[This question paper contains 4 printed pages.]

Your Roll No...............

Sr. No. of Question Paper : 7387 **L**

Unique Paper Code : 2353200019

Name of the Paper : Mathematical Statistics

Name of the Course : **B.A. / B.Sc. Programme - DSE**

Semester : VIII

Duration : 3 Hours Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Attempt **all** questions by selecting any **two** parts from each question.
3. Each question carry equal marks.
4. Use of non-programmable scientific calculators and statistical tables is permitted.

Q1 (a) Define joint probability mass function and marginal probability mass functions.

A consumer electronics store sells extended warranty plans for two types of products: smartphones and laptops. For each product, customers can choose different deductible options for repair claims. For a smartphone warranty, the deductible options are ₹100 and ₹250, whereas for a laptop warranty, the options are ₹0, ₹100, and ₹200. Suppose a customer who has purchased warranty plans for both a smartphone and a laptop is selected at random from the store's database. Let X denote the deductible amount (in ₹) for the smartphone warranty and Y denote the deductible amount (in ₹) for the laptop warranty. The joint probability mass function of (X, Y) is given below:

		Y		
	$p(x,y)$	0	100	200
X	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

(i) Find the marginal probability distributions of X and Y.

P.T.O.

(ii) Define independent random variables and determine whether X and Y are independent.

(iii) Compute $E[X], E[Y], Cov\ (X,Y)$, and the correlation coefficient ρ.

(b) Define the moment generating function (MGF) of a continuous random variable X. State the conditions under which it exists. Obtain the first five moments of a random variable from its moment generating function.

(c) Let X and Y be continuous random variables. The random variable Y has probability density function $f_Y(y) = \{2y, 0 < y < 1;\ 0, otherwise\}$.The conditional distribution of X given Y = y satisfies $E[X \mid Y = y] = y,\ Var(X \mid Y = y) = y^2$

(i) Distinguish between conditional variance and unconditional variance.

(ii) Using the given information, compute $E[Var(X \mid Y)]$ and $Var(E[X \mid Y])$

Q2 Let X_1 and X_2 be two random variables, and let a_1, a_2 and b be arbitrary constants.

(a) Prove that the expectation of a linear combination of random variables is given by:
$E(a_1 X_1 + a_2 X_2 + b) = a_1 E(X_1) + a_2 E(X_2) + b$.

(b) Derive the expression for the variance of a linear combination of two random variables, namely:
$Var(a_1 X_1 + a_2 X_2 + b) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + 2a_1 a_2\ Cov(X_1, X_2)$.

(c) Let X and Y be two continuous random variables with $E(X) = \mu_X$, $E(Y) = \mu_Y$, $Var(X) = \sigma_X^2$, $Var(Y) = \sigma_Y^2$ and $Cov(X,Y) = \sigma_{XY}$

(i) Find the expectation and variance of $Z = 3X + 2Y - 5$.

(ii) If X and Y are independent, simplify the $Var(Z)$.

(iii) If $Var(X) = 4, Var(Y) = 9$, and $Cov(X,Y) = 2$, then compute $Var(2X - Y)$.

Q3(a) The time (in minutes) required to complete an online order follows an exponential distribution with parameter λ. Let $X_1$and X_2 be the order processing times for two independent customers. Define: $T = X_1 + X_2$ as the total processing time for both customers.

(i) Find the cumulative distribution function (cdf) of T, $F_T(t)$.

(ii) Hence, obtain the probability density function (pdf) of T.

(b) Let X represent the amount of time (in minutes) a customer spends browsing in a shopping mall. Suppose that the mean and standard deviation of X are 30 minutes and 8 minutes, respectively.

(i) In a sample of 100 randomly selected customers, what is the approximate probability that the sample mean browsing time is between 29 and 31 minutes?

(ii) In a sample of 100 randomly selected customers, what is the approximate probability that the total time spent by all customers exceeds 3100 minutes?

(c) Define minimum variance unbiased estimator (MVUE) of θ. Prove that if $X_1, \dots, X_n$ be a random sample from a normal distribution with parameters μ and σ. Then the estimator $\hat{\mu} = \bar{X}$ is the MVUE for μ.

Q4(a) (i) State the additive principle of Fisher information.

(ii) Let $X_1, X_2, \dots, X_n$ be a random sample from a Poisson distribution with parameter λ, with $f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,2, \dots$ .Find the Fisher information $I_n(\lambda)$ using the additive principle.

(b) A cybersecurity study tested whether users could correctly identify fake websites. Out of 400 participants, 180 incorrectly identified a phishing website as safe. Let p denote the true proportion of users who would make this mistake. Construct a 95% confidence interval for p using the score interval method.

(c) (i) State the formula for the approximate $100(1-\alpha)\%$confidence interval for a population proportion $p$using the score (Wilson) method.

(ii) A study on online security tested whether users can correctly identify phishing emails. Out of 250 users, 95 users failed to identify a phishing email correctly. Let p denote the true proportion of users who would fail. Construct a 95% confidence interval for p using the score (Wilson) method.

Q5 (a) The heights of 10 males of a given locality are found to be 70, 67, 62, 68, 64, 61, 68, 70, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches? Test at 5% significance level, assuming that for 9 degrees of freedom, P (t > 1.83) = 0.05.

(b) An automobile company is testing whether a new engine design has increased the average fuel efficiency beyond 25 mpg. The standard deviation is known to be 3 mpg. A sample of 36 vehicles is tested. The company will conclude improvement only if the sample mean exceeds 26 mpg.

(i) Formulate the null and alternative hypotheses.

(ii) Explain Type I error in this context.

(iii) Find Power of the test.

(c) A survey claims that 55% of people prefer digital payments. In a sample of 180 people, 90 people prefer digital payments. Test whether the claim is valid.

Q6 (a) Given the regression model: $Y = 25 + 2x + e, e \sim N(0,9)$
(i) Find $E(Y \mid x = 4)$
(ii) Find $Var(Y \mid x = 4)$
(iii) Find $P(Y > 35 \mid x = 4)$
(iv) Interpret the slope coefficient

(b) Given: $\bar{x} = 10,\ \bar{y} = 20,\ S_{xx} = 50,\ S_{xy} = 40,\ S_{yy} = 100$
(i) Find regression line.
(ii) Calculate Sum of Squared Errors(SSE).

(iii) Calculate R^2.

(c) A researcher claims that the distribution of a categorical variable follows a theoretical distribution with equal probabilities. The observed data is:

Category	P	Q	R	S
Observed(O)	22	28	30	20
Expected€	25	25	25	25

Test at the 1% level of significance whether the data fits the given distribution.