CS1 – Actuarial Statistics 1

Syllabus

for the 2023 exams

June 2022

CS1 – Actuarial Statistics

Aim

The aim of the Actuarial Statistics 1 subject is to provide a grounding in mathematical and statistical techniques that are of particular relevance to actuarial work.

Competences

On successful completion of this subject, a candidate will be able to:

1 describe the essential features of statistical distributions.

2 summarise data using appropriate statistical analysis, descriptive statistics and graphical presentation.

3 describe and apply the principles of statistical inference.

4 describe, apply and interpret the results of the linear regression model and generalised linear models.

5 explain the fundamental concepts of Bayesian statistics and use them to compute Bayesian estimators.

Links to other subjects

CS2 builds directly on the material in this subject.

CM1 and CM2 apply the material in this subject to actuarial and financial modelling.

This subject assumes that a candidate will be competent in the following elements of foundational mathematics and basic statistics:

1 Summarise the main features of a data set (exploratory data analysis)

1.1 Summarise a set of data using a table or frequency distribution, and display it graphically using a line plot, a box plot, a bar chart, histogram, stem and leaf plot or another appropriate elementary device.

1.2 Describe the level/location of a set of data using the mean, median, mode, as appropriate.

1.3 Describe the spread/variability of a set of data using the standard deviation, range, interquartile range, as appropriate.

1.4 Explain what is meant by symmetry and skewness for the distribution of a set of data.

2 Probability

2.1 Set functions and sample spaces for an experiment and an event.

2.2 Probability as a set function on a collection of events and its basic properties.

2.3 Calculate probabilities of events in simple situations.

2.4 Derive and use the addition rule for the probability of the union of two events.

2.5 Define and calculate the conditional probability of one event given the occurrence of another event.

2.6 Derive and use Bayes' theorem for events.

2.7 Define independence for two events, and calculate probabilities in situations involving independence.

3 Random variables

3.1 Explain what is meant by a discrete random variable, define the distribution function and the probability function of such a variable, and use these functions to calculate probabilities.

3.2 Explain what is meant by a continuous random variable, define the distribution function and the probability density function of such a variable, and use these functions to calculate probabilities.

3.3 Define the expected value of a function of a random variable, the mean, the variance, the standard deviation, the coefficient of skewness and the moments of a random variable, and calculate such quantities.

3.4 Evaluate probabilities associated with distributions (by calculation or by referring to tables as appropriate).

3.5 Derive the distribution of a function of a random variable from the distribution of the random variable.

Syllabus topics

1 Random variables and distributions (20%)

2 Data analysis (10%)

3 Statistical inference (25%)

4 Regression theory and applications (30%)

5 Bayesian statistics (15%)

These weightings are indicative of the approximate balance of the assessment of this subject between the main syllabus topics, averaged over a number of examination sessions.

The weightings also have a correspondence with the amount of learning material underlying each syllabus topic. However, this will also reflect aspects such as:

- the relative complexity of each topic and hence the amount of explanation and support required for it.

- the need to provide thorough foundation understanding on which to build the other objectives.

- the extent of prior knowledge that is expected.

- the degree to which each topic area is more knowledge- or application-based.

Skill levels

The use of a specific command verb within a syllabus objective does not indicate that this is the only form of question that can be asked on the topic covered by that objective. The Examiners may ask a question on any syllabus topic using any of the agreed command verbs, as are defined in the document 'Command verbs used in the Associate and Fellowship written examinations'.

Questions may be set at any skill level: Knowledge (demonstration of a detailed knowledge and understanding of the topic), Application (demonstration of an ability to apply the principles underlying the topic within a given context) and Higher Order (demonstration of an ability to perform deeper analysis and assessment of situations, including forming judgements, taking into account different points of view, comparing and contrasting situations, suggesting possible solutions and actions and making recommendations).

In the CS subjects, the approximate split of assessment across the three skill types is 20% Knowledge, 65% Application and 15% Higher Order skills.

Detailed syllabus objectives

1 Random variables and distributions (20%)

1.1 Define basic univariate distributions and use them to calculate probabilities, quantiles and moments.

1.1.1 Define and explain the key characteristics of the discrete distributions: geometric, binomial, negative binomial, hypergeometric, Poisson and uniform on a finite set.

1.1.2 Define and explain the key characteristics of the continuous distributions: normal, lognormal, exponential, gamma, chi-square, t, F, beta and uniform on an interval.

1.1.3 Evaluate probabilities and quantiles associated with distributions (by calculation or using statistical software as appropriate).

1.1.4 Define and explain the key characteristics of the Poisson process and explain the connection between the Poisson process and the Poisson distribution.

