**Roll No. …….....** **Total Pages: _ _**

**Paper ID: AI003** **Course Code: CCAD-1-03T**

**Examination (January- 2024)**
**Certificate/ Diploma (Semester-I)**
**Artificial Intelligence and Data Science**

**Statistical Foundation**

**Time Allowed: 2 Hours** **Max. Marks: 70**

**Instructions for the Students**

1. The question paper shall consist of 70 Multiple Choice questions.
2. All questions are compulsory. Each question carries 1 mark.
3. There will be no negative marking.

| | |
|---|---|
| Q1. Which type of data is directly collected from original sources for the first time?<br>a) Secondary Data<br>b) Qualitative Data<br>c) Primary Data<br>d) Nominal Data | Q2. Which approach to probability involves the concept of equally likely outcomes?<br>a) Classical approach<br>b) Empirical approach<br>c) Axiomatic approach<br>d) Theoretical approach |
| Q3. What is the primary purpose of constructing frequency distributions for discrete and continuous data?<br>a) Aesthetic presentation<br>b) Simplifying data entry<br>c) Summarizing and organizing data<br>d) Enhancing data security | Q4. Which of the following is a property of probability?<br>a) Probability is always greater than 1<br>b) Probability is always between 0 and 1<br>c) Probability is a negative value<br>d) Probability is unrelated to events |
| Q5. When is a bar diagram more appropriate than a histogram?<br>a) When data is continuous<br>b) When data is categorical<br>c) When data is highly skewed<br>d) When data is normally distributed | Q6. In the context of data, what does "quantitative" refer to?<br>a) Data with numerical values<br>b) Data with descriptive characteristics<br>c) Data with categories<br>d) Data with subjective opinions |
| Q7. What does conditional probability refer to?<br>a) Probability of an event occurring<br>b) Probability of two independent events occurring<br>c) Probability of an event given that another event has occurred<br>d) Probability of mutually exclusive events | Q8. What is regression analysis primarily used for?<br>a) Descriptive statistics<br>b) Testing hypotheses<br>c) Predicting the value of one variable based on another<br>d) Calculating probabilities |
| Q9. What is the formula for the slope (b) in simple linear regression?<br>a) $(\Sigma xy - \overline{nxy}) / (\Sigma x^2 - \overline{nx^2})$<br>b) $(\Sigma xy - \Sigma x\Sigma y/n) / (\Sigma x^2 - (\Sigma x)^2/n)$<br>c) $(\Sigma xy - \Sigma x\Sigma y) / (\Sigma x^2 - \Sigma x)$<br>d) $(\Sigma xy - \Sigma x\Sigma y) / (\Sigma x^2 - (\Sigma x)^2)$ | Q10. What is the primary emphasis of ordinal data?<br>a) Exact Measurement<br>b) Categories or Labels<br>c) Order or Ranking<br>d) Countable Values |
| Q11. What is the primary purpose of Descriptive | Q12. If two events cannot occur simultaneously, they |

| | |
|---|---|
| Statistics?<br>a) To make predictions about future events<br>b) To summarize and describe the main features of a dataset<br>c) To test hypotheses about population parameters<br>d) To identify patterns and relationships in data | are known as:<br>a) Independent events<br>b) Dependent events<br>c) Mutually exclusive events<br>d) Compound events |
| Q13. What is the primary difference between correlation and regression?<br>a) Correlation measures the strength and direction of a linear relationship, while regression predicts the value of one variable based on another.<br>b) Regression measures the strength and direction of a linear relationship, while correlation predicts the value of one variable based on another.<br>c) Correlation and regression are identical concepts.<br>d) Correlation is used for categorical data, while regression is used for continuous data. | Q14. Which graphical representation is suitable for displaying the distribution of categorical data with non-overlapping bars?<br>a) Histogram<br>b) Frequency Polygon<br>c) Pie Chart<br>d) Ogive Curve |
| Q15. Which theorem relates the joint probability of two events to their individual probabilities?<br>a) Bayes' Theorem<br>b) Multiplication Rule<br>c) Addition Rule<br>d) Complement Rule | Q16. What does the range measure in a set of data?<br>a) Spread of data<br>b) Central tendency<br>c) Dispersion<br>d) Skewness |
| Q17. What is the main concern of the Scope of Statistics?<br>a) Collection of Data<br>b) Interpretation of Data<br>c) Application of Data<br>d) All of the above | Q18. What is the sum of probabilities of all elementary events in a sample space?<br>a) 1<br>b) 0<br>c) 2<br>d) Depends on the experiment |
| Q19. Probability based on observed data and frequency is associated with which approach?<br>a) Classical approach<br>b) Empirical approach<br>c) Axiomatic approach<br>d) Theoretical approach | Q20. Which correlation coefficient is appropriate when dealing with ordinal data?<br>a) Point-biserial correlation coefficient<br>b) Kendall's tau<br>c) Karl Pearson coefficient of correlation<br>d) Phi coefficient |
| Q21. Which of the following is a measure of dispersion?<br>a) Mean<br>b) Median<br>c) Variance<br>d) Skewness | Q22. What is a property of probability that states the probability of the entire sample space is 1?<br>a) Complementarity<br>b) Multiplication Rule<br>c) Addition Rule<br>d) Total Probability Rule |
| Q23. Which theorem provides a way to update the probability of an event based on new evidence?<br>a) Conditional Probability Theorem<br>b) Bayes' Theorem<br>c) Total Probability Theorem<br>d) Independence Theorem | Q24. What does a frequency polygon visually represent in a dataset?<br>a) Individual data points<br>b) Cumulative frequencies<br>c) Relative frequencies<br>d) Mean and standard deviation |
| Q25. In a frequency distribution, what does the cumulative frequency represent?<br>a) Total number of data points | Q26. If events A and B are independent, what is P(A and B)?<br>a) $P(A) \times P(B)$ |

