22/1

2008

STATISTICS

Paper 1

Time : 3 Hours] *[Maximum Marks : 300*

INSTRUCTIONS

Candidates should attempt ***all*** *the questions in Parts A, B & C. However, they have to choose only* ***three*** *questions in Part D.*

Answers must be written in the medium opted (i.e. English or Kannada).

This paper has four parts :

A	20 marks
B	100 marks
C	90 marks
D	90 marks

Marks allotted to each question are indicated in each part.

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

Notations and symbols used are as usual.

PART A

4×5=20

Each question carries 5 marks.

1. (a) If $P(A) = p$ and $P(B) = q$, then show that

$$P(B|A) \geq \frac{p+q-1}{p}.$$

(b) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population with p.d.f.

$$f(x, \theta) = \theta x^{\theta - 1}, \quad 0 < x < 1; \quad \theta > 0.$$

Examine whether $\prod_{i=1}^{n} X_i$ is a sufficient estimator for θ.

(c) Distinguish between simple and composite hypotheses. Illustrate.

(d) Explain the practical advantages of a sequential test procedure.

PART B

10×10=100

Each question carries 10 marks.

2. (a) Let $X_1, X_2, ..., X_n$ be a random sample from a normal population $N(\theta, \sigma^2)$. Show that

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

is an unbiased estimator of σ^2.

(b) Stating the regularity conditions on $f(x, \theta)$, $\theta \in \Omega$, write down the Cramer – Rao lower bound for the variance of an unbiased estimator of $g(\theta)$.

3. (a) The joint probability density function of two random variables X and Y is

$$f(x, y) = \frac{1}{8}(6 - x - y), \quad 0 < x < 2, \quad 2 < y < 4.$$

Find

(i) the marginal density function of X, and

(ii) $P(X < 1, Y < 3)$.

(b) Show that the coefficient of correlation r is independent of change of origin and scale. Also prove that for two independent variables, correlation coefficient $r = 0$. Show by an example that the converse may not be true.

4. Establish Chebychev's inequality.

If X is a random variable such that $E(X) = 3$ and $E(X^2) = 13$, use Chebychev's inequality to determine a lower bound for $P(-2 < X < 8)$.

5. State and prove Lindeberg – Levy Central Limit Theorem.

[Turn over

6. The characteristic function $\varphi_X(t)$ of a discrete random variable X is given by

$$\varphi_X(t) = (p\, e^{it} + q)^n.$$

Find the probability distribution of the random variable X.

7. State and prove Borel – Cantelli lemma.

8. Stating clearly the underlying assumptions, outline the Chi-square test for goodness of fit.

9. Stating the necessary conditions on X and **U**, show that the least squares estimator of $\boldsymbol{\beta}$ in the model $\mathbf{Y} = X\boldsymbol{\beta} + \mathbf{U}$ is (a) unbiased, and (b) consistent.

10. Establish Rao – Blackwell theorem. State the significance of this theorem.

11. If **X** is distributed as $N_p(\boldsymbol{\mu}, \Sigma)$, then find the characteristic function of **X** and prove that every linear function of the components of **X** is univariate normal.

PART C

6×15=90

Each question carries 15 marks.

12. Establish Holder's inequality and deduce Cauchy – Schwartz inequality as a special case.

13. Let $X_1, X_2, \ldots, X_n$ be a random sample from

$$f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}, \; 0 < x < \infty; \; 0 < \theta < \infty.$$

Show that $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is a minimum variance bound estimator of θ and has variance θ^2/n.

14. (a) Define a maximum likelihood estimator (MLE) and state its important properties.

(b) Obtain the MLE of θ based on random samples of n observations from a uniform distribution over the interval $(0, \theta)$.

15. (a) Define the partial correlation coefficient. In usual notations prove that

$$r_{12 \cdot 3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}}.$$

(b) Define orthogonal polynomials. How are these useful in regression analysis ?

16. Outline the following methods of estimation :

(a) Method of moments

(b) Method of minimum chi-square

17. Describe the following tests :

(a) Sign test for location, and

(b) Run test for randomness.

PART D

3×30=90

*Answer any **three** of the following questions. Each question carries 30 marks.*

18. (a) State and prove Chebychev's weak law of large numbers (WLLN).

(b) Examine whether the strong law of large numbers (SLLN) holds for the sequence of mutually independent random variables $\{X_n\}$ with distribution

$$P\left(X_n = 1 + \frac{1}{n}\right) = \frac{1}{2}\left\{1 + \left(1 - \frac{1}{n^2}\right)^{\frac{1}{2}}\right\},$$

$$P\left(X_n = -1 - \frac{1}{n}\right) = \frac{1}{2}\left\{1 - \left(1 - \frac{1}{n^2}\right)^{\frac{1}{2}}\right\}.$$

19. (a) Prove that for any characteristic function φ :

$$1 - \text{Re } \varphi(t) \geq \frac{1}{4} \{1 - \text{Re } \varphi(2t)\},$$

where Re φ is the real part of φ.

(b) Let $\{X_n\}$ be a sequence of random variables with p.m.f.

$$P(X_n = 1) = \frac{1}{n}, \quad P(X_n = 0) = 1 - \frac{1}{n}.$$

Show that $X_n \xrightarrow{P} 0$.

