Indian Forest Service
(Main) Exam, 2020

JKLO-U-STSC

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.

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

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

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

SECTION A

Q1. (a) A printing machine can print n "letters", say $\alpha_1, \alpha_2, ..., \alpha_n$. It is operated by electrical impulses, each letter being produced by a different impulse. Assume that p is the constant probability of printing the correct letter and the impulses are independent. One of the n impulses, chosen at random, was fed into the machine twice and both times the letter α_1 was printed. Compute the probability that the impulse chosen was meant to print α_1. *8*

(b) A positive integer X is selected at random from the first 50 natural numbers. Calculate $P\left(X + \frac{48}{X} > 26\right)$. *8*

(c) Using Central Limit Theorem, show that

$$\lim_{n \to \infty} \sum_{k=n}^{5n} \binom{5n}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{5n-k} = \frac{1}{2}.$$ *8*

(d) (i) Let X_1 and X_2 be iid Poisson (λ) variates. Examine whether the statistic $T = X_1 + 2X_2$ is sufficient for λ.

(ii) Let X be Poisson (λ) and $\psi(\lambda) = e^{-3\lambda}$. Show that $(-2)^X$ is an unbiased estimator for the parametric function $\psi(\lambda)$ and examine whether it is a reasonable estimator. *4+4=8*

(e) Establish the necessary and sufficient condition for an unbiased estimator to be UMVUE. *8*

Q2. (a) (i) Obtain moment estimators of the parameters b and c of the model whose density function is given by

$$f(x; b, c) = \frac{c}{b} x^{c-1} e^{-x^c/b}, \quad x > 0, \; b, c > 0$$

based on a random sample of size n.

(ii) Consider a random sample of size n from Gamma (1, β). It is observed that only k, $0 \leq k \leq n$, of these observations are found to be less than or equal to M, where M is a fixed positive number. Obtain MLE of β, in the above set-up. *7+8=15*

(b) Let the joint probability density function of X and Y be

$$f(x, y) = C \cdot \exp\{-(9x^2 + y^2 - xy)\},$$

where C is a constant.

Find :

(i) E(X) and V(X),

(ii) E(Y) and V(Y), and

(iii) Correlation between X and Y.

Also identify the distributions of X and Y. *15*

(c) How large a sample must be taken in order that the probability will be at least 0·95, that the sample mean will be within 0·5 – neighbourhood of the population mean, provided population standard deviation is 1 ? *10*

Q3. (a) Let $X_1, X_2, ..., X_n$ be iid binomial (1, p). Obtain UMP test, to test $H_0 : p = p_0$ vs $H_1 : p = p_1 \ (> p_0)$. *15*

(b) Suppose a coin is tossed three times. If X denotes the number of heads on first toss and Y is the total number of heads, then write down the probability distribution of Z = X + Y. *15*

(c) Let $X_1, X_2, ..., X_n$ be iid $\sim N(\mu, 1)$. Find $(1 - \alpha)$ level confidence interval for μ. *10*

Q4. (a) (i) Examine whether the strong law of large numbers holds for the sequence $\{X_k\}$ of independent random variables defined as follows :

$$P\left(X_k = -1 - \frac{1}{k}\right) = \frac{1}{2}\left\{1 - \left(1 - \frac{1}{k^2}\right)^{1/2}\right\}$$

and

$$P\left(X_k = 1 + \frac{1}{k}\right) = \frac{1}{2}\left\{1 + \left(1 - \frac{1}{k^2}\right)^{1/2}\right\}.$$

(ii) If the moment generating function of a random variable X is $\left(\frac{2}{3} + \frac{1}{3}e^t\right)^9$, then show that

$$P[\mu - 2\sigma < X < \mu + 2\sigma] = \sum_{x=1}^{5} \binom{9}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x}.$$ *8+7=15*

(b) Derive an SPRT for $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 (> \theta_0)$ in sampling from $N(\mu, \sigma^2)$, σ known. Also find its OC and ASN functions. *15*

(c) Seventeen students were randomly selected to participate in an educational research project. A group of eight students out of them was asked to attend a traditional lecture course for four weeks. The remaining nine students were provided self-instructional material on video cassettes. At the end of the four weeks all the 17 students took the same examination and got the following scores :

Lecture :	75	82	28	82	94	78	76	64	
Self-instruction :	78	95	63	37	48	74	65	77	63

Using normal approximation to Mann-Whitney Wilcoxon Rank Sum Test, verify whether the difference between the scores of the two groups are significant at 5% level (Table value is required but not given intentionally). *10*

SECTION B

Q5. (a) If the cost function is of the form $C = C_0 + \sum_{h=1}^{L} t_h \sqrt{n_h}$, where C_0 and t_h are known, show that

$$n_h \propto \left\{ \frac{N_h^2 \, S_h^2}{t_h} \right\}^{2/3}$$

so that the variance of the estimator for a fixed cost is minimum. *8*

(b) Show that the analysis of data from a completely randomized fixed effects model and a completely randomized random effects model is the same. Under the standard notations, prove that $V(\hat{\tau}_j - \hat{\tau}_{j'})$ is a minimum when all treatments are replicated equal number of times. *8*

(c) Consider the data on number of trees in 10 farms and their respective yields per farm :

Farm	Size of farm (Number of trees)	Yield (Number of fruits)
1	1000	860
2	650	420
3	2100	1350
4	860	520
5	2840	1785
6	1910	1100
7	390	201
8	3200	2198
9	1500	987
10	1200	843

Select a sample of three farms using probability proportional to size of the farms (Use the random numbers 1085, 6251 and 9787 respectively).

