

22/11/2024 CSE-AIML SEM-V C SCHEME DLOC-STAT. FOR AIDS QP CODE: 10067675

**Time: 3-Hour****Max. Marks: 80****N.B**

1. Q.1 is compulsory
2. Attempt any three from the remaining five questions.
3. Assume suitable data, if required and state it clearly.

**Q1. Attempt any 4****20M**

- a. Explain percentiles and Boxplots with example.
- b. Illustrate central limit theorem with a neat diagram.
- c. Explain Null and alternative Hypothesis with example.
- d. What is Histogram. Give its applications.
- e. How to detect outliers?
- f. What is the F-Test used for?

**Q 2.****10M**

- a. What is Chi-Square Test? A study is conducted to examine the relationship between gender and whether a person prefers coffee or tea. The data collected is as follows:

Gender \ Preference	Coffee	Tea	Total
Male	30	10	40
Female	20	30	50
Total	50	40	90

Use chi-square test to find association between gender and beverage preference? (Use  $\alpha=0.05$ ).

- b. Explain Normal and Poisson Distribution. Most graduate schools of business require applicants for admission to take the Graduate Management Admission Council's GMAT examination. Scores on the GMAT are roughly normally distributed with a mean of 527 and a standard deviation of 112. That is the probability of an individual scoring above 500 on the GMAT? How high must an individual score on the GMAT in order to score in the highest 5%?

**10M****Q3.****10M**

- a. Explain single and Multiple linear regression with example and show with suitable plot.
- b. Explain t-Distribution in detail. The CEO of light bulbs manufacturing company claims that a light bulb lasts 300 days. A researcher randomly selects 15 bulbs testing. The sampled bulbs last an average than 290 days? deviation of 50 days. If the CEO's claim were true, what is the probability that 15 randomly selected bulbs would have an average life of no more than 290 days?

**10M****Q 4.****10M**

- a. Explain briefly why use ANOVA? Give difference between one-way and two-way ANOVA test. Solve the following using one way annova

Col1	Col 2	Col 3
82	71	64
93	62	73
61	85	87
74	94	91
69	78	56
70	66	78
53	71	87

- b. Explain Mean, standard deviation, standard error and Confidence Interval. Following table shows values of 10 data points of a sample. Find mean, standard deviation, standard error, and 95% confidence interval for the given sample. **10M**

Data	Value
1	4
2	2
3	2
4	3
5	5
6	2
7	6
8	2
9	7
10	6

- Q.5. a.** Explain how Nonparametric Method are different from parametric methods. **10M**  
 Department of Public Health and safety monitors the measures taken to cleanup drinking water were effective. Trihalomethanes (THMs) at 12 counties drinking water compared before cleanup, 1 week later, and 2 weeks after cleanup. solve the following using the Friedman Test ?

County	Trihalomethanes (THMs)		
	Before Cleanup	Week1	Week2
1	21.1	19.2	18.4
2	24.1	22.3	21.2
3	14.1	12.9	12.9
4	18.1	17.8	17.3
5	15.4	15.1	14.9
6	16.2	15.1	15.1
7	7.4	7.2	6.8
8	7.5	6.7	6.1
9	14.2	13.6	13.1
10	21.3	20.9	20.4
11	9.5	9.8	9.2
12	11.9	10.5	10.1

- b. Define Continuous Probability distribution and Probability Distribution Function (PDF). Consider a random variable  $X$  that is uniformly distributed between 0 and 20. This means that the probability density function (PDF) of  $X$  is given by:  $f(x)=1/20$ , for  $0 \leq x \leq 20$ .
- i) Calculate the probability that  $X$  lies between 5 and 15, i.e.,  $P(5 \leq X \leq 15)$ .
  - ii) Calculate the probability that  $X$  is less than 10, i.e.,  $P(X < 10)$
  - iii) Find the probability that  $X$  is greater than 18, i.e.,  $P(X > 18)$

**10M**

**Q 6.** Write short note

**20M**

- a. Type I and Type II error in detail.
  - b. Histogram and Scatterplot with example.
  - c. Boot strapping Vs Re-sampling.
  - d. Weibull Distribution.
-