

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. Answer any FIVE parts of the following:

12×5=60

- (a) In a two-tier competitive examination, only candidates with a minimum score of x_0 in the preliminary examination are allowed to sit for the main examination. The score distribution in the preliminary examination can be assumed to be normal with mean μ and s.d. σ . Derive the expressions for the mean and the variance of scores in the preliminary examination of the candidates sitting for the main examination.
- (b) 2 sets of n cards, both numbered from 1 to n , are randomly matched. Find $E(X)$ and $V(X)$ where X is the number of matches realised.
- (c) State (without proof) Tchebychev's inequality. How many times a fair coin must be tossed in order that the relative proportion of heads lie between 0.4 and 0.6 with probability at least 0.9? (Upper 5% and 10% points of standard normal variate are 1.645 and 1.282 respectively.)
- (d) State the invariance property of maximum likelihood estimator (m.l.e.). Use this to obtain the m.l.e. of $1/\theta$ in sampling from
- $$f(x; \theta) = \begin{cases} \binom{m}{x} \theta^x (1-\theta)^{m-x}, & x=0(1)m; \\ 0, & \text{otherwise.} \end{cases}$$

- (e) What is a size- α randomised test? For testing the null hypothesis that at most 50% of the 500 ml. packs of hair oil of a certain brand are underweight, against the alternative that the percentage is more, it is desired to examine 8 such packs chosen at random. Use the following information to design a randomised test of size 0.05:

x	0	1	2	3	4	5	6	7	8
$f(x)$	·0039	·0312	·1094	·2188	·2934	·2188	·1094	·0312	·0039

where $f(x)$ is the probability mass function of a binomial variate with parameters $m=8$, $p=0.5$.

- (f) A group of 20 short distance runners were subjected to a month-long training. Discuss how you would examine if the training was at all effective based on their timings to clear 100 meters before and after the training, clearly mentioning the underlying assumptions, if any, the null and the alternative hypotheses, as also the size- α critical region.

2. (a) X_1 and X_2 are i.i.d. random variables with common probability mass function

$$f(x) = \begin{cases} [q^x p, & x=0,1,2,\dots, 0 < p < 1, q=1-p; \\ 0, & \text{otherwise.} \end{cases}$$

Obtain the probability distribution of $Y = \max(X_1, X_2)$. Also derive $E(Y)$.

20

- (b) In sampling from $N(\mu, \sigma^2)$, μ unknown, derive $100(1-\alpha)\%$ shortest confidence interval for σ^2 . 20
- (c) Discuss how you would graphically perform the sequential probability ratio test for a simple hypothesis regarding the mean of a normal population with known standard deviation against a simple alternative. 20
3. (a) Define probability generating function (pgf) of a discrete random variable. Let X be a random variable with p.m.f.
- $$P\{X=j\}=p_j, j=0,1,2,\dots$$
- Set
- $$P\{X>j\}=q_j, j=0,1,2,\dots$$
- and define $Q(s)=\sum_{j=0}^{\infty} q_j s^j$. Show that $Q(s)=[1-P(s)]/[1-s]$, $|s|<1$, where $P(s)$ is the pgf of X . 20
- (b) For the p.d.f.
- $$f(x;\theta)=\{[1/\pi\{1+(x-\theta)^2\}], -\infty<x<\infty; 0, \text{ elsewhere}\}$$
- derive Rao-Cramer lower bound of variance of an unbiased estimator of θ . 20
- (c) Show that the likelihood ratio test, for the hypothesis that the means of K independent normal populations with common variance are identical, boils down to F -test. 20
4. (a) The daily demand (in kg) of fish of a certain variety upto 12 noon with a retailer follows exponential law with mean 25. He makes a profit of Rs. 20 on an average for each kg he is able to sell by 12 noon, and incurs a loss of Rs. 10 on an average for each kg that remains unsold after 12 noon. What is the optimum daily stock he should procure? 20
- (b) For the Pareto distribution with distribution function
- $$F(x)=\{[1-(\theta/x)^\alpha], x\geq\theta; 0, \text{ otherwise}\}$$
- derive the p.d.f. Also obtain its harmonic mean and variance. 20
- (c) Define (i) most powerful test, and (ii) unbiased test of a simple hypothesis against a simple alternative. Show that a most powerful test is necessarily unbiased. 20

