Total No. of Questions : 5]

SEAT No. :

PA-3412

[Total No. of Pages : 3

[5920] - 11

M.Sc. - I

STATISTICS

ST-11 : Basics of Real Analysis & Calculus

(2019 Pattern)(CBCS) (Semester-I) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *All questions are compulsory.*
2) *Figures to the right indicates full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) Answer the following questions. **[5×2=10]**

a) Define field and write the properties of field.

b) Define interior point of a set. Give an example of each of the following.

i) A set with no interior point.

ii) A set with exactly one interior point.

c) For real numbers x, y show that $|x+y| \le |x|+|y|$.

d) Find the radius of convergence of the series $\sum_{n=1}^{\infty}\left(\frac{1}{n^2}x^n\right)$

e) Find infimum and supremum of the set $\{\frac{1}{n}, n \in \mathrm{N}\}$ & check whether it is bounded or not.

Q2) Answer the following questions.(Any 3) **[3×5=15]**

a) For any two real sequence $\{a_n\}\&\{b_n\}$ prove that:

$$\lim_{n\to\infty} \sup(a_n+b_n) \le \lim_{n\to\infty} \sup a_n + \lim_{n\to\infty} \sup b_n$$

P.T.O.

b) Define compact set. Show that every infinite subset E of a compact set K has a limit point in K.

c) Show that every Cauchy sequence in R is converge.

d) Prove or disprove: A set is closed iff it contains all it's limit points.

Q.3) Answer the following questions. (Any 3) **[3×5=15]**

a) Define limit superior (x_n) and limit inferior (x_n) of sequence $\{x_n\}$ or real number & show that a sequence $\{x_n\}$ converges iff limit superior (x_n) = limit inferior (x_n)

b) Test for convergence of the series

i) $\sum_{n=1}^{\infty} n^3 e^{-n}$

ii) $\sum_{n=1}^{\infty} (-1)^n \frac{1}{n^p}$, p is a real number

c) Show that $\lim_{n} \sqrt[n]{n} = 1$

d) Prove or disprove: There exists a rational number which is a square root of 2.

Q4) Answer the following questions.(Any 3) **[3×5=15]**

a) Show that if $a_n \to a$ and $b_n \to b$

i) $a_n b_n \to ab$

ii) $|a_n| \to |a|$

b) State and prove ratio test for the convergence of a series.

c) Check whether the following sequence is convergent:

$$a_1 = 0,\ a_{2m} = \frac{a_{2m} - 1}{2},\ a_{2m+1} = \frac{1}{2} + a_{2m}$$

d) Show that the open set (0,1) is not a compact set.

Q5) A) Answer the following questions.(Any 1) **[15]**

a) Define the following terms with suitable example. **[8]**

i) Subsequence

ii) Absolute convergence

iii) Rearrangement of a sequence

iv) Closed set

b) Show that $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$ **[7]**

B) a) State and prove Mean Value Theorem. **[9]**

b) Show that f $\in$ RS (α) on [a,b] if and only if for every $\in > 0$, there is a partition $P_\in$ of [a,b] such that, $U(P_\in, f, \alpha) - L(P_\in, f, \alpha) < \in$ **[6]**

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PA-3413

[Total No. of Pages : 2

[5920]-12

M.Sc. - I

STATISTICS

ST-12 : Linear Algebra and Numerical Methods

(2019 Pattern) (Semester - I) (4 Credits)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt each of the following. **[5×2=10]**

a) Define the vector space.

b) Define Orthogonal basis of matrix.

c) Define Row space of matrix.

d) Define Idempotent matrix with example.

e) Define Dimensions of vector space.

Q2) Attempt any three of the following. **[3×5=15]**

a) Define trace of a matrix of order n. Let Amxn and Bnxm be two matrices. Show that trace (AB) = trace (BA).

b) If the coefficient matrix A of a system of n equations in n unknowns has an inverse, then prove the system AX = b has the unique solution given by $X = A^{-1} b$.

c) Find moore-penrose inverse of given matrix.

$$M = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 1 \\ 1 & 2 & 0 \end{bmatrix}$$

d) Define a generalized inverse and Moore-Penrose generalized inverse with an examples.

P.T.O.

Q3) Attempt any three of the following. **[3×5=15]**

a) If two rows or two columns of a matrix A is same then prove that det(A) = 0

b) Let A be a $n \times n$ invertible matrix then prove that the inverse matrix of A is unique.

c) Prove that determinant of an orthogonal matrix is ± 1.

d) X is symmetric and idempotent matrix and TX is symmetric then prove that TX = XTX

Q4) Attempt any three of the following. **[3×5=15]**

a) If $\overline{J}n = \frac{1}{n}\overline{J}n$ then prove that $\overline{J}n$ is idempotent.

b) Solve the following system of linear equations by using Gauss Elimination method.

