Institute of Actuaries of India

Subject
CS1 – Actuarial Statistics
Core Principles

For 2024 Examinations

CS1 - Actuarial Statistics

Syllabus for the 2024 Examinations

This syllabus includes information to support the study of this subject. It will guide you through what you need to learn, application of learning as well as the skills that you need to develop. Information regarding the assessment of this subject is also included.

This syllabus includes:

- Aim of the subject
- How this subject links across the Qualifications
- Subject topics and topic weightings
- Subject objectives
- Assessment information

Aim

Actuarial Statistics provides a grounding in mathematical and statistical techniques that are of particular relevance to actuarial work.

Links across the Qualifications

Associateship Qualification

The principles introduced in CS1 are further developed in CS2 Risk Modelling and Survival Analysis. CM1 Actuarial Mathematics and CM2 Financial Engineering and Loss Reserving apply the principles and techniques in CS1 tothe creation of actuarial and financial models

Topics and Topic Weightings

This subject covers the following topics:

1. Data Analysis [10%]
2. Random variables and distributions [20%]
3. Statistical inference [25%]
4. Regression Theory and applications [30%]
5. Bayesian statistics [15%]

Objectives

1 Data analysis [10%]

Production of simple visualisations and statistics from a data set.

1.1 Describe the purpose and function of data analysis

    1.1.1 Aims of a data analysis (e.g. descriptive, inferential and predictive)

    1.1.2 Stages and suitable tools used to conduct a data analysis to solve real-world problems

    1.1.3 Sources of data and their characteristics, including extremely large data sets

    1.1.4 Meaning and value of reproducible research and the elements required to ensure a data analysisis reproducible.

1.2 Complete exploratory data analysis

    1.2.1 Appropriate tools to calculate suitable summary statistics and undertake exploratory datavisualisations

    1.2.2 Interpret and make statistical inferences using Pearson's, Spearman's and Kendall's measuresof correlation for bivariate data

    1.2.3 Principal component analysis to reduce the dimensionality of a complex data set

2 Random variables and distributions [20%]

The basic properties and uses of commonly-used probability distributions and the statistical properties ofdata generated by randomly sampling from a known distribution.

2.1 Understand the characteristics of basic univariate distributions and how to generate samples from them

    2.1.1 Geometric, binomial, negative binomial, hypergeometric, Poisson and uniform discrete distributions on a finite set

    2.1.2 Normal, lognormal, exponential, gamma, chi-square, t, F, beta and uniform continuous distributions on an interval

    2.1.3 Evaluation of probabilities and quantiles associated with these distributions (by calculation orusing statistical software as appropriate)

    2.1.4 Poisson process and the connection between the Poisson process and the Poisson distribution

    2.1.5 Generation of basic discrete and continuous random variables using the inverse transformmethod

    2.1.6 Generation of discrete and continuous random variables using statistical software

2.2 Determine the characteristics of jointly distributed random variables

    2.2.1 Probability function or density function for marginal and conditional distributions of jointlydistributed random variables

    2.2.2 The conditions under which random variables are independent

    2.2.3 Covariance, the correlation and the expected value of a function of two jointly distributed randomvariables

    2.2.4 Mean and variance of linear combinations of random variables

2.3 Evaluate expectations and conditional expectations

    2.3.1 Conditional expectation of one random variable given the value of another random variable

    2.3.2 Mean and variance of a random variable as an expectation of conditional expected values

2.4 Evaluate and apply generating functions

    2.4.1 Moment and cumulant generating functions of a random variable

    2.4.2 Moment calculation via series expansion or differentiation of a generating function

2.5 State and apply the central limit theorem

    2.5.1 Central limit theorem for a sequence of independent, identically distributed random variables

    2.5.2 Comparison of simulated samples from a given distribution with the Normal distribution

2.6 Describe random sampling and the sampling distributions of statistics commonly used in statistical inference

2.6.1 Random samples from a population

2.6.2 The sampling distribution of a statistic

2.6.3 The mean and variance of a sample mean and the mean of a sample variance in terms of thepopulation mean, variance and sample size

2.6.4 Basic sampling distributions for the sample mean and variance for random samples from anormal distribution

2.6.5 The distribution of the t-statistic for random samples from a normal distribution

2.6.6 The F distribution for the ratio of two sample variances from independent samples taken fromnormal distributions

3 Statistical inference [25%]

Use of statistics to make inferences about the process underlying a data set.

