PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST 1

Winter Semester: 2021 - 22

Course Code: CSE 2027

Course Name: Fundamentals of Data Analysis

Program & Sem: B.Tech IV Semester

Date: 27/4/2022

Time: 11.30 am to 12.30 pm

Max Marks: 30

Weightage: 15 %

Instructions:

(i)   Read the all questions carefully and answer accordingly.

(ii)   (ii) Answer all questions

(iii)   (iii) You may use a non-programmable scientific calculator where necessary

Part A [Memory Recall Questions]

Answer all the Questions. Each question carries two marks.   **(5Qx 2M= 10M)**

1. Define Data Analysis. Give any four data analysis tools.   (C.O.1) [Knowledge]

2. Explain the differences between structured & unstructured data.   (C.O.1) [Comprehension]

3. What are the different sources of data?   (C.O.1) [Knowledge]

4. What is simple random sampling? Explain with an example.   (C.O.2) [Knowledge]

5 Calculate the Standard Error of the Mean of {10, 20, 30, 40, 50}.   (C.O.2) [Application]

Part B [Thought Provoking Questions]

Answer both the Questions. Each question carries five marks.   **(2Qx5M=10M)**

6. "What is likely to happen" by using previous data. The simplest data analysis example is like if last year I bought two dresses based on my savings and if this year my salary is increasing double then I can buy four dresses. But of course it's not easy like this because you have to think about other circumstances like chances of prices of clothes is increased this year or maybe instead of dresses you want to buy a new bike, or you need to buy a house! Identify the scenario as to what type of analysis   (C.O.No.1)[Knowledge]

7. P is a measure used to indicate the value below which a given percentage of observation falls is 69, 98, 82, 77, 71, 84, 55, 94, 84, 64, 92, 63, 74, 81, and 76. Calculate the percentiles L_50, the quantiles, the range and the IQR.   (C.O.No.2) [Applications]

Part C [Problem Solving Questions]

Answer the Question. The question carries ten marks.   **(1Qx10M=10 M)**

8. From a statistics standpoint, the standard deviation of a dataset is a measure of the magnitude of deviations between the values of the observations contained in the dataset. From a financial standpoint, the standard deviation can help investors quantify how risky an investment is and determine their minimum required return on the investment. Solve the standard deviation for the following 12 weeks data set.   (C.O.No. 2) [Applications]

Weeks	Expenditure
1	$48.50
2	$87.40
3	$19.98
4	$59.74
5	$40.87
6	$105.51
7	$40.80
8	$23.10
9	$98.10
10	$60.54
11	$64.81
12	$48.01

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST 2

Winter Semester: 2021 - 22

Course Code: CSE 2027

Course Name: Fundamentals of Data Analysis

Program & Sem: B.Tech IV Semester

Date: 2nd June 2022

Time: 11.30AM - 12.30 PM

Max Marks: 30

Weightage: 15 %

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.   **(5Qx 2M=10M)**

1. Write down the advantages & disadvantages for interview method.   (C.O.3)[Comprehension]

2. List out the uses of Correlation Analysis   (C.O.3)[Comprehension]

3. Define Inferential Statistics and write down the types of inferential statistics

(C.O.2)[Comprehension]

4. In a box, there are 8 red, 7 blue and 6 green balls. One ball is picked up randomly. What is the probability that it is neither red nor green?   (C.O.2) [Knowledge]

5. Draw the one tailed t-test and two tailed t-test.   (C.O.2) [Knowledge]

Part B [Thought Provoking Questions]

Answer both the Questions. Each question carries FIVE marks.   **(2Qx5M=10M)**

6. This table shows a random samples of 100 hikers and the areas of hiking they prefer. Hiking area preference.

Gender	The coastline	Near lakes and streams	On mountain peaks	Total
Female	18	16	…………	45
Male	…………..	…………………	14	55
Total		41		

Find

a. P(F AND C)

b. P(F)* P(C)

c. P(F AND C) ≠ P(F)*P(C) check F and C independent or not? (C.O.No.2)[Application]

7. Write down the sample format for both the questionnaire and schedule. (C.O.No.3) [Knowledge]

Part C [Problem Solving Questions]

