



**PRESIDENCY UNIVERSITY  
BENGALURU**

**SCHOOL OF ENGINEERING**

**MAKE UP EXAMINATION – JAN 2023**

**Course Code:** CSE 2027

**Course Name:** Fundamentals of Data Analysis

**Program** : B.Tech

**Date:** 28-JAN-2023

**Time:** 09.30 AM to 12.30 PM

**Max Marks:** 100

**Weightage:**50%

**Instructions:**

- (i) Read the all questions carefully and answer accordingly.
- (ii) Answer all questions
- (iii) You may use a non-programmable scientific calculator where necessary
- (iv)** Statistical table will be provided.

**Part A [Memory Recall Questions]**

**Answer all the Questions. Each question carries TWO marks. (10Qx 2M= 20M)**

1. Give the classification of “Variables”. (C.O.No.1) [Knowledge]
2. Define Data Analysis and explain various types of data analysis. (C.O.No.1) [Comprehension]
3. What is meant by sampling? List the various types of probability sampling methods. (C.O.No.2) [Comprehension]
4. The marks obtained by 40 students out of 50 in a class are given below in the table.

Marks	42	36	30	45	50
Number of Students	7	10	13	8	2

Find the mode of the above data. (C.O.No.2) [Application]

5. Write the significance of line graph, pie chart and scatter plot. (C.O.No.3) [Comprehension]
6. Distinguish between the primary and secondary data. (C.O.No.3) [Knowledge]
7. Differentiate Descriptive vs. Inferential statistics. (C.O.No.4) [Knowledge]
8. Define Correlation Analysis and explain the types of correlation analysis. (C.O.No.4) [Application]
9. Define the predictive model (C.O.No.5) [Knowledge]
10. Illustrate with an example two reasons for variable removal during data cleaning. (C.O.No.3) [Application]

### Part B [Thought Provoking Questions]

Answer all the Questions. Each question carries EIGHT marks. (4Qx8M=32M)

11. Sketch the sampling distribution, assume there is a population size is  $N=4$ . Random variable,  $x$  is age of individuals. *The value* of  $X$  is 18, 20, 22, 24 (years). (C.O.No. 2) [Application level]

12. Identify the contingency table represents all possible outcomes for two binary variables:

**Count:11** The number of observations that were true and predicted to be true.

**Count:10** The number of observations that were false yet predicted to be true.

**Count:01** The number of observations that were true and predicted to be false

**Count:00** The number of observations that were false and predicted to be false.

(C.O.No. 5) [Application level]

13. Interview method of collecting data involves presentation of oral verbal stimuli and deeply in terms of oral- verbal responses. Identify the two ways. (C.O.No. 3) [Application level]

14. List out the typical human-generated unstructured data includes.

(C.O.No. 1) [Knowledge level]

### Part C [Problem Solving Questions]

Answer all the Questions. Each question carries SIXTEEN marks. (3Qx16M=48M)

15. Table showing the data set for Actual and predicted response

Actual response	Predicted response
True(1)	True(1)
False(0)	False(0)
False(0)	False(0)
True(1)	True(1)
True(1)	False(0)
False(0)	True(1)
True(1)	True(1)
False(0)	False(0)
True(1)	True(1)
False(0)	False(0)
False(0)	False(0)

True(1)	True(1)
True(1)	False(0)
True(1)	True(1)
False(0)	False(0)
False(0)	False(0)
True(1)	True(1)
True(1)	True(1)

a) Calculate the contingency table summarizing correct and incorrect predictions

b) Execute the solutions for

- i) Concordance
- ii) Error rate
- iii) Sensitivity
- iv) Specificity

(C.O.No. 5) [Application level]

16. Charting the various charts with the following data sets

<b>Cylinders</b>	<b>Count</b>	<b>Mean(MPG)</b>
3.0	4	20.55
4.0	199	29.28
5.0	3	27.37
6.0	83	19.97
8.0	103	14.96

- i) Categorical Data
  - a) Bar Charts
  - b) Pie Charts
  - c) Pareto Diagram
  
- ii) Numerical Data
  - a) Stem and Leaf Display
  - b) Histogram
  - c) Line graph
  - d) O give
  - e) Scatter plot

(C.O.No. 4) [Application level]

17. Let's say we want to determine if on average girls score more than 600 in the exam. We do not have the information related to variance (or standard deviation) for girls' scores. To perform t-test, we randomly collect the data of 10 girls with their marks and choose our  $\alpha$  value (significance level) to be 0.05 for display the Hypothesis Testing value.

Girls Score: 587, 602, 627, 610, 619, 622, 605, 608, 596, 592

(C.O.No. 2) [Application level]