1.1.5 Generate basic discrete and continuous random variables using the inverse transform method.

1.1.6 Generate discrete and continuous random variables using statistical software.

1.2 Independence, joint and conditional distributions, linear combinations of random variables.

1.2.1 Explain what is meant by jointly distributed random variables, marginal distributions and conditional distributions.

1.2.2 Define the probability function/density function of a marginal distribution and of a conditional distribution.

1.2.3 Specify the conditions under which random variables are independent.

1.2.4 Define the expected value of a function of two jointly distributed random variables, the covariance and correlation coefficient between two variables, and calculate such quantities.

1.2.5 Define the probability function/density function of the sum of two independent random variables as the convolution of two functions.

1.2.6 Derive the mean and variance of linear combinations of random variables.

1.2.7 Use generating functions to establish the distribution of linear combinations of independent random variables.

1.3 Expectations, conditional expectations.

1.3.1 Define the conditional expectation of one random variable given the value of another random variable, and calculate such a quantity.

1.3.2 Show how the mean and variance of a random variable can be obtained from expected values of conditional expected values, and apply this.

1.4 Generating functions.

1.4.1 Define and determine the moment generating function of random variables.

1.4.2 Define and determine the cumulant generating function of random variables.

1.4.3 Use generating functions to determine the moments and cumulants of random variables, by expansion as a series or by differentiation, as appropriate.

1.4.4 Identify the applications for which a moment generating function, a cumulant generating function and cumulants are used and the reasons why they are used.

1.5 Central limit theorem – statement and application.

1.5.1 State the central limit theorem for a sequence of independent, identically distributed random variables.

1.5.2 Generate simulated samples from a given distribution and compare the sampling distribution with the Normal.

2 Data analysis (10%)

2.1 Data analysis.

2.1.1 Describe the possible aims of a data analysis (e.g. descriptive, inferential and predictive).

2.1.2 Describe the stages of conducting a data analysis to solve real-world problems in a scientific manner and describe tools suitable for each stage.

2.1.3 Describe sources of data and explain the characteristics of different data sources, including extremely large data sets.

2.1.4 Explain the meaning and value of reproducible research and describe the elements required to ensure a data analysis is reproducible.

2.2 Exploratory data analysis.

2.2.1 Describe the purpose of exploratory data analysis.

2.2.2 Use appropriate tools to calculate suitable summary statistics and undertake exploratory data visualisations.

2.2.3 Define and calculate Pearson's, Spearman's and Kendall's measures of correlation for bivariate data, explain their interpretation and perform statistical inference as appropriate.

2.2.4 Use principal components analysis to reduce the dimensionality of a complex data set.

2.3 Random sampling and sampling distributions.

2.3.1 Explain what is meant by a sample, a population and statistical inference.

2.3.2 Define a random sample from a distribution of a random variable.

2.3.3 Explain what is meant by a statistic and its sampling distribution.

2.3.4 Determine the mean and variance of a sample mean and the mean of a sample variance in terms of the population mean, variance and sample size.

2.3.5 State and use the basic sampling distributions for the sample mean and the sample variance for random samples from a normal distribution.

2.3.6 State and use the distribution of the t-statistic for random samples from a normal distribution.

2.3.7 State and use the F distribution for the ratio of two sample variances from independent samples taken from normal distributions.

3 Statistical inference (25%)

3.1 Estimation and estimators.

   3.1.1 Describe and apply the method of moments for constructing estimators of population parameters.

   3.1.2 Describe and apply the method of maximum likelihood for constructing estimators of population parameters.

   3.1.3 Define the following terms: efficiency, bias, consistency and mean square error.

   3.1.4 Define and apply the property of unbiasedness of an estimator.

   3.1.5 Define the mean square error of an estimator, and use it to compare estimators.

   3.1.6 Describe and apply the asymptotic distribution of maximum likelihood estimators.

   3.1.7 Use the bootstrap method to estimate properties of an estimator.

3.2 Confidence intervals and prediction intervals.

   3.2.1 Define in general terms a confidence interval for an unknown parameter of a distribution based on a random sample.

   3.2.2 Define in general terms a prediction interval for a future observation based on a model fitted to a random sample.

   3.2.3 Derive a confidence interval for an unknown parameter using a given sampling distribution.

   3.2.4 Calculate confidence intervals for the mean and the variance of a normal distribution.

   3.2.5 Calculate confidence intervals for a binomial probability and a Poisson mean, including the use of the normal approximation in both cases.

   3.2.6 Calculate confidence intervals for two-sample situations involving the normal distribution and the binomial and Poisson distributions using the normal approximation.

