

STATISTICS

(Paper I)

Time Allowed : Three Hours
Maximum Marks : 300

QUESTION PAPER SPECIFIC INSTRUCTIONS

(Please read each of the following instructions carefully before attempting questions)

There are EIGHT questions divided in two Sections.

Candidate has to attempt FIVE questions in all.

Question Nos. 1 and 5 are compulsory and out of the remaining, any THREE are to be attempted choosing at least ONE question from each Section.

The number of marks carried by a question/part is indicated against it.

Wherever any assumptions are made for answering a question, they must be clearly indicated.

Word limit in questions, wherever specified, should be adhered to.

SECTION 'A'

1. (a) Of the under-5 children in a locality, 20% are malnourished but not immunised, while 40% are immunised but not malnourished. If 30% are neither immunised nor malnourished, what percentage of the under-5 children of the locality are malnourished? 12

- (b) For a discrete random variable X

$$P\{X > r\} = (1 - 1/n)(1 - 2/n) \dots (1 - r/n), \quad r = 1, 2, \dots$$

Find the probability mass function of X . 12

- (c) Suppose every packet of the biscuit NICE contains a coupon bearing one of the letters N, I, C or E with equal probability. A customer who has all the four letters, wins a free packet. A housewife buys 8 such packets. What is the probability that she wins (i) one free packet and (ii) two free packets? 12

- (d) Scores of 10 candidates of Civil Service Main Examination of a recent year in Statistics Papers I and II are as under:

Serial No. of Candidates	1	2	3	4	5	6	7	8	9	10
Score in Statistics I	98	172	58	106	185	89	140	168	102	202
Score in Statistics II	69	123	65	120	131	92	96	120	75	174

Use sign test to verify the hypothesis that candidates in the above examination offering both the Papers of Statistics scored higher on an average in Paper I than in Paper II. 12

- (e) Obtain a sufficient statistic for θ in sampling from

$$f(x; \theta) = \{[\theta^x(1-\theta)^{1-x}], \quad x=0,1; 0, \text{ otherwise}\}$$

Check if it is complete sufficient. 12

2. (a) A random variable X has the probability density function

$$f(x) = \{[1/2e^{-|x|}], \quad -\infty < x < \infty; 0, \text{ otherwise}\}$$

Derive the expression for its characteristic function. 20

- (b) State (without proof) the inversion theorem relating to the characteristic function of a continuous random variable. Use this theorem to obtain the probability density function of a random variable having the characteristic function $\phi(t) = e^{-|t|}$. 20

- (c) The random variable X follows uniform distribution in $(0, \theta)$. Let $X_{(1)}, X_{(2)}, X_{(3)}$ be the order statistics in a random sample of size 3 from the population. Show that both $2X_{(2)}$ and $4X_{(3)}/3$ are unbiased for θ .

20

3. (a) A random sample $(X_1, X_2, \dots, X_n)$ is taken from the population with the probability density function

$$f(x) = \begin{cases} kxe^{-x/\theta}, & x > 0, \theta > 0; \\ 0, & \text{otherwise} \end{cases}$$

where k is a constant to be suitably chosen. Derive the maximum likelihood estimator of θ . Also show that

$$S^2 = 1/3n \sum_{i=1}^n X_i^2$$

is an unbiased estimator of the population variance.

20

- (b) It is desired to test if average percentage of family income spent on food is the same in k selected blocks of a district. Assuming the underlying populations to be normal and homoscedastic, indicate how you would use likelihood ratio test for the purpose. Show that the resultant test boils down to the usual F -test.

20

- (c) The joint probability density function of X, Y is given by

$$f(x, y) = \begin{cases} [x(x+y/3)], & 0 < x < 1, 0 < y < 2; \\ 0, & \text{otherwise} \end{cases}$$

Compute (i) $P(X > 1/2)$, (ii) $P(Y < X)$ and (iii) $P(Y < 1/2 | X < 1/2)$. Also derive the regression equation of Y on X .

20

4. (a) For the sequence of random variables $\{X_i\}$

$$P(X_i = 2^i) = P(X_i = -2^i) = 2^{-(2i+1)}$$

and

$$P(X_i = 0) = 1 - 2^{-2i}, i = 1, 2, \dots$$

Check if the sequence obeys the weak law of large numbers and the central limit theorem.

20

- (b) Define Wald's SPRT of strength (α, β) for a simple hypothesis against a simple alternative. Derive the SPRT procedure of strength (α, β) for $H_0: \mu = \mu_0$ against $H_1: \mu = \mu_1$, based on successive observations from $N(\mu, \sigma^2)$, σ^2 known. Indicate how the test can be carried out graphically.

20

- (c) The manufacturers of a certain brand of electrical switches claim that at least 80% of the boxes, containing 12 such switches each, will have no defective. Design a randomised test of exact size 0.05 based on the number of boxes, among a random sample of 8 examined, containing at least one defective.

