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

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

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

d) To calculate the median of a distribution	c) Pie Chart d) Ogive Curve
Q41. Skewness in a dataset refers to: a) The symmetry of the distribution b) The concentration of values around the mean c) The presence of outliers d) The shape of the tails of the distribution	Q42. What is a subset of the sample space? a) Event b) Elementary event c) Random experiment d) Sample point
Q43. Which of the following statements about Moments is correct? a) First moment is the mean, second moment is the variance b) First moment is the median, second moment is the mode c) First moment is the mode, second moment is the mean 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? a) a. Mean b) b. Median c) c. Mode d) d. Standard Deviation
Q45. What type of data is suitable for a histogram? a) Categorical data b) Nominal data c) Discrete and continuous data d) Ordinal data	Q46. Which of the following is a property of the Characteristic Function? a) It is always positive b) It is always negative c) It is always real d) It is always complex
Q47. What are the properties of Mathematical Expectation? a) Linearity, Independence, and Non-negativity b) Additivity, Positivity, and Subtraction c) Homogeneity, Multiplicity, and Positivity d) Linearity, Additivity, and Homogeneity	Q48. Which of the following is a visual representation of the cumulative frequencies in a dataset? a) Bar Diagram b) Frequency Polygon c) Ogive Curve d) Histogram
Q49. Which correlation coefficient is suitable for measuring the strength and direction of a linear relationship between two continuous variables? a) Kendall's tau b) Spearman's rank correlation coefficient c) Karl Pearson coefficient of correlation d) Point-biserial correlation coefficient	Q50. Moments of higher order provide information about the: a) Spread of the distribution b) Skewness of the distribution c) Shape of the distribution d) Central tendency of the distribution
Q51. Which term represents a single outcome of a random experiment? a) Sample space b) Event c) Sample point d) Compound event	Q52. What is the range of the correlation coefficient? a) -1 to 1 b) 0 to 1 c) 1 to 100 d) -100 to 100
Q53. Continuous data is best described as: a) Data that can only take distinct values b) Data that is measured in categories c) Data that can take any real value within a range d) Data that is collected through surveys	Q54. What is a measure of central tendency that is most affected by outliers? a) Mean b) Median c) Mode d) Range
Q55. What is Data Visualization? a) The process of collecting data b) The representation of data through visual elements c) The analysis of data using statistical techniques	Q56. What is the primary purpose of Statistical Inference? a) To summarize data 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 d) To measure the central tendency of a distribution
Q57. What type of correlation coefficient is suitable for ranked data? a) Karl Pearson coefficient b) Spearman's rank correlation coefficient c) Multiple correlation coefficient d) D) Partial correlation coefficient	Q58. What is the primary purpose of scatter plots in statistical analysis? a) To show the relationship between two variables graphically b) To calculate correlation coefficients c) To perform regression analysis d) To display the mean and standard deviation
Q59. What is the purpose of an Ogive Curve in data representation? a) Displaying individual data points b) Showing cumulative frequencies c) Highlighting outliers d) Comparing multiple datasets	Q60. What is the collection of all possible outcomes of a random experiment? a) Sample point b) Event c) Sample space d) Elementary event
Q61. Which type of probability is based on observed frequencies or past data? a) Classical probability b) Empirical probability c) Axiomatic probability d) Mathematical probability	Q62. Which term is used to describe the measure of the spread of a dataset? a) Central tendency b) Variability c) Frequency d) Correlation
Q63. Which type of graph is suitable for representing the relationship between two continuous variables? a) Bar chart b) Pie chart c) Scatter diagram d) Histogram	Q64. In statistical inference, what is the concept related to a numerical summary of a random variable? a) Probability mass function b) Mathematical expectation c) Moments d) Characteristic function
Q65. What is the formula for calculating the coefficient of variation? a) (Mean - Median) / Standard Deviation b) Standard Deviation / Mean c) Range / Mean d) (Mean - Mode) / Standard Deviation	Q66. Which graphical representation is effective for showing the trend in a dataset over time? a) Histogram b) Frequency Polygon c) Bar Diagram d) Ogive Curve
Q67. What is the purpose of a Grouped Data representation? a) To simplify large datasets b) To complicate data analysis c) To increase data variability d) To reduce the need for data visualization	Q68. Which term is used to describe the spread or dispersion of a set of values? a) Skewness b) Range c) Kurtosis d) Median
Q69. In a Scatter diagram, what does a strong positive correlation between two variables indicate? a) An inverse relationship b) No relationship c) A direct relationship d) A random relationship	Q70. The coefficient of variation is a measure of: a) Central tendency b) Dispersion c) Skewness d) Kurtosis