3.1 Construct estimators and discuss their properties

3.1.1 Method of moments for constructing estimators of population parameters

3.1.2 Method of maximum likelihood for constructing estimators of population parameters

3.1.3 Efficiency, bias, consistency and mean square error of an estimator

3.1.4 Comparison of estimators using their mean square error and bias or unbiasedness

3.1.5 Asymptotic distribution of maximum likelihood estimators

3.1.6 Bootstrap method for estimating properties of an estimator

3.2 Calculate confidence intervals and prediction intervals

3.2.1 Confidence interval for an unknown parameter of a distribution based on a random sample

3.2.2 Prediction interval for a future observation based on a model fitted to a random sample

3.2.3 Confidence interval for an unknown parameter using a given sampling distribution

3.2.4 Confidence intervals for the mean and the variance of a normal distribution

3.2.5 Confidence intervals for a binomial probability and a Poisson mean, including the use of thenormal approximation in both cases

3.2.6 Confidence intervals for two-sample situations involving the normal distribution and thebinomial and Poisson distributions using the normal approximation

3.2.7 Confidence intervals for a difference between two means from paired data

3.2.8 Bootstrap method to obtain confidence intervals

3.3 Apply the concepts of hypothesis testing and goodness of fit

3.3.1 Understand the concepts of Null and alternative hypotheses, simple and composite hypotheses, type I and type II errors, sensitivity, specificity, test statistic, likelihood ratio, critical region, levelof significance, probability value and power of a test

3.3.2 Use of basic tests for the one-sample and two-sample situations involving the normal, binomialand Poisson distributions, and apply basic tests for paired data

3.3.3 The permutation approach to non-parametric hypothesis tests

3.3.4 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 A contingency (or two-way) table, and use of a chi-square test to test the independence of twoclassification criteria

4 Regression theory and applications [30%]

Use of statistics to examine and make inferences about the relationships between two or more data sets.

4.1 Understand and use linear regression models

 4.1.1 Response and explanatory variables

 4.1.2 Simple regression model (with a single explanatory variable) and multiple linear regression model (with several explanatory variables)

 4.1.3 Least squares estimates of the slope and intercept parameters in a simple linear regression model

 4.1.4 Use of appropriate software to fit a 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 Measures of model fit to select an appropriate set of explanatory variables

4.2 Understand and use generalised linear models

 4.2.1 Binomial, Poisson, exponential, gamma and normal distributions as an exponential family

 4.2.2 Mean, variance, variance function and scale parameter for a generalised linear model binomial, Poisson, exponential, gamma and normal distributions. Evaluate these quantities for the distributions in 4.2.1

 4.2.3 The link function and the canonical link function, referring to the distributions in 4.2.1

 4.2.4 Variables, factors taking categorical values and interaction terms.

 4.2.5 Definition of the linear predictor, including its form for simple models, including polynomial models and models involving factors

 4.2.6 Deviance scaled deviance and estimation of the parameters of a generalised linear model

 4.2.7 Choice of a suitable model using an analysis of deviance and examination of the significance of the parameters

 4.2.8 Pearson and deviance residuals and their use

 4.2.9 Statistical tests to determine the acceptability of a fitted model: Pearson's chi-square test and the likelihood-ratio test

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

5 Bayesian statistics [15%]

Use of Bayesian statistics to update prior beliefs about a data-generating process.

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

 5.1.1 Use of Bayes' theorem to calculate simple conditional probabilities

 5.1.2 Prior distribution, posterior distribution and conjugate prior distribution

 5.1.3 Posterior distribution for a parameter in simple cases

5.1.4 Use of simple loss functions to derive Bayesian estimates of parameters

5.1.5 Credible intervals in simple cases

5.1.6 Credibility premium formula and the role played by the credibility factor

5.1.7 Bayesian approach to credibility theory and its use for calculating credibility premiums in simplecases

5.1.8 Empirical Bayes approach to credibility theory and its use for deriving credibility premiums insimple cases

5.1.9 Understanding the differences between the two approaches (Bayes v Empirical Bayes) and theassumptions underlying each of them

Assessment

Assessment consists of a combination of a one-hour and forty-five-minutes computer-based data analysis and statistical modelling assignment and a three-hours and fifteen-minutes written examination

END