$12X_1 + 3X_2 - 5X_3 = 1$

$X_1 + 5X_2 + 3X_3 = 28$

$3X_1 + 7X_2 + 13X_3 = 76$

c) Consider the quadratic form

$X^2 + Y^2 - 3Z^2 + 2XY + 3XZ + 6YZ$

i) Write the above quadratic form in matrix form.

ii) Find Eigen value of the matrix of Quadratic form. Also examine nature of the quadratic form.

d) Use Define algebraic multiplicity and geometric multiplicity of an eigen value. Give an example with justification of a matrix having an eigen value with algebraic multiplicity and geometric multiplicity are equal.

Q5) Attempt any one of the following. **[1×15=15]**

a) i) If V is an inner product space over the field F and if $u, v \in$ V then prove that $||u + v|| \leq ||u|| + ||v||$ and $|||u|| - ||v||| \leq ||u - v||$ **[7]**

ii) Explain Helmert matrix and find the Helmert matrix of order 4(i.e. n=4)**[8]**

b) i) State and prove caley - Hamilton theorem. Explain how it can be used to find inverse. **[8]**

ii) Define adjoint matrix and explain how it can be used to find inverse.**[7]**

✪ ✪ ✪

Total No. of Questions : 5]

SEAT No. :

PA-3414

[Total No. of Pages : 3

[5920]-13
M.A./M.Sc.
STATISTICS
ST - 13 : Probability Distributions
(2019 Pattern) (Semester - I)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:
1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt each of the following. **[5×2=10]**

a) Define probability measure. Verify whether the following is probability measure or not.

$\Omega = \{a, b, c\}$, $F = \{\phi, \{a\}, \{b\}, \{c\}, \Omega\}$ $P\{a\} = 3$, $P\{b\} = \frac{2}{3}$, $P\{c\} = 0$.

b) Define the following terms.
   i) Probability generating function.
   ii) Degenerate random variable.

c) Find $P(X^2 \leq u, Y^2 \leq v)$ if joint probability density function of (X, Y) is

$$f(x,y) = \begin{cases} \dfrac{1+xy}{4} & , \quad |X| \leq |, \quad |Y| \leq | \\ 0 & , \quad \text{otherwise} \end{cases}$$

d) State non-central t-distribution and give its probability density function.

e) Prove that the random variable X defined in probability space $(\Omega, \&, \mathbb{P})$ induces a probability space $(\mathbb{R}, \mathbb{B}, \mathbb{Q})$ by means of correspondence.

Q2) Attempt any three of the following. **[3×5=15]**

Let X be a uniform (–2, 2) continuous random variable. Define Y = *g* (X),

where the function *g* (X) is defined as, $g_x(x) = \begin{cases} 1 & , \quad x > 1 \\ x & , \quad 0 \leq X \leq 1 \\ 0 & , \quad \text{otherwise} \end{cases}$

Find cumulative distribution and probability density function of y.

P.T.O.

b) State and prove characteristic properties of bivariate distribution function.

c) Define symmetric random variable about point '*a*' let X be a random variable having probability density function $F(X)=\dfrac{e^{-x}}{\left(1+e^{-x}\right)^{2}}, -\infty < X < \infty$

Check whether F (X) is symmetric about zero or not.

d) i) If X is a random variable with probability density function $F(X)=\begin{cases}\alpha e^{-\alpha x} & , \quad X>0, \alpha>0 \\ 0 & , \quad \text{otherwise}\end{cases}$

Find probability distribution of $Y = \alpha X$

ii) Let $X \sim N(\mu, \sigma^2)$, then find pdf of X^2.

Q3) Attempt any three of the following. **[3×5=15]**

Let X be a random variable with probability density function,

$F(X)=\dfrac{c}{\left(1+X^{2}\right)^{m}}, x \in \mathbb{R}, m \geq 1$ where $c=\dfrac{\sqrt{m}}{\sqrt{\frac{1}{2}}\sqrt{m-\frac{1}{2}}}$. Show that E (X^{2r}) exists if and only if $2r < 2m - 1$.

b) If $\tau(\underline{X})=\sum$ then show that, $1-\rho^2_{1.23........p} \dfrac{\left|\sum\right|}{\sigma_{11}\left|\sum_{22}\right|}$ in standard notations.

c) Let $F(X, Y) = \begin{cases} 21\, X^3 Y^3, & 0 < X, Y < 1 \\ 0 & , \quad \text{otherwise} \end{cases}$

Obtain marginal probability distribution of X and Y. Also, check whether it is linear function or not.

d) i) Let X_i, $i \geq 1$ be independent and identical random variables with generating function $P(S) = ES^x$ and N be discrete random variable with probability mass function, $P(N = i) = p$; $i \geq 1$ which is independently distributed of X_i's with generating function $g(S)$. Show that generating function of $S_N = \Sigma X_i$, is $g(P(S))$.

ii) Let $X_i \sim$ Binomial (1, p) and $S_N = \Sigma X_i$, where $N \sim$ poisson (λ). Show that $S_N | N{=}n \sim \text{Bin}(n,p)$.

