17849

# BRAINWARE UNIVERSITY



**Term End Examination 2024-2025**
**Programme – BBA(BA)-Hons-2024**
**Course Name – Foundations of Data Science**
**Course Code - BBB20002**
**( Semester II )**

**Full Marks : 60** **Time : 2:30 Hours**

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

## Group-A
(Multiple Choice Type Question) 1 x 15=15

1. *Choose the correct alternative from the following :*

(i) Select the primary goal of Data Science.
a) To store large amounts of data
b) To extract insights and knowledge from data
c) To design computer hardware
d) To manually process data

(ii) Select the key characteristic of Big Data.
a) Small volume
b) Structured only
c) High volume, velocity, and variety
d) Requires no processing

(iii) Select the main challenge in handling Big Data.
a) Limited storage capacity
b) Data being too structured
c) Difficulty in processing and analyzing large volumes
d) Lack of data generation

(iv) Examine which step is crucial in statistical inference.
a) Making predictions based on sample data
b) Collecting all data points in the population
c) Ignoring random variations
d) Using only qualitative data

(v) Examine which method ensures every individual in a population has an equal chance of selection.
a) Convenience sampling
b) Snowball sampling
c) Simple random sampling
d) Judgmental sampling

(vi) Select the definition of probability in statistical terms.
a) The likelihood of an event occurring
b) The total number of outcomes
c) The average of all possible events
d) The difference between two events

(vii) Select the probability value of an impossible event.

a) 1 b) 0
c) 0.5 d) Infinity

(viii) Select the probability rule that applies when two independent events occur together.

a) Addition rule b) Multiplication rule
c) Bayes' theorem d) Central limit theorem

(ix) Cite the key principle of Bayes' Theorem.

a) It describes the probability of an event based on prior knowledge b) It calculates the probability of independent events
c) It determines the mean of a dataset d) It only applies to mutually exclusive events

(x) Cite the type of probability that Bayes' Theorem helps to compute.

a) Prior probability b) Conditional probability
c) Joint probability d) Marginal probability

(xi) Select the primary purpose of a Chi-Square test.

a) To compare categorical variables for independence b) To estimate population means
c) To analyze time-series data d) To test for linear correlation

(xii) Examine the primary purpose of simple linear regression.

a) To establish causality between variables b) To predict the dependent variable using one independent variable
c) To analyze non-linear relationships d) To compare multiple regression models

(xiii) Examine how the coefficient of determination ($R^2$) is interpreted in regression analysis.

a) It measures the strength and direction of correlation b) It represents the proportion of variance in the dependent variable explained by the independent variables
c) It determines the statistical significance of coefficients d) It identifies outliers in the dataset

(xiv) Examine what happens when an independent variable is added to a multiple regression model.

a) Adjusted $R^2$ always increases b) The model complexity remains the same
c) The variance explained by the model may increase or decrease d) Multicollinearity is automatically eliminated

(xv) Examine the role of the intercept in a simple regression equation.

a) It represents the expected value of the dependent variable when all independent variables are zero b) It measures the slope of the regression line
c) It identifies outliers in the dataset d) It determines the strength of correlation

## Group-B

(Short Answer Type Questions) 3 x 5=15

2. Define the role of a data scientist. (3)
3. Describe the importance of programming languages in data science. (3)
4. Discuss the Chi-Square distribution and its significance in statistics (3)
5. Discuss the concept of data cleaning and why it is essential. (3)
6. Contrast descriptive and inferential statistics in EDA. (3)

**OR**

Contrast SPSS's GUI-based interface with syntax-based operations. (3)

## Group-C

(Long Answer Type Questions) 5 x 6=30

7. Justify the characteristics of the t-distribution and how it differs from the normal distribution. (5)
8. Classify the Different Types of Statistical Distributions with Examples. (5)
9. Describe the principles of statistical inference and its applications. (5)
10. Define datafication and explain its impact on businesses and society. (5)
11. Describe the role of datafication in artificial intelligence and automation. (5)
12. Explain the basic structure of a neural network (5)

**OR**

Explain how Python and R are used for data cleaning and preprocessing (5)

**********