Sl. No. C–JTT–M–TUA

STATISTICS—I

Time Allowed : Three Hours	Maximum Marks : 200

INSTRUCTIONS

Candidates should attempt FIVE questions in ALL including Question Nos. 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, unless otherwise indicated.)

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

A Normal Distribution Table and a 't' Table are attached with this question paper.

All parts and sub-parts of a question being attempted must be completed before moving on to the next question.

Section—A

1. *(a)* You are given the following information :

(i) In random testing, you test positive for a disease.

(ii) In 5% of cases, the test shows positive even when the subject does not have the disease.

(iii) In the population at large, one person in 1000 has the disease.

What is the conditional probability that you have the disease given that you have been tested positive, assuming that if someone has the disease, he will test positive with probability 1?

(b) Items from a large lot are examined one by one until r items with a rare manufacturing defect are found. The proportion of items with this type of defect in the lot is known to be p. Let X denote the number of items needed to be examined. Derive the probability distribution of X, and find $E(X)$.

(c) A 4-digit number is formed by selecting 4 digits from the set $(0, 1, 2, \cdots, 9)$ at random, 0 at the left-most position(s) being permissible. Find the probabilities that—

(i) all 4 digits will be alike;

(ii) 3 will be alike, 1 different;

(iii) there will be 2 pairs of identical digits;

(iv) 2 will be alike, 2 different;

(v) all 4 will be different.

(d) 12·3% of the candidates in a public examination score at least 70%, while another 6·3% score at most 30%. Assuming the underlying distribution to be normal, estimate the percentage of candidates scoring 80% or more.

(e) Let $\{X_n\}$ be a sequence of random variables with

$$P\{X_n = -2^n\} = P\{X_n = +2^n\} = \frac{1}{2},\ n = 1, 2, \cdots$$

Examine if the sequence obeys Weak Law of large numbers and Central Limit Theorem. 8×5=40

2. (a) Of three independent events A, B and C, A only happens with probability $\frac{1}{4}$, B only happens with probability $\frac{1}{8}$ and C only happens with probability $\frac{1}{12}$. Find the probability that at least one of these three events happens.

(b) For the Cauchy distribution given by

$$f(x) = \frac{k}{\sigma^2 + (x-\mu)^2},\ -\infty < x < \infty$$

where k is a constant to be suitably chosen, derive the expression for the distribution function. Hence obtain a measure of central tendency and a measure of dispersion. What are the points of inflexion of the distribution?

(c) The ith box contains $2i$ white balls and $6-2i$ black balls, $i = 1(1)3$. A fair die is cast once. 3 balls are taken at random from box 1, box 2 or box 3 according as the die shows up face 1, any of 2 and 3, or any of 4, 5 and 6, respectively. Let X denote the number of white balls drawn. Find $E(X)$.

(d) Write down the probability mass function of geometric distribution. State and prove its 'lack of memory property'. Find also the mean and the variance of the distribution. 10×4=40

3. (a) Two persons Amal and Bimal come to the club at random points of time between 6 p.m. and 7 p.m., and each stays for 10 minutes. What is the chance that they will meet?

(b) Show that in a sequence of $2s$ Bernoullian trials with success probability $p = \frac{1}{2}$, the most probable number of successes is s, and the corresponding probability is

$$p_s = \frac{1 \cdot 3 \cdot 5 \cdot \cdots 2s-1}{2 \cdot 4 \cdot 6 \cdot \cdots 2s} < \frac{1}{\sqrt{2s+1}}$$

(c) Find $E(X)$ and $V(X)$ for the random variable X having the probability density function

$$f(x) = \begin{cases} c \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} & , \ \mu - k\sigma \le x \le \mu + k\sigma \\ 0 & , \ \text{otherwise} \end{cases}$$

where c is a constant to be suitably chosen.

(d) Let X be a random variable with $F(x)$ as the distribution function. m is a suitable measure of central tendency and M is the mean square error about m. Show that

$$F(m+c) - F(m-c) \ge 1 - M/c^2 \quad \forall c > 0$$

10×4=40

4. (a) An urn contains a white and b black balls. After a ball is drawn at random, it is to be returned to the urn if it is white; if it is black, it is to be replaced by a white ball from another urn. What is the probability of drawing a white ball from the urn after the foregoing operation has been repeated n times?

(b) For a discrete random variable X assuming the values $0, 1, 2, \cdots$, the probability mass function satisfies the recursion relation

$$f(x) = \frac{\alpha + \beta x}{x} f(x-1), \ x = 1, 2, \cdots$$

Find the mean and the mean deviation about the mean of X.

(c) Let $X_1, X_2, \cdots, X_n$ be random variables such that $E(X_1) = \mu$, $E(X_r \mid X_{r-1}) = X_{r-1}$, $r = 2(1)n$. Also

$$E(X_1 - \mu)^2 = E[(X_r - X_{r-1})^2 \mid X_{r-1}] = \sigma^2, \; r = 2(1)n$$

Find $E(X_n)$ and $V(X_n)$.

