C-MNS-U-STC

I.F.S. EXAM-2016

STATISTICS

Paper – I

Time Allowed : Three Hours

Maximum Marks : 200

Question Paper Specific Instructions

Please read each of the following instructions carefully before attempting questions :

*There are **EIGHT** questions in all, out of which **FIVE** are to be attempted.*

*Questions no. **1** and **5** are compulsory. Out of the remaining **SIX** questions, **THREE** are to be attempted selecting at least **ONE** question from each of the two Sections A and B.*

Attempts of questions shall be counted in sequential order. Unless struck off, attempt of a question shall be counted even if attempted partly. Any page or portion of the page left blank in the Question-cum-Answer Booklet must be clearly struck off.

All questions carry equal marks. The number of marks carried by a question/part is indicated against it.

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

Unless otherwise mentioned, symbols and notations have their usual standard meanings.

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

SECTION A

Q1. (a) (i) Show that for arbitrary events $A_1, A_2, ..., A_n$,

$$P\left(\bigcup_{1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i).$$

(ii) If P_1 and P_2 are probability measures, prove that

$P = \alpha P_1 + (1 - \alpha) P_2$ is also a probability measure, $(0 \le \alpha \le 1)$. *8*

(b) Let (X_1, X_2) denote quantitative scores on test 1, and (Y_1, Y_2) be verbal scores on test 2. If $Cov(X_1, Y_1) = 5$, $Cov(X_1, Y_2) = 1$, $Cov(X_2, Y_1) = 2$ and $Cov(X_2, Y_2) = 8$, compute the covariance between total quantitative score $(X_1 + X_2)$ and total verbal score $(Y_1 + Y_2)$. *8*

(c) Derive the lower bound for the variance of an unbiased estimator of σ^2 of $N(\mu, \sigma^2)$. Hence show that $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$ does not attain this bound. Why is it still the best unbiased estimator ? *8*

(d) Obtain $100(1 - \alpha)\%$ confidence interval for the difference $(p_1 - p_2)$ of success probabilities of two independent Bernoulli distributions. *8*

(e) Show that the characteristic function (ch. fn.) of a random variable determines its distribution function (d.f.) uniquely. *8*

Q2. (a) If g is a continuous function and $X_n \xrightarrow{P} X$, then show that $g(X_n) \xrightarrow{P} g(X)$. *10*

(b) For the Pareto distribution with pdf

$$f(x \mid \alpha) = \frac{\alpha}{x_0} \left(\frac{x_0}{x_1} \right)^{1+\alpha}, x > 0, \alpha > 0,$$

show that the d.f. is $\left(1 - \frac{x_0}{x}\right)^{\alpha}$ and sketch it for $\alpha = \frac{1}{2}$ and $\alpha = 2$. Also show that Var(X) does not exist for $\alpha \le 2$. (Sketches to be shown on plain paper). *10*

(c) Given a random sample from

$$f(x \mid \theta) = \theta\, x^{\theta - 1} e^{-x^{\theta}}, x > 0, \theta > 0.$$

Show that the loglikelihood equation has a unique root and that it provides a strongly consistent estimator. *10*

(d) Let X have a Poisson distribution with mean θ. Assume that θ has a $\Gamma(p, \sigma)$ distribution, that is,

$$\pi(\theta) = \frac{\theta^{p-1} \sigma^p e^{-\sigma\theta}}{\Gamma(p)}, \theta > 0, \sigma > 0, p \ge 1.$$

Show that the posterior distribution given $(x_1, ..., x_n)$ is Gamma $(p + \sum_{1}^{n} x_i, \sigma + n)$. Hence obtain the Bayes estimator of θ with respect to a squared-error loss. *10*

Q3. (a) Derive the likelihood ratio test of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ given a random sample from

$$f(x \mid \theta) = \frac{1}{2} \exp\{-\frac{1}{2}(x-\theta)\},\ x > \theta,\ \theta > 0, \text{ considering } \alpha = 0{\cdot}05.$$ *10*

(b) Consider a random sample $X_1, X_2, \ldots, X_n$ from $U(0, \theta)$. Show that the range $R = X_{(n)} - X_{(1)}$ is an ancillary statistic. *10*

(c) Obtain the level-α UMPU test of $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$ of $N(\mu, \sigma^2)$ and its power function (μ is unknown). *10*

(d) Using an appropriate pivotal quantity, derive a $100(1-\alpha)\%$ confidence interval for μ of $N(\mu, \sigma^2)$, where σ^2 is unknown. Is the C.I. also UMAU ? *10*

Q4. (a) Briefly explain the implementation of SPRT and state its optimal property. Illustrate it for testing $\theta = \frac{1}{3}$ versus $\theta = \frac{2}{3}$ for the distribution

$$f(x \mid \theta) = \theta^{\frac{1-x}{2}} (1-\theta)^{\frac{1+x}{2}},\ x = -1, 1.$$ *10*

(b) A fair coin is tossed n times and S_n, the sum of number of heads, is noted. Use it to find a lower bound for the probability that $\frac{S_n}{n}$ differs from $\frac{1}{2}$ by less than 0·1 when n = 100, and show that $\frac{S_n}{n} \xrightarrow{P} \frac{1}{2}$ as $n \to \infty$. *10*

(c) Stating the hypothesis, explain the two-sample Wilcoxon-Mann-Whitney test and derive the mean of the test statistic. *10*

