website to ensure you are using the most up-to-date version of this document.