| | |
|---|---|
| b) Frequency of each data point<br>c) Cumulative sum of frequencies up to a certain point<br>d) Average frequency of the data set | b) P(A) + P(B)<br>c) P(A) - P(B)<br>d) P(A) / P(B) |
| Q27. What is the purpose of the intercept (a) in simple linear regression?<br>a) To shift the regression line horizontally<br>b) To determine the slope of the line<br>c) To shift the regression line vertically<br>d) To calculate the mean of the dependent variable | Q28. The graphical representation of data using bars is known as:<br>a) Histogram<br>b) Pie chart<br>c) Line graph<br>d) Scatter plot |
| Q29. In the context of data, what does "discrete" refer to?<br>a) Data that can take any real value within a range<br>b) Data that can only take distinct, separate values<br>c) Data collected from primary sources<br>d) Data that is continuously changing | Q30. If the probability of event A is 0.6 and the probability of event B is 0.4, what is the probability of either A or B occurring?<br>a) 0.6 + 0.4<br>b) 0.6 * 0.4<br>c) 0.6 - 0.4<br>d) 0.6 / 0.4 |
| Q31. In a random experiment, if every outcome is equally likely, what type of probability distribution is it?<br>a) Normal distribution<br>b) Uniform distribution<br>c) Binomial distribution<br>d) Exponential distribution | Q32. Which measure of skewness indicates a longer tail on the right side of the distribution?<br>a) Positive skewness<br>b) Negative skewness<br>c) Zero skewness<br>d) Mode skewness |
| Q33. What is Exploratory Data Analysis (EDA)?<br>a) A method to make predictions with high accuracy<br>b) An approach to summarize data using inferential statistics<br>c) An initial investigation of data to discover patterns and insights<br>d) A technique to test hypotheses about population parameters | Q34. Probability Mass Function (PMF) is associated with:<br>a) Continuous random variables<br>b) Discrete random variables<br>c) Both A and B<br>d) Neither A nor B |
| Q35. What does the Density Function represent in probability theory?<br>a) The probability of a discrete event<br>b) The probability of a continuous event<br>c) The sum of probabilities in a distribution<br>d) The average of a set of values | Q36. In which type of data measurement do the categories have a meaningful order?<br>a) Nominal<br>b) Ordinal<br>c) Quantitative<br>d) Qualitative |
| Q37. What does the Coefficient of Variation (CV) express?<br>a) The spread of data around the mean<br>b) The percentage difference between the mean and median<br>c) The proportion of data within a certain range<br>d) The relative variability of data relative to the mean | Q38. Mathematical Expectation is also known as:<br>a) Probability Expectation<br>b) Expectation Value<br>c) Mean<br>d) Median |
| Q39. What is the Moment Generating Function used for?<br>a) To calculate the mean of a distribution<br>b) To calculate the variance of a distribution<br>c) To calculate moments of a distribution | Q40. Which graphical representation is effective for comparing the proportions of different categories within a whole?<br>a) Histogram<br>b) Frequency Polygon |

