I.E.S./ISS-09

№ 000041

C-JTT-J-TA

STATISTICS – I

Time Allowed : Three Hours | *Maximum Marks : 200*

INSTRUCTIONS

Candidates should attempt FIVE questions in ALL including Questions No. 1 and 5 which are compulsory. The remaining THREE questions should be answered by choosing at least ONE question each from Section A and Section B.

The number of marks carried by each question is indicated against each.

Answers must be written only in ENGLISH.

(Symbols and abbreviations are as usual)

Any essential data assumed by candidates for answering questions must be clearly stated.

SECTION A

1. Answer any *five* of the following : *5×8=40*

(a) Prove that for $n \geq 2$,

$$\sum_{j=1}^{n} P\left[A_j\right] \geq P\left[\bigcup_{j=1}^{n}\right] \geq \sum_{j=1} P\left[A_j\right] - \sum_{1 \leq j \leq k \leq n} P\left[A_j \cap A_k\right]$$

(b) A die is rolled twice. Let A, B, C denote the events respectively that the sum of scores is 6, the sum of scores is 7, and the first score is 4. Are A and C independent ? Are B and C independent ?

(c) Given the distribution function

$$F(x) = \frac{1}{2} + \frac{x}{2(1+|x|)} \qquad -\infty < x < \infty$$

find its probability density function.

(d) Let X have the probability density function

$$f(x) = \begin{cases} (\alpha+1)x^{\alpha}, & 0 < x < 1,\ \alpha > 1 \\ 0 & , \text{otherwise.} \end{cases}$$

Find the mean, geometric mean and harmonic mean.

(e) (i) Let X be a random variable such that $P[X < 0] = 0$ and $E[x]$ exist. Show that $P(X \leq 2E[x]) \geq \frac{1}{2}$.

(ii) Let $E[X] = 0$ and $E[X^2]$ be finite. Show that $P(X^2 \leq 9E[X^2]) > \frac{8}{9}$.

(f) Let $\{X_n\}$ be a sequence of random variables such that

$$P(X_n = 1) = \frac{1}{n} = 1 - P(X_n = 0),\ n \geq 1.$$

Show that $X_n \to 0$ *in prob.* and $X_n \nrightarrow 0$ *a.s.*

2. (a) Given the joint density of (X_1, X_2),

$$f(x_1, x_2) = \begin{cases} \frac{1}{8}\left(x_1^2 - x_2^2\right) e^{-x_1}, & 0 < x_1 < \infty, \ |x_2| < x_1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the marginal densities of X_1 and X_2. Also find $E[X_1]$.

(b) Suppose that the random variable X has a normal distribution with mean μ and variance σ^2. Let ϕ be the distribution function of a standard normal variate. Find the density of $\phi\left(\frac{X-\mu}{\sigma}\right)$.

Also find $E\left[\phi\left(\frac{X-\mu}{\sigma}\right)\right]$.

(c) Find the generating function of X whose probability density function is

$P[X = r] = pq^{r-1}$, $r = 1, 2, \ldots, 0 \le p \le 1, q = 1 - p$.

(d) Explain "Memoryless property" of a distribution. Show that the exponential distribution has memoryless property. *4×10=40*

3. (a) Let

$$P_{X,Y}(u, v) = e^{-\lambda_1 - \lambda_2 - \mu + \lambda_1 u + \lambda_2 v + \mu uv}$$

be the probability generating function of the random variables X and Y.

(i) Find the marginal densities of X and Y.

(ii) Obtain the generating function of $Z = X + Y$.

(iii) Interpret the case when $\mu = 0$.

(b) Let $X \sim BIN(100, 0{\cdot}2)$. Compute $P[10 \leq X \leq 30]$.

(c) The joint density of (X, Y) is

$$f(x, y) = \begin{cases} x^2 + \dfrac{xy}{3}, & 0 < x < 1, \quad 0 < y < 2 \\ 0, & \text{otherwise.} \end{cases}$$

Find the conditional densities and $E[X \mid Y = 1{\cdot}5]$.

(d) Find the density, if its characteristic function is

$$\phi(t) = \begin{cases} 1 - |t|, & |t| < 1 \\ 0, & \text{otherwise.} \end{cases}$$

4×10=40

4. (a) Let $X_1, X_2, \ldots, X_n$ be a sequence of *iid* $N(0, 1)$ variables. Find the limiting distribution of

$$\frac{\left(X_1 + X_2 + \ldots + X_n\right)}{X_1^2 + X_2^2 + \ldots + X_n^2}$$

(b) Show that

$$\sum_{k=0}^{n} \frac{e^{-n}\, n^k}{k!} \to \frac{1}{2}, \quad \text{as } n \to \infty$$

(c) Let $\{X_n\}$ be *iid* with density function

$$f_n(x) = \begin{cases} \dfrac{1}{|x|^5}, & |x| > 1 \\ 0, & \text{otherwise.} \end{cases}$$

Examine whether Law of Large Numbers holds good for $\{X_n\}$.

