BHARATI VIDYAPEETH INSTITUTE OF TECHNOLOGY
QUESTION BANK
Unit Test-I

Program: - Information Technology/Computer Technology	Program Code:- IF/CM
Course Title: - DATA ANALYTICS	Semester: - Fifth
Course Abbr &Code:- DAN (315326)	Scheme: K

Chapter 1: INTRODUCTION TO DATA ANALYTICS (CO1)

2 Marks

1. Define Data Analytics.
2. List any two types of Data Analytics.
3. What is meant by Quality and Quantity of data?
4. Define Mean, Median and Mode with its formula.
5. What is Central Limit Theorem (CLT)?
6. What is Sampling Variation?
7. Mention any two applications of Predictive Analytics.
8. Define Confidence Interval.
9. Differentiate between Nominal and Ordinal data .
10. What is the purpose of a Sampling Funnel?
11. Define variance and standard deviation with formula.

4 Marks

1. Explain the importance of Data Analytics in modern businesses.
2. Differentiate between Descriptive and Predictive Analytics with examples.
3. Explain the phases in Data Analytics Life Cycle.

4. Explain Measures of Central Tendency with formulas.
5. Discuss types of Data Analytics with suitable examples.
6. Explain types of data with examples.
7. What is Confidence Interval? How is it useful in sampling?
8. Interpret the concept of Sampling Funnel with a diagram.
9. Explain the role of measurement scales in data analytics.
10. Explain the Central Limit Theorem with proof.
11. Find 1) Mean 2)Median 3)Mode
    a) 50, 70, 40, 75, 30, 60, 80, 90, 100, 55.
    b) 5, 8, 12, 14, 18
12. Find variance and standard deviation of following data set

    i) 49,63,46,59,65,52,60,54

    ii)

Marks	5	15	25	35	45	55
No. of students	10	20	30	50	40	30

Chapter 2 – STATISTICAL ANALYSIS (CO2)

2 Marks

1. Define the following term

   a) Correlation b) Regression c) Skewness d) Kurtosis
2. Define ANOVA. State its Types
3. Name different type of Graphical Technique
4. What is Degrees of Freedom and how it is calculated?
5. What is the difference between linear regression and logistic regression?
6. Define Chi-square Test. State its formula
7. Define Box Plot and Scatter diagram

4 Marks

1. What is data cleaning? Why it is important in the Analytics. What are the different steps in Data cleaning Process?

2. Create a box plot using dataset of student test scores: 78, 85, 88, 92, 67, 73, 85, 90, 95, 100

3. A biologist samples 12 red oak trees in a forest plot and counts the number of caterpillars on each tree. The following is a list of the number of caterpillars on each tree:

   32, 47, 2, 15, 58, 24, 21, 11, 20, 50, 29, 36.

   Calculate the median, 1st and 3rd quartile, IQR and plot the Box Plot.

4. Find the regression lines for the following data:

   i)

x	6	2	10	4	8
y	9	11	5	8	7

   ii)

x	2	3	5	7	9
y	4	5	7	10	15

5. For 10 observations on price (x) and supply (y), the following data were obtained:

$$\Sigma x = 130, \quad \Sigma y = 220, \quad \Sigma x^2 = 2288, \quad \Sigma y^2 = 5506, \quad \Sigma xy = 3467$$

   Obtain the line of reflection of y on x and estimate the supply, when the price is 16 units.

6. Explain Scatter Diagram with example

7. Write Short Note on:

   i) Correlation and Regression Analysis

   ii) Skewness and Kurtosis

   iii) Imputation Technique

   iv) Probability Distribution

   v) Chi Square Test

8. A company conducted an employee satisfaction survey across three departments: Sales, IT, and HR. Each employee was asked to rate their job satisfaction as Satisfied, Neutral, or Dissatisfied. The responses are summarized in the following table:

	Satisfied	Neutral	Dissatisfied	Total
Sales	15	10	5	30
IT	10	15	5	30
HR	5	5	20	30
Total	30	30	30	90

Using a chi-square test for independence, determine whether there is a significant association between department and employee satisfaction level at the (level of significance 5 % , degree of freedom is 4 , critical value will be 9.488)

9. In a malaria campaign in a certain area, quinine was administered to 812 persons out of a total population of 3248. The number of fever cases is shown below. level of significance 5 % , degree of freedom is 1 , critical value will be 3.84)

Treatment	Fever	No Fever	Total
Quinine	140	30	170
No Quinine	60	20	80
Total	**200**	**50**	**250**

10. A certain drug was administered to 500 people out of a total of 800 included in a sample to test its efficiency against typhoid. The results are given below: level of significance 5 % , degree of freedom is 1 , critical value will be 3.84)

Contingency Table:

Treatment	Typhoid	No Typhoid	Total
Drug	200	300	500
No Drug	280	20	300
Total	480	320	800

Calculate value of $\chi 2$ (chi-square), and state whether the drug is effective in preventing typhoid.

11. State the procedure for Hypothesis testing.

12. Write short Note on ANOVA and its types.

13. A study was conducted to compare the effectiveness of three different exercise programs on weight loss over a period of 12 weeks. The weight loss (in pounds) for participants in each group is recorded below:

Group 1 (Aerobic)	Group 2 (Strength)	Group 3 (Yoga)
8	6	4
10	7	5
9	5	3
7	8	6
6	6	4

Using this data, test if there is a significant difference in the mean weight loss among the three exercise groups at the 5% significance level.

14. Test whether the sample having the values 63, 63, 64, 55, 66, 69, 70, 70, 71 has been chosen from a population with mean 65 at 5% level of significance (Consider t value=2.306)

15. Maths test score of group-A of 16 students showed a mean of 107 with a variance of 100. Another group-B of 11 students showed a mean score of 98 with a variance of 225.

    Is there a significant difference between the scores for the two groups at 0.05 (5 %) level of

significance? (Consider t value=2.12)

16. A manufacturer claims that the average life of their electric light bulbs is 2000 hours. A random sample of 64 bulbs is tested and the life, x, in hours recorded. The results are as below:

$$\sum x = 127808 \quad \sum(\bar{x} - x)^2 = 9694.6$$

Is there sufficient evidence, at the 1% level, that the manufacturer is overestimating the length of the life of the light bulbs?(

17. Test the significant difference between the means of the samples, drawn from two normally distributed populations with the same S.D., using the following information.

	Size	Mean	S.D.
Sample-1	100	61	4
Sample-2	200	63	6

18. Three unbiased coins are tossed. What is the probability of getting:

(a) all heads, (b) exactly one head, (c) exactly two heads, (d) at least one head

19. A fair die is rolled. What is the probability that the number on the die is

(a) 5, (b) an odd number, (c) a number greater than 1, (d) a multiple of 4?

20. A card is taken out at random from a pack of well shuffled pack of 52 playing cards.

What is the probability that it is

(a) a seven, (b) a heart, (c) a red card, (d) a red six.