(c) If X is a random variable such that $E(e^{aX})$ exists for $a > 0$, then prove that

$$E(|X| \geq \varepsilon) \leq E(e^{aX})/e^{a\varepsilon}.$$

20. (a) Define a most powerful test. A sample of size 1 is taken from probability density function

$$f(x, \theta) = \frac{2(\theta - x)}{\theta^2}, \quad 0 < x < \theta.$$

Find most powerful test of $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, $\theta_0 > \theta_1$, at level α.

(b) Define a U.M.P. test. Let $X_1, X_2, ..., X_n$ be a random sample from the uniform distribution on $(0, \theta)$. Obtain an U.M.P. test for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$.

21. (a) Explain the concept of one-way classified data. Write the linear model for it. Obtain the least squares estimate of the parameters involved in it.

(b) If (α, β) is the strength of a Sequential Probability Ratio Test with boundary points (A, B), prove that

$$A \leq \frac{1-\beta}{\alpha} \quad \text{and} \quad B \geq \frac{\beta}{1-\alpha}.$$

22. (a) Define Mahalanobis D^2. Show that it is invariant under any non-singular linear transformation.

(b) Define Hotelling's T^2-statistic. Develop a test for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ on the basis of a random sample of size N drawn from $N_p(\boldsymbol{\mu}, \Sigma)$, where Σ is unknown.

22/2

2008

STATISTICS

Paper 2

Time : 3 Hours] *[Maximum Marks : 300*

INSTRUCTIONS

Candidates should attempt ***all*** *the questions in Parts A, B & C. However, they have to choose only* ***three*** *questions in Part D.*

Answers must be written in the medium opted (i.e. English or Kannada).

SEAL

This paper has four parts :

A	20 marks
B	100 marks
C	90 marks
D	90 marks

Marks allotted to each question are indicated in each part.

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

Notations and symbols used are as usual.

PART A

$4 \times 5 = 20$

Each question carries 5 marks.

1. (a) Describe ratio method of estimation. When is this method more precise than simple random sampling ?

   (b) Explain 'reliability' and 'hazard rate' and establish a relation between them.

   (c) Explain duality in linear programming and give its economic interpretation.

   (d) Describe the logistic model of population growth.

PART B

10×10=100

Each question carries 10 marks.

2. Explain PPS sampling. Give an unbiased estimator for the population total under PPSWR and derive its variance. How do you estimate this variance ?

3. Explain a method of analyzing the data obtained from a randomized block design.

4. Explain the basis and construction of control charts for 'defectives' and 'defects'.

5. Derive the reliability of a system with n independent components having exponential life length when the components are connected in series.

6. Derive the probability distribution of number of customers in the $M|M|1:FIFO$ queueing system in steady state conditions. Hence determine the mean queue length.

7. Find an initial solution by the least cost method for the following transportation problem and then solve :

	D_1	D_2	D_3	Availability
O_1	2	3	4	30
O_2	6	2	5	50
O_3	3	2	7	35
Requirement	45	25	45	

8. Describe the method of constructing an abridged life table.

9. What is multicollinearity ? What are its consequences ? How do you overcome the problem of multicollinearity ?

10. Write a program in FORTRAN to fit a straight line $y = \alpha + \beta x$ to the given data $\{(x_i, y_i), i = 1, 2, ... n\}$ using the method of least squares.

11. Explain the need for scaling techniques in psychological studies. Explain any two scaling techniques.

PART C

6×15=90

Each question carries 15 marks.

12. Describe two-stage sampling. Explain its advantages. Outline a method of determining the optimum sampling and sub-sampling fractions.

13. Describe the Yate's technique of analyzing 2^3 factorial experiment. Explain the meaning of confounding in factorial experiments.

14. Outline the basis and construction of $\overline{X}$ and R charts. Derive the OC of R-chart.

15. Explain the components of a time-series. Outline the 'ratio to trend' and 'link relative' methods to measure seasonal fluctuations.

16. Solve the following linear programming problem :

Maximize $Z = 5x_1 + 3x_2$

subject to $2x_1 + x_2 \leq 1$

$x_1 + 4x_2 \geq 6$

$x_1, x_2 \geq 0.$

17. (a) What is an index number ? Explain the construction of Paasche's and Fisher's price index numbers and compare them.

(b) Explain 'stable population' and 'stationary population'.

PART D

3×30=90

*Answer any **three** of the following questions. Each question carries 30 marks.*

18. (a) Explain (i) cluster sampling, (ii) two-phase sampling.

(b) Define recurrent and transient states. Show that a state j is recurrent iff $\sum_{n=0}^{\infty} p_{jj}^{(n)}$ is divergent.

19. (a) Explain the analysis of data in a balanced incomplete block design.

(b) Compare Shewhart control charts and CUSUM control charts. Outline the V-mask procedure.

20. (a) Explain the truncated life testing for exponential model.

(b) Describe the lot-by-lot double sampling plan for attributes. Derive its expressions for OC, ASN and AOQ.

21. (a) Stating the assumptions, derive the optimum inventory policy when the demand is probabilistic.

(b) Outline the Leontief's method of fitting demand curve from time series data.

22. (a) What is the problem of identification in a simultaneous equation model ? Establish a necessary condition for identifiability.

(b) Explain

(i) infant mortality rate

(ii) general fertility rate

(iii) net reproduction rate.