   3.2.7 Calculate confidence intervals for a difference between two means from paired data.

   3.2.8 Use the bootstrap method to obtain confidence intervals.

3.3 Hypothesis testing and goodness of fit.

   3.3.1 Explain what is meant by the following terms: null and alternative hypotheses, simple and composite hypotheses, type I and type II errors, sensitivity, specificity, test statistic, likelihood ratio, critical region, level of significance, probability value and power of a test.

   3.3.2 Apply basic tests for the one-sample and two-sample situations involving the normal, binomial and Poisson distributions, and apply basic tests for paired data.

   3.3.3 Apply the permutation approach to non-parametric hypothesis tests.

   3.3.4 Use a chi-square test to test the hypothesis that a random sample is from a particular distribution, including cases where parameters are unknown.

   3.3.5 Explain what is meant by a contingency (or two-way) table, and use a chi-square test to test the independence of two classification criteria.

4 Regression theory and applications (30%)

4.1 Linear regression.

   4.1.1 Explain what is meant by response and explanatory variables.

   4.1.2 State the simple regression model (with a single explanatory variable).

   4.1.3 Derive the least squares estimates of the slope and intercept parameters in a simple linear regression model.

   4.1.4 Use appropriate software to fit a simple linear regression model to a data set and interpret the output:

      • Perform statistical inference on the slope parameter.

      • Describe the use of measures of goodness of fit of a linear regression model.

      • Use a fitted linear relationship to predict a mean response or an individual response with confidence limits.

      • Use residuals to check the suitability and validity of a linear regression model.

   4.1.5 State the multiple linear regression model (with several explanatory variables).

   4.1.6 Use appropriate software to fit a multiple linear regression model to a data set and interpret the output.

   4.1.7 Use measures of model fit to select an appropriate set of explanatory variables.

4.2 Generalised linear models.

4.2.1 Define an exponential family of distributions. Show that the following distributions may be written in this form: binomial, Poisson, exponential, gamma, normal.

4.2.2 State the mean and variance for an exponential family, and define the variance function and the scale parameter. Derive these quantities for the distributions above.

4.2.3 Explain what is meant by the link function and the canonical link function, referring to the distributions above.

4.2.4 Explain what is meant by a variable, a factor taking categorical values and an interaction term. Define the linear predictor, illustrating its form for simple models, including polynomial models and models involving factors.

4.2.5 Define the deviance and scaled deviance and state how the parameters of a generalised linear model may be estimated. Describe how a suitable model may be chosen by using an analysis of deviance and by examining the significance of the parameters.

4.2.6 Define the Pearson and deviance residuals and describe how they may be used.

4.2.7 Apply statistical tests to determine the acceptability of a fitted model: Pearson's chi-square test and the likelihood-ratio test.

4.2.8 Fit a generalised linear model to a data set and interpret the output.

5 Bayesian statistics (15%)

5.1 Explain the fundamental concepts of Bayesian statistics and use these concepts to calculate Bayesian estimators.

5.1.1 Use Bayes' theorem to calculate simple conditional probabilities.

5.1.2 Explain what is meant by a prior distribution, a posterior distribution and a conjugate prior distribution.

5.1.3 Derive the posterior distribution for a parameter in simple cases.

5.1.4 Explain what is meant by a loss function.

5.1.5 Use simple loss functions to derive Bayesian estimates of parameters.

5.1.6 Derive credible intervals in simple cases.

5.1.7 Explain what is meant by the credibility premium formula and describe the role played by the credibility factor.

5.1.8 Explain the Bayesian approach to credibility theory and use it to derive credibility premiums in simple cases.

5.1.9 Explain the empirical Bayes approach to credibility theory and use it to derive credibility premiums in simple cases.

5.1.10 Explain the differences between the two approaches and state the assumptions underlying each of them.

Assessment

The assessment of this subject will consist of two examinations, CS1 Paper A (CS1A) and CS1 Paper B (CS1B).

CS1A will include a range of statistical questions, including numerical questions with varying marks. Expect questions to be set in line with the above syllabus topic weightings and skill levels. This examination will be three hours and twenty minutes, timed and online.

CS1B is a problem-based examination, focusing on computer-based data analysis and statistical modelling skills. Candidates are expected to include the R code that they have used to obtain the answers, together with the main R output produced, such as charts or tables. This examination is one hour and fifty minutes, timed and online.

The candidate must sit both CS1A and CS1B in the same sitting, and to pass, the weighted average of the candidate's final marks achieved in both papers must exceed the pass mark for the subject.

Please read A Guide to CS1 and CS2 Examinations, as well as the latest version of the IFoA Examinations Handbook and IFoA Examination Regulations on the IFoA website, before sitting any IFoA examination.

END