(d) The continuous random variable X has the distribution function $F(x)$. X_1 and X_2 are two randomly selected values of X. Show that

$$E|X_1 - X_2| = 4\int_{-\infty}^{\infty} x\left[F(x) - \frac{1}{2}\right] dF(x)$$

10×4=40

Section—B

5. (a) The three sides x_1, x_2 and x_3 (in decimeter) of a solid rectangular parallelopiped satisfy the relation $x_1 + 2x_2 + 4x_3 = 12$. Exploiting the inequality relation between two well-known measures of central tendency, determine the maximum possible volume of this parallelopiped.

(b) In respect of the following observations, arranged in a non-decreasing sequence

8, 10, 10, x, 12, 14, 14, 16

where x is unknown, indicate, with a brief justification, if the following statements are correct :

(i) x is necessarily larger than the standard deviation of the entire set including x.

(ii) The mean deviation about the median is necessarily smaller than that about the mean.

(c) Solve the linear difference equation

$$u_n = pu_{n+1} + qu_{n-1}$$

with $p+q=1$ and initial conditions $u_0 = 0$ and $u_{a+b} = 1$.

(d) (i) In a ticket counter, at some point of time the sequence of males (M) and females (F) was found as

MMFMFFFMFMMFFFFMFMMMM

Use runs test to examine if the sequence is random (5% critical value of the number r of runs with $n_1 = 11$, $n_2 = 10$ is 6).

(ii) Name two non-parametric tests for comparing locations of two correlated populations.

(e) 125 out of 285 college-going male students in 1995 from a city were found to be smokers. Another sample of 325 such students from the same city in 2012 included 95 smokers. Examine at 5% level of significance if smoking habit among college-going students is on the decrease in this city. 8×5=40

6. *(a)* Show, by proving all intermediate results, that for a set of n unsorted data, the measure of skewness based on mean, median and standard deviation, necessarily lies between -3 and $+3$.

(b) Scores in a UPSC examination in Statistics Paper—I and Paper—II awarded to 10 candidates were as under :

Candidate	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Score in Paper—I	91	65	80	120	80	140	105	80	42	72
Score in Paper—II	86	38	75	135	85	135	94	101	45	65

Calculate Spearman's rank correlation coefficient between the scores in the two papers.

(c) Solve the equation $f(x) = 0$ by using a suitable interpolation formula on the following values :

x	3	4	5	6
$f(x)$	$-2 \cdot 8$	$-1 \cdot 2$	$-0 \cdot 3$	$1 \cdot 8$

(d) Let (X, Y) follow the bivariate normal distribution $N_2(0, 0, 1, 1, p)$. Show that the expected value of the absolute difference between X and Y is $2\sqrt{(1-p)/\pi}$.

10×4=40

7. *(a)* Consider a 3-point symmetrical distribution having the values $x_0 - h$, x_0, $x_0 + h$ with the corresponding relative frequencies f, $1-2f$ and f. Calculate the coefficient of kurtosis b_2 and find its limiting values as $f \to 0$ and $f \to \frac{1}{2}$. Hence comment on the suitability of b_2 as a measure of unimodality versus bimodality.

(b) For a set of 10 pairs of observations (x_i, y_i), $i = 1(1)10$, the following calculations are available :

$$\Sigma x_i = 61,\ \Sigma x_i^2 = 412,\ \Sigma y_i = 45,\ \Sigma y_i^2 = 308$$
$$\text{and}\quad \Sigma x_i y_i = 295$$

Examine at 5% level of significance if the two variables are uncorrelated in the population.

(c) An unknown function u_x has been tabulated below for some selected values of x. Use Newton's divided difference formula on these to find an approximate value of u_3 :

x	0	2	5	10
u_x	3	19	73	223

(d) Define partial correlation coefficient $r_{12\cdot3}$ between x_1 and x_2 eliminating from each the effect of x_3. Derive the expression for $r_{12\cdot3}$ and comment on its usefulness in multivariate data analysis. 10×4=40

8. *(a)* Mention two measures—one of skewness and the other of kurtosis—based on central moments. State and prove an inequality relation involving these two measures and a constant term. Give an example of a distribution for which equality holds in this relation.

(b) The speed y (in km/hr) of a car at different points of time x between 10:00 a.m. and 10:40 a.m. on some day was recorded as follows :

Time x (a.m.)	10:00	10:10	10:20	10:30	10:40
Speed y (in km/hr)	24·2	35·0	41·3	42·8	39·2

Calculate the approximate distance covered by the car between 10:00 a.m. and 10:40 a.m. on that day using Simpson's one-third formula for numerical integration.

(c) Indicate how you would test the hypothesis that the means of k independent normal populations are identical, clearly mentioning the null and the alternative hypotheses, the assumptions made, the test statistic used, and the critical region.

(d) Ten short-distance runners were put to a rigorous training for two months. Times taken by them to clear 100 metres before and after the training were as follows :

Sl. no. of runner	1	2	3	4	5	6	7	8	9	10
Time (in sec) *before training*	10·6	10·9	10·1	10·5	11·0	11·2	10·7	10·2	10·9	10·6
Time (in sec) *after training*	10·1	10·7	9·9	10·0	11·1	10·9	10·6	10·3	10·5	10·8

Use Wilcoxon's paired sample signed rank test to examine at 1% level if the training was at all effective. [The critical value of Wilcoxon's statistic at 1% level of significance for $n = 10$ is 5] 10×4=40

★ ★ ★