(d) Define convergence in distribution. Show that the maximum of a random sample from $U(0, \theta)$ converges to an exponential distribution. *10*

SECTION B

Q5. (a) Given the Gauss-Markov linear model

$(\mathbf{Y}, A\boldsymbol{\theta}, \sigma^2 I)$ with $E(Y_1) = \theta_1 + \theta_2 - \theta_3$,

$E(Y_2) = 2\theta_1 - \theta_2 - \theta_3$, $E(Y_3) = \theta_1 - 2\theta_2$ and

$E(Y_4) = 3\theta_1 - 2\theta_3$. Find a necessary and sufficient condition for

$\mathbf{a}'\boldsymbol{\theta} = a_1\theta_1 + a_2\theta_2 + a_3\theta_3$ to be estimable. Hence examine the estimability of $\theta_1 + \theta_2 - 2\theta_3$. *8*

(b) Given $\mathbf{X} \sim N_3 \left(\mu' = (2, 4, 3), \sum = \begin{bmatrix} 8 & 2 & 3 \\ 2 & 4 & 1 \\ 3 & 1 & 3 \end{bmatrix}\right)$,

(i) find the regression function of X_1 on X_2 and X_3, and

(ii) compute the conditional variance of X_1 given X_2 and X_3. *8*

(c) Define connectedness of a block design. Examine if the following design is connected : *8*

$B_1 : (T_1, T_2, T_3)$, $B_2 : (T_3, T_4, T_5)$, $B_3 : (T_3, T_4)$,

$B_4 : (T_1, T_2, T_2)$, $B_5 : (T_4, T_4, T_5)$.

(d) Describe two-stage sampling with a real life example (stating clearly the first and second stage units) and compare it with cluster sampling. *8*

(e) What is the importance of confounding in factorial experiments ? Give a layout of a 2^3 factorial experiment with 2 replicates so as to confound partially the interaction effects ABC and BC. *8*

Q6. (a) Define multiple and partial correlation coefficients. Suppose that the correlation matrix of $(X_1, X_2, X_3, X_4)'$ is

$$R = \begin{bmatrix} 1 & 0{\cdot}6 & -0{\cdot}36 & -0{\cdot}53 \\ & 1 & -0{\cdot}46 & -0{\cdot}37 \\ & & 1 & 0{\cdot}37 \\ & & & 1 \end{bmatrix}.$$

Find

(i) the multiple correlation coefficient between X_1 and the other 3 variables, and

(ii) the partial correlation coefficient $\rho_{X_1X_2 \cdot X_3X_4}$. *10*

(b) Suppose that Y_i, i = 1, ..., 6 is uncorrelated with common variance σ^2 and $E(Y_1) = \beta_1 = E(Y_2)$, $E(Y_3) = \beta_2 = E(Y_4)$ and $E(Y_5) = \beta_1 + \beta_2 = E(Y_6)$. Obtain four mutually orthogonal contrasts of (Y_i) belonging to the error space. Find BLUE of β_1 and β_2. *10*

(c) If **Q** denotes the vector of adjusted treatment totals in a block design, find (i) **Q′J** (J is a vector of 1's), (ii) E (**Q**), and (iii) covariance matrix of **Q**. *10*

(d) Define a regression estimator of population mean. Show that this estimator is more efficient than the sample mean under SRSWOR using large sample approximation. *10*

7. (a) Discuss the main effects and interaction effects of a 2^3 factorial experiment as orthogonal contrasts. Also describe how to compute the sum of squares due to these contrasts. *10*

(b) Explain the concept and utility of principal components. Find the first principal component and its variance of the random vector with covariance matrix $\sum = \begin{bmatrix} 9 & 3 & 3 \\ & 9 & 3 \\ & & 9 \end{bmatrix}$. *10*

(c) Discuss the need for PPS Sampling illustrating with an example from real life. Define Horvitz-Thompson estimator and show that it is unbiased for the population total. *10*

(d) Show that a split-plot design in an RBD with p main plot treatments and q sub-plot treatments provides more precise comparisons among sub-plot treatments relative to an RBD with pq plots. *10*

8. (a) Suppose that a variable of interest is distributed as U(b, b + h), b > 0, h > 0. Let the range be divided into L strata of equal sizes. An SRS of size $\frac{n}{L}$ is selected from each stratum. Denoting by V and V_1, the variances under an SRS of size n, and a stratified random sample as described above, respectively, show that $\frac{V_1}{V} = \frac{1}{L^2}$. Interpret this result. *10*

(b) Define multiple linear regression model. Find the unbiased estimators of the parameters of this model. State the sampling distribution of these estimators under the assumption of normality. *10*

(c) Explain how one missing value in an RBD is estimated so as to minimise the error sum of squares. Set up the ANOVA table for analysis in this situation. *10*

(d) Suppose that a sample of size N = 20 is drawn from $\mathbf{X} \sim N_3 (\boldsymbol{\mu}, \Sigma)$ which yields the sample mean vector and sample covariance matrix as

$$\bar{\mathbf{x}} = (21{\cdot}05,\ 21{\cdot}65,\ 28{\cdot}95)'$$

$$S = \begin{pmatrix} 2{\cdot}2605 & 2{\cdot}1763 & 1{\cdot}6342 \\ & 2{\cdot}6605 & 1{\cdot}8237 \\ & & 2{\cdot}4710 \end{pmatrix}.$$

Construct the Likelihood Ratio Test of the hypothesis

$$H_0 : \begin{pmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$ *10*