(d) If the exponent of a bivariate normal density function is

$$-\frac{2}{3}[x^2 + 9y^2 - 3xy - 13x + 60y + 103],$$

find its means, variances, conditional means, conditional variances and correlation coefficient.

4×10=40

SECTION B

5. Answer any ***five*** of the following : *5×8=40*

(a) Explain the advantages of fitting orthogonal polynomials to a given paired data over the other methods.

(b) Show that the mean and variance computed for a random sample drawn from a normal distribution are independent.

(c) The data on lives of two models of refrigerators are given below :

Life	Number of Model A	Number of Model B
0 – 2	5	2
2 – 4	16	7
4 – 6	13	12
6 – 8	7	19
8 – 10	5	9
10 – 12	4	1

Which model is more reliable ?

(d) Two drugs were given to two batches of 6 students. The numbers of days to get a complete cure are given below :

Drug A	6	7	8	9	12	16
Drug B	10	11	13	14	15	17

Using Mann – Whitney test decide whether the median days for cure by the two drugs are equal.

(Table values of U – Statistic at 0·05 level are :

$U_{5,5} = 2$, $U_{5,6} = 3$, and $U_{6,6} = 5$.)

(e) Given the following data :

x :	0	1	3
f(x) :	1	3	55

find a polynomial P(x) of degree 2 or less so that P(x) = f(x) at the tabulated values of x. Hence approximate f(2).

(f) A train travels y(t) kilometers at time t = 0(1)6. Using the following table of values for (t, y(t)), compute the speed and acceleration of the train at t = 1.

t :	0	1	2	3	4	5	6
y(t) :	0	1	8	27	64	125	216

6. (a) Let the correlation coefficient between X_i and X_j be ρ for $i, j = 1, 2, \ldots, n$, $i \neq j$. Show that the square of the multiple correlation coefficient satisfies

$$R^2 = \frac{(p-1)\rho^2}{1+(p-1)\rho}.$$

(b) A manufacturer of alkaline batteries expects that only 5% of his products are defective. A random sample of 300 batteries contained 10 defectives. Can we conclude the proportion of defectives in the entire lot is less than 0·5 at 5% level of significance ?

(c) The two regression lines between X and Y are $8X - 10Y + 66 = 0$, $40X - 18Y = 214$. The variance of X is 9. Find $\overline{X}$, $\overline{Y}$, σ_Y, and ρ.

(d) Explain the method of testing normality by using chi-squared test. *4×10=40*

7. (a) If $Cov(X_i, X_j) = r_{ij}$, $i, j = 1, 2, 3$, $i \neq j$, then show that

(i) $r_{12} + r_{23} + r_{13} \geq -\dfrac{3}{2}$

(ii) $r_{12}^2 + r_{23}^2 + r_{13}^2 \leq 1 + 2\, r_{12}\, r_{23}\, r_{13}$

(b) (i) Show that the arithmetic mean of (positive) regression coefficients is greater than the correlation coefficient.

(ii) What is the value of the product of geometric mean of variances and the geometric mean of regression coefficients ?

(c) Define 'minimum mean square error' and 'best linear predictor'. Show that they coincide for the normal distribution.

(d) Two set of students were given different teaching methods. Their IQ's are given below :

Set I	77	74	82	73	87	69	66	80
Set II	72	68	76	68	84	68	61	76

Test whether the two teaching methods differ significantly at 5% level of significance.

(Assume critical value of test statistic to be 1·96) *4×10=40*

8. (a) A sample of 100 records on lengths of stay of patients in a hospital gave a standard deviation of days of stay as 4·9. In order to estimate the mean number of days of stay within 0·25 day with 95% confidence, what should be the sample size ?

(b) Find the value of

$$\int_0^1 \frac{1}{1+x^2}$$

by taking 5 subintervals and using the Trapezoidal rule.

(c) Define correlation ratio η. Show that

$$0 \le \rho^2 \le \eta^2 \le 1.$$

(d) Ten experts have given scores of effectiveness on two teams A and B as given below.

Expert	Score on A	Score on B
1	7	9
2	4	5
3	8	8
4	9	8
5	3	6
6	6	10
7	8	9
8	10	8
9	9	4
10	5	9

Are the distribution of scores for group A above that of group B ? *4×10=40*