Answer the Question. The question carries TEN marks. **(1Qx10M=10 M)**

8. let's say we want to determine if on average, boys score 15 marks more than girls in the exam. We do not have the information related to variance (or standard deviation) for girls' scores or boys' scores. To perform a t-test. we randomly collect the data of 10 girls and boys with their marks. We choose our α value (significance level) to be 0.05 as the criteria for Hypothesis Testing (Two sample **T Test**). (C.O.No. 2) [Applications]

Girls	Boys
587	626
602	643
627	647
610	634
619	630
622	649
605	625
608	623
596	617
592	607

Roll No | | | | | | | | | | |

PRESIDENCY UNIVERSITY
BENGALURU
SCHOOL OF ENGINEERING

ENDTERM EXAMINATION

Winter Semester: 2021 - 22

Course Code: CSE 2027

Course Name: Fundamentals of Data Analysis

Program & Sem: 4th sem B.E

Date: 30th June 2022

Time: 09.30 AM to 12.30 PM

Max Marks: 100

Weightage: 50%

Instructions:
(i) *Read the all questions carefully and answer accordingly.*
(ii) *You may use a non-programmable scientific calculator where necessary*

Part A [Memory Recall Questions]

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

1. Define Data Analysis?   (C.O.1) [Knowledge]
2. List various types of Data Analysis.   (C.O.1) [Knowledge]
3. Define Sampling. And list the types of Sampling Techniques.   (C.O.2) [Comprehension]
4. Define Data Visualization. List the different types of charts used for visualization of   data.
                                                              (C.O.4) [Comprehension]
5. "The response is a categorical variable & Continuous variable" Identify the type of Model.
                                                              (C.O.5) [Knowledge]
6. List the various Vs that are used to describe the nature of Big Data.   (C.O.1) [Knowledge]
7. An agriculturalist wants to find out if there exists a correlation between the chemical composition of his soil samples and turbid water bodies in the area. List the three stages of analysis that the agriculturalist has to perform.   (C.O.3) [Knowledge]
8. Explain the significance of the metric 'Residual' used in Regression Analysis.
                                                              (C.O.5) [Comprehension]
9. Name another term used for significance level in inferential statistics and give the generally chosen value for the same.   (C.O.2) [Knowledge]
10. Define scales of data and provide a brief explanation on the same.   (C.O.1) [Knowledge]

Part B [Thought Provoking Questions]

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

11. List and Explain the types of Observation Method.   (C.O.No.3)[Knowledge]
12. Explain the methodology involved in Data Collection using Questionnaire & Schedule and give the comparison.   (C.O.No.3) [Application]

13. Data is not always collected for purpose of Analysis. Explain with appropriate examples the other reasons for data collection which eventually become sources of data.

(C.O.No.3) [Comprehension]

14. Compute the sampling distribution for age variable of values: 18, 20, 22, 24 in a step wise manner

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

Part C [Problem Solving Questions]

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

15. Identify the type of Interview Method in questions a. and b. and provide appropriate answers for parts c. and d.

a. "Focus attention on the given experience of the respondent and its effect. The interviewer may ask questions in any manner or sequence with the aim to explore reasons and motives of the respondent". [4M]

b. "It is concerned with broad underlying feeling and motives or individuals' life experience which are used as method to collect information under this method at the interviewer direction". [4M]

c. List the advantages and Limitations of Interview Method. [4M]

d. In what manner is interview method different from other data collection methods [4M]

(C.O.No. 3) [Comprehension]

16. Let us assume we want to know the average performance score more than 600 employee. We do not have the information related to variance (or standard deviation) for employees. To a perform t-test, we randomly collect the data of 10 employees with their performance score and choose our α value (significance level) to be 0.05 for Hypothesis Testing.

(C.O.No. 2) [Knowledge]

Employee score 587,602,627,610,619,622,605,608,596,592

a) Find the Mean value & Standard Deviation [4M]

b) Find the t value [4M]

c) Find the degree of freedom [2M]

d) Find the critical value [2M]

e) Find the P value [2M]

f) Find the conclusion [2M]

17. An analyst randomly selects 20 values from a weighing machine and records the daily high values:

24, 35, 17, 21, 24, 37, 26, 46, 58, 30, 32, 13, 12, 38, 41, 43, 44, 27, 53, 27.

Construct a frequency distribution table for these recorded observations.

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