SECTION 'B'

5. Answer any FIVE of the following parts:–

12×5=60

- (a) A committee of 5 is to be formed from amongst 86 officers, numbered serially from 1 to 86. Draw a circular systematic sample for this purpose, giving the procedure you have followed in detail. You may use the following set of random numbers: 1349 0417 9311 9787 1284 0769 8422 1077.
- (b) Examine, after deducing necessary result, if the following set of correlation coefficients are internally consistent: $r_{12}=0.62$, $r_{13}=0.55$ and $r_{23}=-0.42$.
- (c) Locate 5 points at random on a rectangular paddy field measuring 30 m × 20 m for conducting a crop-cutting experiment. The points are to be located using rectangular co-ordinates to nearest decimeters with one corner of the plot as origin. The following set of random numbers may be used: 2292 2933 6125 2464 1038 3163 3569 7155 2029 2538 7080 3027 6215 3125 5856 9543
- (d) For pps sampling without replacement, give the Horvitz-Thompson estimator of the population total. Show that it is unbiased. Also obtain its variance.
- (e) Suggest a balanced confounded design of a 2^4 experiment in 3 replicates, each replicate containing 4 incomplete blocks of size 4 each, retaining full information on the main effects and the 2nd order interactions. Give the complete lay-out of one such replicate indicating the factorial effects confounded in it.

- (f) Define a BIBD. Show that for a BIBD with parameters v, b, r, k and λ , (i) $\lambda(v-1)=r(k-1)$ and (ii) $b \geq v+r-k$.
6. (a) The vector variable $X_{\sim}$ follows p -variate normal distribution with mean vector $\mu_{\sim}$ and dispersion matrix Σ . Obtain the distribution of $Q=(X_{\sim}-\mu_{\sim})'\Sigma^{-1}(X_{\sim}-\mu_{\sim})$. 20
- (b) Based on a set of data $(x_i, y_{ij}), j=1(1)n_i, i=1(1)p$, discuss how you would use the technique of Analysis of Variance for testing the hypothesis that the regression equation of y on x is linear. 20
- (c) Use the following set of random numbers to obtain the lay-out of a randomised block design with 5 treatments and 4 blocks, giving the outline of the procedure you have followed: 5711 7343 7539 3684 9397 5335 4031 1486 2588 3301 0553 2427 3598 2580 7017 9176 20
7. (a) The joint p.d.f. of X and Y is given by

$$f(x,y)=\left\{\begin{aligned} & \frac{n!}{(x! y! (n-x-y)!)} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}, \quad x,y=0(1)n, x+y \leq n, 0 < p_1, p_2 < 1, p_1+p_2 < 1; \\ & 0, \quad \text{otherwise.} \end{aligned}\right.$$

 Find the correlation coefficient between X and Y , as also the regression equation of Y on X . 20
- (b) A rural block in a district was divided into 3 strata. The following table gives some relevant information:
- | Stratum No. (h) | Number of Villages (N_h) | Area X_h under wheat (hectares) | S.D. S_{X_h} (hectares) |
|---------------------|------------------------------|-----------------------------------|---------------------------|
| 1 | 20 | 620 | 100.5 |
| 2 | 15 | 580 | 90.8 |
| 3 | 25 | 825 | 140.7 |
- A sample of 10 villages is to be drawn. How many villages should be selected from different strata using (i) proportional allocation, (ii) optimum allocation, and (iii) allocation proportional to area under crop? 20
- (c) 3 factors, each at 2 levels, are to be tested in a single experiment using r randomised blocks. Using standard notation, give the expressions for the sums of squares due to different factorial effects. Also give the ANOVA table. 20
8. (a) You are to estimate the proportion of school-going students (Classes V to XII) using mobile phones in a metropolitan city. Suggest a two-stage sampling scheme. Using a suitable notation, give the estimator to be used, and check if the estimator is unbiased. 20
- (b) Write down the random effects model for a set of two-way classified data with one observation per cell, stating the underlying assumptions completely. What are the hypotheses that can be tested using the ANOVA technique? Give the ANOVA table with explicit expressions for the different sums of squares, and the test statistics for the above hypotheses. 20
- (c) Compare completely Randomised Design, Randomised Block Design and Latin Square Design in terms of flexibility for number of treatments and number of replicates per treatment on the one hand, and error-control on the other. 20