
Certificate in Actuarial Science

Mathematical Statistics

Mathematical Statistics is a branch of mathematics that deals with the collection, analysis, interpretation, presentation, and organization of data. It involves applying mathematical techniques to analyze and interpret data to make decisions, predictions, and conclusions. In the course Certificate in Actuarial Science, a solid understanding of mathematical statistics is essential for actuaries to assess risk, make projections, and inform financial decisions.

Key Terms and Vocabulary:

1. **Data:** Data refers to information collected or observed from the real world. It can be quantitative (numerical) or qualitative (categorical).
2. **Population:** The population is the complete set of individuals, objects, or events being studied. It is the entire group that a researcher wants to draw conclusions about.
3. **Sample:** A sample is a subset of the population that is selected for study. It is used to make inferences or generalizations about the population.
4. **Descriptive Statistics:** Descriptive statistics are used to describe and summarize data. Measures such as mean, median, mode, variance, and standard deviation fall under descriptive statistics.
5. **Inferential Statistics:** Inferential statistics are used to make predictions or inferences about a population based on sample data. It involves hypothesis testing, confidence intervals, and regression analysis.
6. **Variable:** A variable is a characteristic or attribute that can take on different values. It can be independent (predictor) or dependent (response) in statistical analysis.
7. **Random Variable:** A random variable is a variable whose possible values are outcomes of a random phenomenon. It can be discrete or continuous.
8. **Probability:** Probability is the measure of the likelihood that an event will occur. It ranges from 0 (impossible) to 1 (certain).
9. **Probability Distribution:** A probability distribution is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment.
10. **Normal Distribution:** The normal distribution is a symmetric, bell-shaped distribution that is commonly used in statistics. It is characterized by its mean and standard deviation.
11. **Central Limit Theorem:** The central limit theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the shape of the population distribution.

-
12. Hypothesis Testing: Hypothesis testing is a statistical method used to make inferences about a population based on sample data. It involves formulating null and alternative hypotheses and conducting tests to determine the validity of the null hypothesis.
13. Confidence Interval: A confidence interval is a range of values within which the true population parameter is estimated to lie with a certain level of confidence.
14. Regression Analysis: Regression analysis is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. It is used for prediction and inference.
15. Bayesian Statistics: Bayesian statistics is an approach to statistical inference that uses Bayes' theorem to update the probability of a hypothesis as new evidence becomes available.
16. ANOVA (Analysis of Variance): ANOVA is a statistical technique used to test for differences in means among two or more groups. It assesses whether there are statistically significant differences between group means.
17. Chi-Square Test: The chi-square test is a statistical test used to determine whether there is a significant association between two categorical variables.
18. Probability Density Function (PDF): The probability density function is a function that describes the likelihood of a continuous random variable falling within a particular range.
19. Cumulative Distribution Function (CDF): The cumulative distribution function is a function that gives the probability that a random variable takes on a value less than or equal to a given value.
20. Sampling Distribution: The sampling distribution is the probability distribution of a sample statistic based on all possible samples of a certain size from a population.
21. Standard Error: The standard error is a measure of the variability of a sample statistic. It is used to quantify the uncertainty in estimating a population parameter.
22. Correlation: Correlation measures the strength and direction of the relationship between two variables. It ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation).
23. Regression Coefficient: The regression coefficient is a measure of the change in the dependent variable for a one-unit change in the independent variable in regression analysis.
24. Outlier: An outlier is an observation that lies significantly distant from other observations in a dataset. It can affect the results of statistical analysis.
25. Type I Error: A Type I error occurs when a true null hypothesis is rejected. It is also known as a false positive.
26. Type II Error: A Type II error occurs when a false null hypothesis is not rejected. It is also known as a false negative.
-

