Indian Institute of Information Technology Sri City

6th September 2023

Name of the Subject: IDA Duration: 90 mins Max Marks: 20

Roll Number: S20210020328

Instructions:

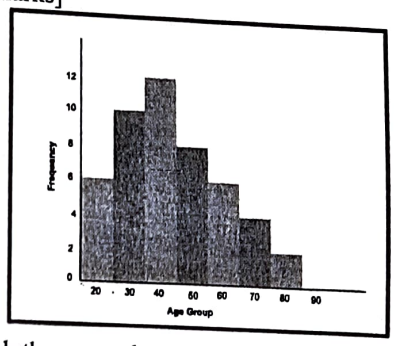
1. *This is a closed book, closed notes exam.*
2. *For descriptive questions be brief and to-the-point. Answers must be given in ball point pen only. Answers in pencils will not be checked.*
3. ***For Part-B, mark the answers in the question paper and attach the question paper with the answer sheet.***
4. *You are allowed to use calculators.*
5. *Follow all other instructions given by the faculty during the exam.*

PART – A

Question 1 [5 marks]

A) A person had to wait for a bus to work on 10 working days. The following are number of minutes he waited. 10, 1, 13, 9, 5, 2, 10, 3, 8, 6. Obtain the five-number summary for the given data and construct the box-plot. [2 marks]

B) From the given graph explain the following



Explain whether the given data is skewed or not. What is the right measure of location for the given data and calculate the same. [2 marks]

C) Describe the given attribute as either discrete or continuous. Also classify them as nominal, ordinal, interval or ratio. Some cases may have more than one interpretation. So briefly indicate your reasoning if you think there may be some ambiguity. [1 mark]

i) Military rank
ii) Anomaly detection in bank

Question 2 [5 marks]

A) Explain different Linear data transformation techniques. [1.5 marks]

B) For the following pairs of observations from a population, Calculate the Pearson Product Moment Correlation Coefficient and conclude whether it is positively correlated or negatively correlated. [1.5 marks]

Age	43	21	25	42	57	59
Glucose Level	99	65	79	75	87	81

C) A tire manufacturer wants to determine the inner diameter of a certain grade of tire. Ideally, the diameter would be 570 mm. The data are as follows: [2 marks]

572, 572, 573, 568, 569, 575, 565, 570.

i. Find the sample mean and median.
ii. Find the sample variance, standard deviation, and range.
iii. Using the calculated statistics in parts (i) and (ii), can you comment on the quality of the tires?

6th September 2023

PART – B [10*1 = 10 marks]

1. The concentration of DDT, in milligrams per litre, is:
   a) a nominal variable b) an ordinal variable
   c) an interval variable d) a ratio variable.

2. Dataset with 100 (non-zero) observations are taken for inference. If the largest value is doubled in the dataset, which of the following option is false?
   a) the variance increases b) the mean increases
   c) the median increases d) the range increases

3. A set of data points follow a simple linear relation $y = 3x + 2$, where x is any integer number. The mean of the values of y for all values of x in the range [1 ... 100] is
   a) 50 b) 50.5 c) 152 d) 153.5

4. Which cannot be measured with "Categorical" data?
   a) Mean b) Median c) Mode d) All of the above

5. In a right-skewed histogram, the value of which of the following is the lowest?
   a) Mean b) Median c) Mode d) All three are equal

6. The ratings of four movies of an actor are 8, 10, 15 and 20. What is the z-score normalized value of this data?
   a) -1.14, -4.7, 0.3, 1.4 b) -1.11, 1.32, 5.10, 1.06
   c) -0.97, -0.60, 0.32, 1.25 d) -5.25, -3.25, 1.75, 6.75

7. A sample of 100 IQ scores produced the following statistics: mean = 95; lower quartile = 70; median = 100; upper quartile = 120; mode = 75; standard deviation = 30. Which statement(s) is (are) correct?
   a) Half of the scores are less than 95
   b) The middle 50% of scores are between 100 and 120
   c) One-quarter of the scores are greater than 120
   d) The most common score is 95

8. If the standard deviation is zero, you can conclude that
   a) All values of the variable appear with equal frequency
   b) All values of the variable have the same value
   c) The mean of the values is also zero
   d) None of the above is correct

9. Which is not related to the characteristics of Big data?
   a) Speed b) Complexity c) Size d) Computability

10. The GM of the following data will be calculated as X = [50, 125, 70, 56, 49, 98]
    a) 101 b) 74 c) 100 d) 70

---*All the best*---