1·F·S-2009

Serial No. 0995

B–JGT–J–TUA

STATISTICS

Paper—I

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

INSTRUCTIONS

Candidates should attempt Question Nos. 1 and 5 which are compulsory, and ***THREE*** *of the remaining questions, selecting at least* ***ONE*** *question from each Section.*

The number of marks carried by each question is indicated at the end of the question.

Answers must be written in ***ENGLISH.***

Assume suitable data, if considered necessary, and indicate the same clearly.

(Notations and symbols are as usual)

SECTION—A

1. Answer any **FOUR** parts :— [4×10=40]

(a) 5 coins with probability of head equal to $\frac{2}{3}$ each and 4 coins with probability of head equal to $\frac{2}{5}$ each are tossed simultaneously. Use the probability generating function of binomial distribution to find the probability of getting 3 heads altogether.

(b) State Weak Law of Large Numbers (WLLN). Examine if WLLN and Central Limit Theorem hold for the sequence of random variables $\{X_k\}$ where

$$P\{X_k = K\} = P\{X_k = -K\} = \frac{1}{2\sqrt{K}},$$

$$P\{X_k = 0\} = 1 - \frac{1}{\sqrt{K}}, \quad K = 1, 2, \ldots\ldots .$$

(c) Show that $\overline{X}^2$ is consistent for μ^2 where $\overline{X}$ is the mean of a random sample of size n from N (μ, σ^2).

(d) Design a randomised sign test of size 0.05, based on a random sample of size 5, for testing the null hypothesis that 50% of the new born babies in a hospital are females against the alternative that the percentage is less.

(e) Define a family of Uniformly Most Accurate Unbiased (UMAU) confidence sets of level $1-\alpha$ for a parameter θ.

Starting from the acceptance region of a suitable test for $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$ where σ^2 is the variance of a normal population with unknown mean, obtain the UMAU confidence set of level $1 - \alpha$.

2. (a) A discrete random variable assuming the values 0, 1, 2 ... has a probability mass function f(x) satisfying

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

where α and β are constants.

Find the arithmetic mean μ and the mean deviation about μ for the distribution.

Hence obtain the mean deviation about μ for the binomial distribution with parameters m and p.

(b) Derive the moment generating function of the negative binomial distribution, and use this to obtain its mean and variance.

(c) State and prove the lack of memory property of exponential distribution. For what type of failure rate can exponential probability law be used to model the underlying life distribution ?

(d) 50% of the mangoes of a certain variety in a large lot have weight 320 gms or more, while another 6.68% have weight of at least 380 gms. Assuming the underlying distribution to be normal, estimate the percentage of such mangoes weighing between 280 gms and 400 gms.

(You may use the information that for a standard normal variate X, $P\{X \le K\} = 0.8413$, 0.9332, and 0.9772 for $K = 1.0$, 1.5 and 2.0 respectively).

[10+10+10+10]

(Contd.)

3. (a) Define Mean Square Error (MSE) of an estimator T for a parameter θ.

Let $(X_1, X_2, \ldots, X_n)$ be a random sample from the population $N(\mu, \sigma^2)$, and $S^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2/n$ where $\overline{X} = \sum_{i=1}^{n} X_i/n$. Find a constant C such that CS^2 is the minimum MSE estimator for σ^2.

(b) A random sample

− 2·1, 6·4, −3·5, 2·8, 0, 4·5, −5·2, 5·8 is taken from the population with p.d.f.

$$f(x) = \begin{cases} \frac{1}{2\theta} & , -\theta \le x \le \theta \\ 0 & , \text{otherwise} \end{cases}$$

Find the maximum likelihood estimate of θ justifying your answer.

(c) In a large forest there are N number of trees of a certain species haphazardly located throughout the forest and numbered from 1 to N. A stranger, while roaming about in the forest one day, identifies n such trees with serial numbers $x_1, x_2, \ldots, x_n$.

Indicate how he would use these observed serial numbers to estimate N, unknown to him. Give an unbiased estimate of the variance of his estimate, when N is large.

(d) Use a suitable theorem, to be clearly stated, to obtain the unique minimum variance unbiased estimator for the Poisson parameter λ, based on a random sample from the population. [10+10+10+10]

4. (a) Random samples from 3 species of trees in a forest were examined for incidence of a certain disease. The following results emerged :

Species	*Developed the disease ?*		
	Yes	*No*	*Total*
A	18	65	83
B	13	91	104
C	25	49	74
Total	56	205	261

Test at 5% level of significance if the species behave differently in respect of proneness to the disease.