-
27. Power: Power is the probability of correctly rejecting a false null hypothesis. It is influenced by sample size, effect size, and significance level.
28. Confounding Variable: A confounding variable is a variable that influences both the dependent and independent variables, leading to erroneous conclusions.
29. Statistical Significance: Statistical significance indicates that the results of a study are unlikely to have occurred by chance. It is typically assessed using p-values.
30. Skewness: Skewness measures the asymmetry of the distribution of data. Positive skewness indicates a longer right tail, while negative skewness indicates a longer left tail.
31. Kurtosis: Kurtosis measures the peakedness of the distribution of data. High kurtosis indicates a sharp peak, while low kurtosis indicates a flat peak.
32. Residual: A residual is the difference between the observed value and the predicted value in regression analysis. Residual analysis is used to assess the goodness of fit of a model.
33. Degrees of Freedom: Degrees of freedom represent the number of independent observations in a sample. It is used in statistical tests to determine the variability of estimates.
34. Multicollinearity: Multicollinearity occurs when independent variables in regression analysis are highly correlated with each other. It can lead to unstable estimates and unreliable results.
35. Interquartile Range (IQR): The interquartile range is the range of values between the first and third quartiles of a dataset. It is a measure of variability that is resistant to outliers.
36. Normality Test: Normality tests are used to determine whether a dataset follows a normal distribution. Popular tests include the Shapiro-Wilk test and the Kolmogorov-Smirnov test.
37. Time Series Analysis: Time series analysis is a statistical method used to analyze data collected over time. It involves identifying patterns, trends, and seasonality in time series data.
38. Survival Analysis: Survival analysis is a statistical technique used to analyze time-to-event data. It is commonly used in medical research, engineering, and actuarial science.
39. Covariance: Covariance measures the relationship between two random variables. Positive covariance indicates a direct relationship, while negative covariance indicates an inverse relationship.
40. Autocorrelation: Autocorrelation measures the correlation of a variable with itself over different time lags in time series data. It is important for identifying patterns and trends.
41. Moment: Moments are statistical measures that describe the shape, center, and spread of a distribution. The first moment is the mean, the second moment is the variance, and so on.
42. Sampling Bias: Sampling bias occurs when the sample selected is not representative of the population, leading to erroneous conclusions. It can result from non-random sampling methods.
-

-
43. Bayesian Inference: Bayesian inference is a method of statistical inference that uses Bayes' theorem to update the probability of a hypothesis based on prior knowledge and new evidence.
44. Estimation: Estimation is the process of estimating unknown population parameters based on sample data. Point estimation provides a single value estimate, while interval estimation provides a range of values.
45. Statistical Model: A statistical model is a mathematical representation of a real-world process or phenomenon. It describes the relationship between variables and is used for prediction and inference.
46. Overfitting: Overfitting occurs when a statistical model is overly complex and fits the training data too closely, leading to poor generalization to new data.
47. Underfitting: Underfitting occurs when a statistical model is too simple and fails to capture the underlying patterns in the data, leading to poor predictive performance.
48. Random Sampling: Random sampling is a sampling method in which every member of the population has an equal chance of being selected for the sample. It ensures the sample is representative of the population.
49. Statistical Test: A statistical test is a method used to make decisions about a population based on sample data. Common tests include t-tests, chi-square tests, and ANOVA.
50. Statistical Inference: Statistical inference is the process of drawing conclusions about a population based on sample data. It involves estimation, hypothesis testing, and prediction.
51. Confidence Level: The confidence level is the probability that a confidence interval contains the true population parameter. Common levels include 90%, 95%, and 99%.
52. P-value: The p-value is the probability of observing a test statistic as extreme as, or more extreme than, the actual observed statistic, assuming the null hypothesis is true. It is used to determine statistical significance.
53. Statistical Power: Statistical power is the probability of correctly rejecting a false null hypothesis. It is influenced by sample size, effect size, and significance level.
54. Statistical Software: Statistical software is computer software used for statistical analysis. Popular tools include R, Python, SAS, SPSS, and Excel.
55. Statistical Programming: Statistical programming involves writing code to perform statistical analysis, data visualization, and modeling. It is essential for data analysis and research.
56. Longitudinal Data: Longitudinal data is data collected over time from the same individuals or subjects. It is used in longitudinal studies to analyze changes and trends over time.
57. Factor Analysis: Factor analysis is a statistical method used to identify underlying factors or latent variables that explain the correlations among observed variables.
-

-
58. **Principal Component Analysis (PCA):** PCA is a dimensionality reduction technique that converts correlated variables into linearly uncorrelated variables called principal components. It is used for data compression and visualization.
59. **Bootstrapping:** Bootstrapping is a resampling technique used to estimate the sampling distribution of a statistic by repeatedly sampling with replacement from the original sample.
60. **Monte Carlo Simulation:** Monte Carlo simulation is a computational technique that uses random sampling to estimate the distribution of an unknown quantity. It is used for risk analysis and decision-making.
61. **Statistical Paradox:** A statistical paradox is a counterintuitive result that arises from the misinterpretation or misuse of statistical methods. Examples include Simpson's paradox and the birthday paradox.
62. **Statistical Learning:** Statistical learning is a field that combines statistics and machine learning to develop algorithms that can learn from data and make predictions and decisions.
63. **Survival Function:** The survival function is a probability function that gives the probability of survival beyond a certain time point in survival analysis.
64. **Actuarial Science:** Actuarial science is the discipline that applies mathematical and statistical methods to assess risk in insurance, finance, and other industries. Actuaries use probability theory and statistics to analyze and manage risk.
65. **Loss Function:** A loss function is a function that quantifies the cost or penalty associated with the difference between predicted and actual values in statistical modeling and machine learning.
66. **Markov Chain:** A Markov chain is a stochastic process that evolves over time with the property that the future state depends only on the current state, not on the past states.
67. **Time Series Forecasting:** Time series forecasting is the process of predicting future values of a time series based on historical data. It is used in finance, economics, meteorology, and other fields.
68. **Actuarial Risk:** Actuarial risk refers to the uncertainty associated with predicting future events and outcomes, such as insurance claims, mortality rates, and investment returns.
69. **Loss Reserving:** Loss reserving is the process of estimating the future liabilities of an insurance company for claims that have been reported but not yet settled.
70. **Claim Frequency:** Claim frequency is the number of insurance claims reported over a specific period, such as a year. It is used to assess the risk and profitability of insurance policies.
71. **Claim Severity:** Claim severity is the amount of money paid out for each insurance claim. It is used to estimate the expected cost of claims and set insurance premiums.
72. **Frequency-Severity Model:** The frequency-severity model is a common actuarial model that combines claim frequency and claim severity to estimate the total expected claims for an insurance portfolio.
-