| | |
|---|---|
| d) To calculate the median of a distribution | c) Pie Chart<br>d) Ogive Curve |
| Q41. Skewness in a dataset refers to:<br>a) The symmetry of the distribution<br>b) The concentration of values around the mean<br>c) The presence of outliers<br>d) The shape of the tails of the distribution | Q42. What is a subset of the sample space?<br>a) Event<br>b) Elementary event<br>c) Random experiment<br>d) Sample point |
| Q43. Which of the following statements about Moments is correct?<br>a) First moment is the mean, second moment is the variance<br>b) First moment is the median, second moment is the mode<br>c) First moment is the mode, second moment is the mean<br>d) First moment is the variance, second moment is the mean | Q44. What is the measure of central tendency that is not affected by extreme values?<br>a) a. Mean<br>b) b. Median<br>c) c. Mode<br>d) d. Standard Deviation |
| Q45. What type of data is suitable for a histogram?<br>a) Categorical data<br>b) Nominal data<br>c) Discrete and continuous data<br>d) Ordinal data | Q46. Which of the following is a property of the Characteristic Function?<br>a) It is always positive<br>b) It is always negative<br>c) It is always real<br>d) It is always complex |
| Q47. What are the properties of Mathematical Expectation?<br>a) Linearity, Independence, and Non-negativity<br>b) Additivity, Positivity, and Subtraction<br>c) Homogeneity, Multiplicity, and Positivity<br>d) Linearity, Additivity, and Homogeneity | Q48. Which of the following is a visual representation of the cumulative frequencies in a dataset?<br>a) Bar Diagram<br>b) Frequency Polygon<br>c) Ogive Curve<br>d) Histogram |
| Q49. Which correlation coefficient is suitable for measuring the strength and direction of a linear relationship between two continuous variables?<br>a) Kendall's tau<br>b) Spearman's rank correlation coefficient<br>c) Karl Pearson coefficient of correlation<br>d) Point-biserial correlation coefficient | Q50. Moments of higher order provide information about the:<br>a) Spread of the distribution<br>b) Skewness of the distribution<br>c) Shape of the distribution<br>d) Central tendency of the distribution |
| Q51. Which term represents a single outcome of a random experiment?<br>a) Sample space<br>b) Event<br>c) Sample point<br>d) Compound event | Q52. What is the range of the correlation coefficient?<br>a) -1 to 1<br>b) 0 to 1<br>c) 1 to 100<br>d) -100 to 100 |
| Q53. Continuous data is best described as:<br>a) Data that can only take distinct values<br>b) Data that is measured in categories<br>c) Data that can take any real value within a range<br>d) Data that is collected through surveys | Q54. What is a measure of central tendency that is most affected by outliers?<br>a) Mean<br>b) Median<br>c) Mode<br>d) Range |
| Q55. What is Data Visualization?<br>a) The process of collecting data<br>b) The representation of data through visual elements<br>c) The analysis of data using statistical techniques | Q56. What is the primary purpose of Statistical Inference?<br>a) To summarize data<br>b) To make predictions about a population based on a sample |

| | |
|---|---|
| d) The process of cleaning and preparing data | c) To describe the characteristics of a population<br>d) To measure the central tendency of a distribution |
| Q57. What type of correlation coefficient is suitable for ranked data?<br>a) Karl Pearson coefficient<br>b) Spearman's rank correlation coefficient<br>c) Multiple correlation coefficient<br>d) D) Partial correlation coefficient | Q58. What is the primary purpose of scatter plots in statistical analysis?<br>a) To show the relationship between two variables graphically<br>b) To calculate correlation coefficients<br>c) To perform regression analysis<br>d) To display the mean and standard deviation |
| Q59. What is the purpose of an Ogive Curve in data representation?<br>a) Displaying individual data points<br>b) Showing cumulative frequencies<br>c) Highlighting outliers<br>d) Comparing multiple datasets | Q60. What is the collection of all possible outcomes of a random experiment?<br>a) Sample point<br>b) Event<br>c) Sample space<br>d) Elementary event |
| Q61. Which type of probability is based on observed frequencies or past data?<br>a) Classical probability<br>b) Empirical probability<br>c) Axiomatic probability<br>d) Mathematical probability | Q62. Which term is used to describe the measure of the spread of a dataset?<br>a) Central tendency<br>b) Variability<br>c) Frequency<br>d) Correlation |
| Q63. Which type of graph is suitable for representing the relationship between two continuous variables?<br>a) Bar chart<br>b) Pie chart<br>c) Scatter diagram<br>d) Histogram | Q64. In statistical inference, what is the concept related to a numerical summary of a random variable?<br>a) Probability mass function<br>b) Mathematical expectation<br>c) Moments<br>d) Characteristic function |
| Q65. What is the formula for calculating the coefficient of variation?<br>a) (Mean - Median) / Standard Deviation<br>b) Standard Deviation / Mean<br>c) Range / Mean<br>d) (Mean - Mode) / Standard Deviation | Q66. Which graphical representation is effective for showing the trend in a dataset over time?<br>a) Histogram<br>b) Frequency Polygon<br>c) Bar Diagram<br>d) Ogive Curve |
| Q67. What is the purpose of a Grouped Data representation?<br>a) To simplify large datasets<br>b) To complicate data analysis<br>c) To increase data variability<br>d) To reduce the need for data visualization | Q68. Which term is used to describe the spread or dispersion of a set of values?<br>a) Skewness<br>b) Range<br>c) Kurtosis<br>d) Median |
| Q69. In a Scatter diagram, what does a strong positive correlation between two variables indicate?<br>a) An inverse relationship<br>b) No relationship<br>c) A direct relationship<br>d) A random relationship | Q70. The coefficient of variation is a measure of:<br>a) Central tendency<br>b) Dispersion<br>c) Skewness<br>d) Kurtosis |