

## MSc Sem.-1 Examination

401

Data Science

February-2025

Time : 2-30 Hours]

[Max. Marks : 70

**Instructions:** All questions are compulsory. Use of non-programmable scientific calculators is allowed.

- Q.1** (a) A medical test is used to detect a rare disease that affects 1 in 1,000 people. The test is 99% accurate (07)  
 If a person has the disease, there is a 99% probability that the test will be positive (true positive rate).  
 If a person does not have the disease, there is a 99% probability that the test will be negative (true negative rate).  
 If a person tests positive, what is the probability that they actually have the disease? Using Bayes' Theorem.
- (b) Explain Cumulative Distribution Function with properties. (07)

OR

- (a) Consider a standard deck of 52 playing cards. What is the probability of drawing a King given that the card drawn is a face card (Jack, Queen, or King)? (07)  
 Total number of face cards = 12 (3 face cards per suit  $\times$  4 suits) and  
 Number of Kings = 4.  
 The probability of drawing a King given that the card is a face card?
- (b) Explain Moment Generating Function with properties. (07)
- Q.2** (a) Assume you have a dataset of returns on a particular investment. The average return is 12%, and the standard deviation is 6%. You want to determine the probability that a return falls within 2 standard deviations (12%) of the mean using Chebyshev's inequality. (07)
- b] Find Cumulant Generating Function of Normal Distribution (07)

OR

- (a) Let  $X$  be a continuous random variable with the probability density function (PDF) (07)  

$$\begin{cases} 2(1-x) & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$
- Derive the Moment Generating Function (MGF) of  $X$ .
- b) Find Probability Generating Functions of Poisson Distribution. (07)

- Q.3 (a) Suppose X and Y are continuous random variables with a joint PDF defined as: (07)

$$f_{X,Y}(x, y) = \begin{cases} c \cdot (x + y) & \text{for } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

where  $c$  is a constant. Find  $c$  and verify that  $f_{X,Y}(x, y)$  is a valid PDF.  
Find Joint Probability Density Function and determine the value of  $C$ .

- (b) Find Mean and Variance of Negative Binomial Distribution (07)

OR

- (a) A group of students claims that their average test score is 75. However, a teacher suspects this is an overestimate. To test this, she takes a random sample of 10 students' scores, obtaining: (07)  
72, 74, 78, 71, 73, 75, 76, 70, 74, and 73.  
Table value is: 2.262

- (b) Find Mean and Variance of Geometric Distribution. (07)

- Q.4 (a) Consider two samples representing the weights (in grams) of a particular species of fruit from two different orchards: (07)

**Orchard A:** 150, 160, 165, 155, 170 and **Orchard B:** 180, 190, 200, 185, 195

Check that two populations have equal variances against the alternative hypothesis that they do not.

Table value is 6.338

- (b) Explain properties and Find Mean and Variance of Exponential Distribution (07)

OR

- (a) Explain properties and Find Mean and Variance of Lognormal Distribution (07)

- (b) In the pharmaceutical industry, the F-test is often used to compare the variances of two different batches of a drug or formulation to check if their variability is significantly different. (07)

A pharmaceutical company is testing the variance in the dissolution rates of two different drug batches:

- **Batch 1:** variance = 600, sample size = 8
- **Batch 2:** variance = 400, sample size = 10

Ftable value is 3.293

- Q.5** Attempt any **SEVEN** out of **TWELVE (True or False)** **(14)**
- (1) The Cumulative Distribution Function (CDF) of a random variable is non-decreasing.
  - (2) Chebyshev's inequality provides an upper bound on the probability that the value of a random variable deviates from its mean.
  - (3) The Probability Density Function (PDF) can take negative values.
  - (4) The expectation of a random variable is always equal to its median.
  - (5) If two random variables are independent, their joint probability density function equals the product of their marginal density functions.
  - (6) The F-distribution is used to compare the variances of two independent populations and is always positively skewed.
  - (7) When conducting a test of significance for the difference between two population means with known variances, the z-test is used.
  - (8) Type II error occurs when the null hypothesis is incorrectly rejected when it is actually true.
  - (9) Random sampling techniques ensure that every member of the population has an equal chance of being selected in the sample.
  - (10) The Hypergeometric distribution is used when sampling is done with replacement.
  - (11) The Normal distribution is characterized by its bell-shaped curve and is defined by its mean and variance.
  - (12) The Chi-square distribution is used primarily in hypothesis testing and is always symmetric.
  - (13) The t-distribution is used when the sample size is small and the population standard deviation is unknown.
  - (14) The Bernoulli distribution models a single trial with two possible outcomes: success and failure.
-