Also compute the Hansen-Hurwitz estimator for the total number of fruits. *8*

(d) List out four important types of departure on assumptions of the simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

which can be investigated using residuals graphically. Mention any four residual plots that can be treated as diagnostic tools. *8*

(e) Given $\mathbf{X} \sim N_3(\mu, \Sigma)$, where $\mu = (2, 4, 3)'$ and $\sum = \begin{pmatrix} 8 & 2 & 3 \\ 2 & 4 & 1 \\ 3 & 1 & 3 \end{pmatrix}$,

(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

Q6. (a) A sample S of size n = 2 was drawn using simple random sampling from a population $\{y_1, y_2, y_3\}$ of size 3. The estimator proposed for the population mean is

$$\hat{\bar{y}}^* = \begin{cases} \dfrac{y_1 + y_2}{2} & \text{if} \quad S = \{y_1, y_2\} \\ \dfrac{3y_1 + 4y_2}{6} & \text{if} \quad S = \{y_1, y_3\} \\ \dfrac{3y_2 + 2y_3}{6} & \text{if} \quad S = \{y_2, y_3\} \end{cases}$$

Show that $\hat{\bar{y}}^*$ is unbiased for $\bar{y}$. Prove that $V(\hat{\bar{y}}) > V(\hat{\bar{y}}^*)$ whenever $y_3 [3y_2 - 3y_1 - y_3] > 0$, where $\hat{\bar{y}}$ is the sample mean. *15*

(b) Give the design matrix of 2^3 factorial design. Write down the contrasts for average effect of each factor, with sum of squares and associated degrees of freedom. How will the ANOVA table be modified if you are replicating the full factorial twice but in two different blocks ? *15*

(c) For a linear regression model

$$Y = \beta_0 + \beta_1 X + \varepsilon_i$$

where $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$, $Cov(\varepsilon_i, \varepsilon_j) = 0$, prove that the least square estimators of β_0 and β_1 are unbiased and have minimum variance among all unbiased linear estimators. *10*

Q7. (a) Suppose $n_1 = 11$ and $n_2 = 12$ observations are made on two random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ which are assumed to have bivariate normal distribution with a common covariance matrix Σ, but possibly different mean vectors μ_1 and μ_2. The sample mean vectors and pooled covariance matrix are

$$\overline{\mathbf{X}}_1 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \ \overline{\mathbf{X}}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \ S_{\text{pooled}} = \begin{pmatrix} 7 & -1 \\ -1 & 5 \end{pmatrix}.$$

Obtain Mahalanobis sample distance D^2 and Fisher's linear discriminant function. Assign the observation $\mathbf{X}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to either population π_1 or π_2. *15*

(b) (i) If $\mathbf{X}$ is distributed as $N_p(\mu, \Sigma)$, then prove that $\mathbf{X}' \Sigma^{-1} \mathbf{X}$ is distributed as a non-central chi-square with p degrees of freedom and non-centrality parameter $\mu' \Sigma^{-1} \mu$.

(ii) If $X_1 = Y_1 + Y_2$, $X_2 = Y_2 + Y_3$, $X_3 = Y_3 + Y_1$, where Y_1, Y_2 and Y_3 are uncorrelated random variables and each of which has zero mean and unit standard deviation, find the multiple correlation coefficient between X_2 and X_1, X_3. *8+7=15*

(c) The table below gives a simple randomized block design. Here one observation y_{11} is missing. Set up the regression model approach to the missing value problem in RBD model. Explicitly state the reduced model for accessing block effects and treatment effects. Give the breakup of degrees of freedom. *10*

	Blocks		
Treatments	1	2	3
1	X	y_{12}	y_{13}
2	y_{21}	y_{22}	y_{23}
3	y_{31}	y_{32}	y_{33}

Q8. (a) Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ has covariance matrix Σ with eigenvalue – eigenvector pairs

$(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$.

If $Y_i = \mathbf{e}_i' \mathbf{X}$, $i = 1, 2, \ldots, p$, are the principal components obtained from Σ, then show that the correlation coefficient between Y_i and X_k is given by

$$\rho(Y_i, X_k) = e_{ik} \frac{\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, \quad i, k = 1, 2, \ldots, p.$$ 15

(b) In a multiple regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

discuss any five consequences of X_1 and X_2 being perfectly correlated. 10

(c) Define ratio estimator for a population mean of character y giving its conditions. Derive the expressions for its bias and variance. When will the regression estimator and ratio estimator be equally efficient ? 15