20

SECTION 'B'

5. (a) The p levels of a factor A are the only levels the experimenter is interested in, while the q levels of the factor B chosen form a random sample from the totality of levels available. Only one observation per cell $A_i \times B_j$, $i = 1(1)p$, $j = 1(1)q$, is taken on the response variable. Write down the appropriate linear model, the hypotheses to be tested, the test statistics and the level- α critical regions.

12

- (b) Assuming the model

$$[x_i = m_1 + e_{1i}; y_i = m_2 + e_{2i}; z_i = m_1 + m_2 + e_{3i}, i = 1(1)n]$$

where e_{1i} , e_{2i} and e_{3i} are the errors, obtain least square estimates of m_1 and m_2 .

12

- (c) 6 days are to be selected from the 366 calendar days of 2012 for recording the maximum daily temperature in a metropolitan city. Using the following set of random numbers, select these days, mentioning the month and the date for each, and giving the procedure in detail: 6503 0085 3822 2193 5392 4635 0495 3296 1348

12

- (d) Suggest the confounding scheme of a 2^4 -design to be carried out in 4 replicates each comprising 4 incomplete blocks of size 4, such that full information is retained on the main

12

effects, each of the first-order and second-order interactions is confounded only in one replicate, while the third-order interaction is confounded in two replicates. Give the complete layout of one such replicate indicating the factorial effects confounded in it.

- (e) Suppose, with usual notation $r_{12}=0.36$, $r_{13}=0.29$, $r_{14}=-0.62$, $r_{23}=0.41$, $r_{24}=-0.24$, $r_{34}=-0.52$. For predicting x_1 , the variable x_4 has already been included in the regression equation. Which one of x_2 and x_3 is worth including in the regression equation in addition to x_4 ? 12
6. (a) Give the outline of the procedure, using ANOVA technique, for testing if the response variable y is dependent on a fixed set of variables $x_1, x_2, \dots, x_k$, based on an observed set of data on these variables. 20
- (b) Determine the sample size n , such that the estimated value, based on this sample, of the mean SO_2 level in air in 1000 petrol pumps in a large town, differs from the true value by at most $5 \mu\text{g}/\text{m}^3$ with probability 0.95, assuming the standard deviation to be $35 \mu\text{g}/\text{m}^3$. [2.5% and 5% points of a standard normal variate are 1.960 and 1.645 respectively.] 20
- (c) Introduce the concepts of confounding, total confounding, partial confounding and balanced confounding in the context of a factorial design. A 3^2 -design has been conducted using 4 replicates each comprising 3 incomplete blocks of size 3 each. In two of the replicates, AB has been confounded, while in the remaining two, AB^2 has been confounded. Write down the ANOVA table. Indicate how the sums of squares due to different factorial effects will be calculated. 20
7. (a) Show that a necessary and sufficient condition for $X_{\sim p}$ to be distributed as $N_p(\mu, \Sigma)$ is that every linear function $Y=I'X_{\sim}$ is distributed as univariate normal $N(I'\mu_{\sim}, I'\Sigma I)$. 20
- (b) The total population size of a district in 2001 census is available. The population sizes of a random sample of n villages from the district individually in both 2001 and 2011 censuses are also available. Suggest a suitable estimator of the total population of the district in 2011, based on these figures using the ratio method of estimation. Derive an expression for the bias of the estimator and deduce the condition under which the bias is negligible. 20
- (c) Construct a pair of orthogonal Latin squares with symbols A, B, C and α, β, γ respectively. Superimpose one onto the other. Numbering the small squares from 1 to 9, use this arrangement to construct a balanced incomplete block design with parameters $v=9$, $b=12$, $r=4$, $k=3$ and $\lambda=1$. 20
8. (a) Does the observed correlation between two variables X_1 and X_2 always imply causal relationship between them? Give an example. Define partial correlation coefficient $r_{12.3}$ between X_1 and X_2 eliminating from each the effect of X_3 . Why is $r_{12.3}$ more suitable for throwing light on the net correlation between X_1 and X_2 ? Based on the observed data set x_{ij} , $i=1, 2, 3$ and $j=1(1)n$, derive the expression for $r_{12.3}$. 20
- (b) Derive the regression equation of X_1 on X_2 , when X_1 and X_2 jointly follow trinomial distribution with parameters n , P_1 and P_2 . Tube lights of a certain brand are graded as III, II or I according as they fail during the first 6000 hours and the next 6000 hours or survive the first 12000 hours respectively. 12 such new tube lights are put to life-testing simultaneously. 5 of them are found to be of grade III. How many of the remaining are expected to be of grade I? 20
- (c) In a planned town, there are N blocks each containing M residential buildings. A simple random sample of n blocks is drawn, and the numbers of working women in all the residential buildings belonging to these selected blocks are counted. Based on these figures, suggest an unbiased (to be shown) estimator of the total number of working women in the town, and derive the standard error of the estimator. 20