-
73. Risk Premium: The risk premium is the additional amount charged by an insurer to cover the expected losses and expenses associated with an insurance policy.
74. Loss Ratio: The loss ratio is the ratio of incurred losses and expenses to earned premiums. It is used to measure the profitability of an insurance company.
75. Actuarial Valuation: Actuarial valuation is the process of estimating the present value of future cash flows and liabilities for pension plans, insurance companies, and other financial institutions.
76. Stochastic Model: A stochastic model is a mathematical model that incorporates random variables to account for uncertainty in predictions and simulations.
77. Actuarial Assumptions: Actuarial assumptions are the key assumptions and parameters used in actuarial calculations, such as mortality rates, interest rates, and inflation rates.
78. Credibility Theory: Credibility theory is a branch of actuarial science that focuses on estimating risks for small or insufficient data sets. It uses historical data and industry experience to make predictions.
79. Loss Distribution: The loss distribution is the probability distribution of potential losses that an insurer may experience due to claims, catastrophes, or other events.
80. Risk Management: Risk management is the process of identifying, assessing, and mitigating risks to minimize the impact of uncertain events on an organization's objectives.
81. Actuarial Tables: Actuarial tables are statistical tables that provide mortality rates, life expectancies, and other demographic information used in actuarial calculations.
82. Survival Probability: Survival probability is the likelihood that an individual will survive beyond a certain age or time period, as estimated by actuarial tables.
83. Capital Adequacy: Capital adequacy refers to the sufficiency of an insurer's capital reserves to cover potential losses and meet regulatory requirements.
84. Risk Assessment: Risk assessment is the process of evaluating the likelihood and impact of risks to determine the best course of action to manage them effectively.
85. Actuarial Report: An actuarial report is a formal document prepared by an actuary that presents the findings, assumptions, and recommendations related to a specific actuarial analysis.
86. Actuarial Science Society: The Actuarial Science Society is a professional organization that promotes education, research, and networking opportunities for actuaries and actuarial students.
87. Actuarial Exam: Actuarial exams are rigorous examinations that test the knowledge and skills of aspiring actuaries in various areas of actuarial science, such as probability, statistics, and finance.
88. Actuarial Internship: An actuarial internship is a temporary work experience program that provides students with hands-on training and exposure to the actuarial profession.
-

-
89. Actuarial Modeling: Actuarial modeling involves developing mathematical and statistical models to analyze and predict future events and outcomes in insurance, finance, and other industries.
90. Actuarial Pricing: Actuarial pricing is the process of determining the appropriate premiums for insurance policies based on the expected claims, expenses, and profitability.
91. Actuarial Science Ethics: Actuarial science ethics are the professional standards and guidelines that actuaries must adhere to in their practice to ensure integrity, honesty, and fairness.
92. Actuarial Software: Actuarial software is specialized software used by actuaries to perform complex calculations, modeling, and analysis in actuarial science.
93. Actuarial Tables: Actuarial tables are statistical tables that provide mortality rates, life expectancies, and other demographic information used in actuarial calculations.
94. Actuarial Valuation: Actuarial valuation is the process of estimating the present value of future cash flows and liabilities for pension plans, insurance companies, and other financial institutions.
95. Bayesian Statistics: Bayesian statistics is an approach to statistical inference that uses Bayes' theorem to update the probability of a hypothesis as new evidence becomes available.
96. Bayesian Inference: Bayesian inference is a method of statistical inference that uses Bayes' theorem to update the probability of a hypothesis based on prior knowledge and new evidence.
97. Central Limit Theorem: The central limit theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the shape of the population distribution.
98. Chi-Square Test: The chi-square test is a statistical test used to determine whether there is a significant association between two categorical variables.
99. Confidence Interval: A confidence interval is a range of values within which the true population parameter