Indian Statistical Institute, Bangalore
B. Math I, Second Semester, 2024-25
Final Examination
Introduction to Statistics and Computation with Data

28.04.25 **Maximum Score 100** **Duration: 3 Hours**

Students are allowed to bring calculator and two one-sided pages of notes.
Values from the normal distribution qnorm(0.975)=1.959964, 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

1. (10) Let $X_1, \cdots, X_n$ be a random sample from a uniform(θ_1, θ_2) distribution. Find the method of moments estimators of θ_1 and θ_2.

2. (5+5=10) Answer any two of the following questions related to the class presentations.
   (a) Describe some procedures to test the ANOVA assumptions like Levene's test, Brown-Forsythe test, Welch's test.
   (b) Explain Simpson's paradox
   (c) Describe the set-up of linear regression with multiple predictors and the corresponding least squares approach.

3. (10+10=20) Consider a random variable Y with triangular distribution whose probability density function (pdf) is given by
$$f(y) = \begin{cases} 4y & \text{if} \quad 0 < y < 1/2 \\ 4 - 4y & \text{if} \quad 1/2 < y < 1 \end{cases}$$
   Assume that you can generate observations on uniform [0, 1].
   (a) How would you draw observations on Y using probability integral transform?
   (b) If U and V are independent uniform [0, 1] then obtain the distribution of $(U+V)/2$. Use this result to describe an alternative procedure to generate observations on Y.

4. (10) The lifetime of each of 10 randomly selected batteries, in weeks, produced by a certain company is listed below.
$$84.5, 80.0, 77.3, 81.0, 80.2, 80.1, 78.3, 79.9, 77.8, 75.4$$
   Find a 95% confidence interval for the average lifetime of all batteries produced by the company, stating all assumptions.

5. (2+2+4+4+4+4=20) Consider the following R code and output.

```
> x<-c(45,92,287,98)
> chisq.test(x,p=c(0.2,0.2,0.4,0.2))

Chi-squared test for given probabilities

data:  x
X-squared = 64.949, df = 3, p-value = 5.143e-14
```

   (a) What is the underlying model for generating the data x?
   (b) State the null and alternative hypotheses.
   (c) State the test statistic and its distribution under the null hypothesis.
   (d) Is the distribution in the previous part exact or approximate? In the latter case, what is the condition for the approximation to work?
   (e) What is the value of the test statistic and the corresponding p-value?
   (f) Is the null hypothesis rejected at α=5%? What is the interpretation in terms of parameters of the model in part (a)?

6. (12+3=15) Find the covariance between $\hat{\alpha}$ and $\hat{\beta}$, the least squares estimators of α and β in simple linear regression. State the model and all assumptions.

7. (2+3+6+4=15) The interest is in testing if the median daily ammonia concentrations (parts per million) at a particular location is more than 1.5. It is known from previous studies that the distribution is not normal. The daily ammonia concentration on eight randomly selected days are given below.

$$1.53, 1.50, 1.37, 1.51, 1.55, 1.42, 1.41, 1.48$$

   (a) Set up the null and alternative hypotheses for the test.
   (b) Find the value of the test statistic.
   (c) What is hte distribution of the test statistic?
   (d) Find the p-value of the test.
   (e) Give the appropriate conclusion in terms of the median ammonia concentration at 5 % level.