[You may use the following values of $\chi^2_{\cdot 05, n}$:

$\chi^2_{\cdot 05,1} = 3\cdot841$ $\chi^2_{\cdot 05,2} = 5\cdot991$ $\chi^2_{\cdot 05,5} = 11\cdot070$]

(b) Describe the median test for comparing the locations of two independent populations, clearly mentioning the underlying assumptions, if any, the precise hypothesis to be tested, the test-statistic and the critical region.

(c) Derive Sequential Probability Ratio Test for $H_0 : \mu = \mu_0$ against $H : \mu = \mu_1$ regarding the parameter μ in N (μ, 1). Indicate how the test can be carried out graphically.

(d) Define, using suitable notation, the size-α minimax test from amongst the rival tests $w_1, w_2, \ldots, w_k$ for $H_0 : \theta = \theta_0$ against $H : \theta > \theta_0$. [10+10+15+5]

SECTION—B

5. Attempt any **FOUR** parts :— [4×10=40]

(a) Use the following set of random numbers, from a random starting point, to draw a simple random sample of 5 trees without replacement from the 324 trees in a forest :

0912	8257	4015	5933	5520
6917	4650	6927	6472	7279
5364	9901	7710	3750	7940

(b) The relation between the dependent variable y, subject to some observational errors, and the independent variable x is known to be of the form :

$$y = ax^b.$$

Using a set of observed data (x_i, y_i), $i = 1, 2, \ldots, n$, indicate how you would estimate the above equation by the method of least squares. Derive the normal equations to be solved for estimating the parameters.

(c) Construct a Balanced Incomplete Block Design (BIBD) with parameters $v = b = 11$, $r = k = 5$, $\lambda = 2$. Indicate how another BIBD can be constructed from this design.

(d) Discuss how you would estimate the percentage of candidates in a public examination adopting unfair means, using randomised response technique. Derive the variance of the estimator.

(e) Define Wishart distribution. When does this reduce to chi-square distribution ? State its reproductive property.

6. (a) The vector variable $(X_1, X_2, X_3)'$ follows a trivariate normal distribution with mean vector $(49, 52, 46)'$ and dispersion matrix :

$$\begin{pmatrix} 64 & 28 & 10 \\ 28 & 49 & 14 \\ 10 & 14 & 25 \end{pmatrix}.$$

Find :

(i) the multiple correlation coefficient $\rho_{1\cdot 23}$, and

(ii) the conditional mean of X_1 given $X_2 = x_2$ and $X_3 = x_3$.

(b) Write down the Gauss-Markov linear model. When is a linear parametric function said to be estimable under the model ? State and prove a necessary and sufficient condition for a linear parametric function to be linearly estimable.

(c) Suggest a measure of efficacy of the least squares bivariate linear regression equation as a prediction formula, and establish its relationship with the product moment correlation coefficient between the variables under study. [15+15+10]

7. (a) Show that, subject to a certain assumption to be specified by you, $V_3 \leq V_2 \leq V_1$, where V_1, V_2 and V_3 denote the variances of estimators of the population mean using an unrestricted simple random sample, a stratified random sample with proportional allocation and a stratified random sample with optimum (Neyman) allocation, respectively, the total sample size being identical in each case.

(b) What is ratio method of estimation of the population mean ? Derive the variance of the usual ratio estimator of the population mean to the first order of approximation and deduce the condition under which this estimator is more efficient than a comparable mean per unit estimator.

(c) What is probability proportional to size sampling with replacement (ppswr) ? For ppswr sampling, give the Hansen-Hurwitz estimator of the population total. Show that the estimator is unbiased. Obtain the variance and the variance-estimator of this estimator.

[15+15+10]

8. (a) Write down the fixed-effects model for a 2-way classified data with m (≥ 2) observations per cell. Indicate how the total sum of squares can be decomposed into various components. What are the hypotheses that can be tested using these data ? Mention the test statistic for each of these hypotheses.

(b) Obtain the lay-out of a randomised block design for testing the treatments A, B, C, D and E each to be replicated 4 times, using the following set of random numbers and explaining the procedure used :

7474	6525	1350	3148	4576
0516	2710	3523	4508	8105
1341	6045	2877	0132	5507

(c) A 3^2 design is to be laid out in 3 incomplete blocks of size 3 each per replicate. The composition of one such incomplete block in a replicate, using standard notation, is :

00	11	22

Identify the factorial effect confounded, and give the compositions of the other two incomplete blocks of the replicate. If this basic pattern is repeated r times, write down the first two columns of the ANOVA table to be used for analysing the data.

(d) For testing v varieties using a $v \times v$ Latin square design, the crop of a corner plot is found to be destroyed by cattle. Discuss how you would proceed to validly analyse the data by appropriately estimating the yield of the missing plot. [10+10+10+10]