**Indian Statistical Institute, Bangalore**
**B. Math II, First Semester, 2024-25**
**Back Paper Examination**
**Introduction to Statistics and Computation with Data**
**05.06.25** **Maximum Score 70** **Duration: 2 Hours**

Values from the normal distribution qnorm(0.9)=1.281552, qnorm(0.995)=2.575829, qnorm(0.95)=1.644854
Values from the t distribution qt(.95,9)=1.833113, qt(.95,10)=1.81, qt(.975,9)=2.26, qt(.975,10)=2.23

Students are allowed to bring calculator and two one-sided pages of notes.

1. (12+3=15) Let $X_1, \cdots, X_n$ be a random sample from a distribution with pdf

$$f(x) = \lambda \exp(-\lambda(x-\theta)), \quad \text{if} \quad x > \theta$$

where $\lambda \in R+$ and $\theta \in R$.

   (a) Find the method of moments estimators of $\theta$ and $\lambda$.

   (b) Are the estimators consistent?

2. (5+5=10) Explain any two of the following methods/concepts related to the class presentations.

   (a) Variance Inflation Factor and Cook's Distance

   (b) Decision tree and Random forest

   (c) Maximum Likelihood Estimator, Fisher Information Matrix and their relation.

3. (5+5+5=15) Let $U \sim Unif(0,1)$ and $W = \tan(\pi(U - 1/2))$.

   (a) Find the distribution of $W$.

   (b) Use this to describe how to generate a random variable from the Cauchy distribution with parameters 10 and 5.

   (c) Find the mean, median and first quartile of the distribution in part (b).

4. (5+5+5=15) From a bivariate dataset $(x_1, y_1), \cdots, (x_n, y_n)$, we obtain two least squares regression lines $2y = 3x + 5$ and $3y = 5x + 7$. One line is for the regression of $Y$ on $X$ and the other is for the regression of $X$ on $Y$.

   (a) What are the sample means of $X$ and $Y$ ?

   (b) What is the value of the correlation coefficient between $X$ and $Y$?

   (c) Which one of the two lines is the regression line of $Y$ on $X$? Explain.

5. (2+4+2+4+3=15)A certain company of battery claims that the average life of their battery is 80 weeks. The average life of each of 10 randomly selected batteries is listed below.

$$83.5, 81.0, 78.3, 80.0, 81.2, 81.1, 78.3, 78.9, 77.2, 75.8$$

Assume the battery life distribution is normal. Using $\alpha = 0.05$ perform the appropriate test of hypothesis to determine if the average life is smaller than 80 weeks.

(a) State the null and alternative hypotheses.

(b) State the test statistic and its distribution under the null hypothesis.

(c) State all assumptions

(d) Find the critical region and compute the value of the test statistic.

(e) Is the null hypothesis rejected? What is the conclusion regarding the mean battery life?