Paper / Subject Code: 48815 / Statistics for Artificial Intelligence & Data Science (DLOC - I)

Dec'2023

TE Sem V R-2019 (C Scheme) "AI & DS"

04/12/2023

Duration: 3hrs **[Max Marks:80]**

(1) Question No 1 is Compulsory.
(2) Attempt any three questions out of the remaining five.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

1 Attempt any **four** **[20]**

a) Write a short note on hypothesis testing.
b) What is Fisher's exact test?
c) Write a short note Simple Linear Regression
d) Write a short note on Random sampling
e) What is the empirical CDF function?

2 a) Construct a frequency distribution table for the following weights (in gm) of 30 oranges using the equal class intervals, one of them is 40-45 (45 not included). The weights are: 31, 41, 46, 33, 44, 51, 56, 63, 71, 71, 62, 63, 54, 53, 51, 43, 36, 38, 54, 56, 66, 71, 74, 75, 46, 47, 59, 60, 61, 63. **[10]**

(a) What is the class mark of the class intervals 50-55?
(b) What is the range of the above weights?
(c) How many class intervals are there?
(d) Which class interval has the lowest frequency?

b) What is the primary purpose of conducting a one-way ANOVA. Explain the key components of a one-way ANOVA, including the dependent variable, independent variable, and factors. **[10]**

3 a) Find the standard error of the estimate for the average number of children in a household in your city by using the data collected from a sample of households in your city. Then find a 95% confidence interval for the data. **[10]**

Household	No. of children
1	2
2	3
3	1
4	0
5	5
6	2
7	1
8	4

b) What is the concept of correlation in statistics, how is it different from regression? **[10]**

4 a) A radar unit is used to measure speeds of cars on a motorway. The speeds are normally distributed with a mean of 90 km/hr and a standard deviation of 10 km/hr. What is the probability that a car picked at random is travelling at more than 100 km/hr? [10]

b) Explain Numerical and Categorical data types with appropriate examples [10]

5 a) Duracell manufactures batteries that the CEO claims will last an average of 300 hours under normal use. A researcher randomly selected 20 batteries from the production line and tested these batteries. The tested batteries had a mean life span of 270 hours with a standard deviation of 50 hours. Do we have enough evidence to suggest that the claim of an average lifetime of 300 hours is false? [10]

b) Explain linear least square regression (LLSR) along with it's advantages and disadvantages. [10]

6 a) A farmer is trying out a planting technique that he hopes will increase the yield on his pea plants. The average number of pods on one of his pea plantsis 145 pods with a standard deviation of 100 pods. This year, after trying his new planting technique, he takes a random sample of his plants and finds theaverage number of pods to be 147. He wonders whether or not this is a statistically significant increase. What are his hypotheses and the test statistic? [10]

b) What is the Chi-Square Test in statistics, and in what kind of situations or research scenarios is it commonly used? [10]

****

2AE962E453B326368F9D42C43A4D1AA2