[This question paper contains 4 printed pages.]

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

Sr. No. of Question Paper : 8065 **L**

Unique Paper Code : 2353200019

Name of the Paper : Mathematical Statistics DSE

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

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 market has both an express checkout line and a super express checkout line. Let X_1 denote the number of customers in line at the express checkout at a particular time of day, and let X_2 denote the number of customers in line at the super express checkout at the same time. Suppose the joint PMF of X_1 and X_2 is as given in the following table:

$X_1 \setminus X_2$	0	1	2	3
0	.08	.07	.04	.00
1	.06	.15	.05	.04
2	.05	.04	.10	.06
3	.00	.03	.04	.07
4	.01	.01	.05	.06

(a) What is P $(X_1 = 1, X_2 = 1)$ and P $(X_1 = X_2)$? What is the probability that the total number of customers in the two lines is at least four?

P.T.O.

(b) Let A denote the event that there are at least two more customers in one line than in the other. Express A in terms of X_1 and X_2, and calculate the probability of this event.

(c) Determine the marginal pmf of X_1 and X_2. Hence, calculate the expected number of customers in line at the express checkout. By inspection of $P(X_1 = 4)$, $P(X_2 = 0)$, and $P(X_1 = 4, X_2 = 0)$, are X_1 and X_2 independent random variables? Explain your reasoning.

Q2 Let X and Y be two random variables.

(a) Define covariance and correlation between two random variables. Prove the following properties:

(i) $Cov(X,Y) = Cov(Y,X)$
(ii) $Cov(X,Y) = E(XY) - \mu_X \mu_Y$, where μ_X and μ_Y are the mean of random variables X and Y respectively.

(b) State and verify the law of total expectation, i.e., show that

$$E[X] = E(E[X \mid Y]),$$ where X and Y are continuous random variables.

(c) Show that $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$ and hence derive the formula for $Corr(X,Y)$.

Q3(a) A company has three branches, each employing two workers. The monthly sales figures (in thousands of units) for the six employees are given below:

Branch	1	1	2	2	3	3
Employee	1	2	3	4	5	6
Sales	12.5	14.0	13.2	15.1	11.8	14.4

Suppose two employees are randomly selected from the six (without replacement). Determine the sampling distribution of the sample mean $\bar{X}$.

(b) (i) State Central Limit Theorem.

(ii) A factory produces metal rods, and the length of each rod is a random variable with mean 25cm and standard deviation 4cm. If a random sample of $n = 64$ rods is selected, using Central Limit theorem, find the approximate probability that the sample mean length $\bar{X}$ lies between 24cm and 26cm.

(c) (i) Define the unbiased estimator of a parameter of θ.

(ii) Prove that the sample variance S^2 is an unbiased estimator of σ^2 for any population distribution.

Q4 (a) (i) Define Fisher Information for a parameter θ.

(ii) Let X be a Bernoulli random variable with parameter p, with

$$f(x;p) = p^x(1-p)^{1-x}, x = 0,1$$

Find the Fisher Information $I(p)$ for a single observation.

(b) A survey conducted by a market research firm in 2022 asked 1200 consumers how much they planned to spend on online shopping during a festive sale. The mean response was ₹5200 with a standard deviation of ₹6800. Construct an approximate 99% confidence interval for μ, the mean amount all consumers plan to spend with the assumption that the sample size is large enough for the normal approximation to be valid.

(c) Determine the following:

(i) The 99th percentile of the χ^2_6 distribution

(ii) The 1st percentile of the χ^2_6 distribution

(iii) $P(Y < 10.645 \text{ or } Y > 36.123)$, where $Y \sim \chi^2_{20}$.

Q5 (a) A manufacturer claims that the mean activation temperature of a sprinkler system is 130°F. A sample of 9 systems gives a mean temperature of 131.08°F. Assume that the population is normally distributed with standard deviation σ = 1.5°F. Test at the 1% level of significance whether the manufacturer's claim is valid.

(b) A pharmaceutical company claims that a medicine reduces recovery time to an average of 10 days. The standard deviation is known to be 2 days. A researcher suspects that the recovery time is actually longer. A sample of 36 patients is taken. The null hypothesis is rejected if the sample mean exceeds 11 days.

(i) State the null and alternative hypotheses.

(ii) Interpret Type I error in this context.

(iii) Interpret Type II error in this context.

(c) A company claims that the proportion of products passing quality control is 75%. In a sample of 160 products, 110 products pass the test. At the 1% level of significance, test whether the population proportion is as claimed.

Q6 (a) Given the regression model: $Y = 80 - 4x + e, e \sim N(0,36)$
(i) Find $E(Y \mid x = 5)$
(ii) Find standard deviation
(iii) Find $P(Y > 70 \mid x = 5)$

(b) Given: $\bar{x} = 10$, $\bar{y} = 15$, $S_{xx} = 50$, $S_{xy} = 40$, $S_{yy} = 90$
(i) Find regression line.
(ii) Predict Y at $x = 12$.
(iii) Find R^2.

(c) A company records the preferences of customers for three colors. The observed and expected frequencies are given below:

Colour	Red	Blue	Green
Observed (O)	40	35	25
Expected (E)	33	33	34

Test at the 5% level of significance whether the observed data fits the expected distribution.