Q4) Attempt any three of the following. **[3 × 5 = 15]**

a) Define non-central chi-square distribution. Obtain its moment generating function.

b) Prove that, two quadratic forms $Q_1 = \underline{X}'A\underline{X}$ and $Q_2 = \underline{X}'B\underline{X}$ are said to be independent if and only if AB = 0.

c) Let X_1, X_2, X_3 be independent gamma random variable with shape parameter λ_1, λ_2, λ_3 and scale parameter ($\alpha = 1$). Let $Y_3 = X_1 + X_2 + X_3$ then obtain joint probability density function of (y_1, y_2, y_3) and (y_1, y_2).

d) Define Marshall - Olkins exponential distribution. Identify marginal distribution of variables involve in it.

Q5) Attempt any one of the following : **[1×15=15]**

a) i) Let X_1, X_2, ..., X_n be a random sample from a population with continuous density. Show that $Y_1 = \min(X_1, X_2, \ldots, X_n)$ is exponential with parameter $n\lambda$ if and only if each X_i is exponential with parameter λ.

ii) Define Freund's bivariate exponential model.

iii) Describe Kolmogorov - Smirnov test procedure for testing the hypothesis $H_0 : F(X) = F_0(X)$ against $H_1 : F(X) \neq F_0(X)$.

iv) Let (X, Y) be random variables with common probability density function, $F(X) = \begin{cases} \alpha e^{-x}, & x > 0, \alpha > 0 \\ 0, & \text{o.w.} \end{cases}$ Find the probability distribution of X + Y using convolution method.

[5+2+4+4]

b) i) Obtain the probability density function of $X_{(r)}$ in a random sample of size n from the exponential distribution with parameter (α) and show that $X_{(r)}$ and $W_{rs} = X_{(s)} - X_{(r)}$, $r < s$ are independently distributed. What is the distribution of $W_1 = X_{(r+1)} - X_{(r)}$.

ii) State and prove Fisher - Cochran Theorem.

[7+8]

✡ ✡ ✡

Total No. of Questions : 5]

SEAT No. :

PA-3415

[Total No. of Pages : 2

[5920]-14

M.A./M.Sc. - I (Statistics)

ST - 14 : SAMPLING THEORY

(2019 Pattern) (Semester - I) (4 Credits)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All the questions are compulsory.***
2) ***Figures to right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt the following questions: **[2 each]**

a) Define the following with illustration.

   i) First stage units and second stage units.

   ii) Stratum in stratified random sampling

b) Explain the method of selecting systematic sample if the population size is not multiple of the sample size.

c) Explain the cluster sampling and give one real life application in which cluster sampling is use.

d) State the confidence interval of the population in case of SRSWR and SRSWOR method, when S^2 is not known.

e) State any two differences between inverse sampling and simple random sampling for proportion.

Q2) Attempt any three of the following: **[5 each]**

a) Prove that for the simple random sampling probability that specified unit is selected at any given draw is $\frac{1}{N}$, where N is the population size.

b) In simple random sampling without replacement for attribute, obtain the minimum sample size for the desired precision.

c) Explain in brief Lahiri's method of selecting a sample of size n by SRSWR with Probability Proportional to Size (PPS) and state its limitations.

d) Prove that in simple random sampling without replacement sample mean sum of square is an unbiased estimator of population mean sum of square.

P.T.O.

Q3) Attempt any three of the following: **[5 each]**

a) Show that for the Probability Proportional to Size without Replacement (PPSWOR) design, Horvitz Thomson estimator is unbiased for the population mean and obtain its variance.

b) For the inverse sampling define the estimator of the population mean. Hence find variance of the estimator.

c) Write short note on Circular Systematic Sampling.

d) Describe optimum allocation for fixed variance and to minimize cost of the survey in stratified random sampling. Hence, Obtain the variance of estimator in optimum allocation.

Q4) Attempt any three of the following: **[5 each]**

a) Explain Cantered systematic sampling with an illustration.

b) Define ratio estimator of the population mean. Obtain its mean squared error.

c) Prove that an estimator of relative efficiency from the sample information for a large number of clusters is given by

$$\text{Est.}(\text{R.E.}) = \left\{s_b^2 + \left(1 - \text{M}^{-1}\right)s_w^2\right\} / \text{M}\, s_b^2$$

d) Obtain the optimum subsample size for two stage sampling with equal cluster size.

Q5) Attempt any one of the following:

a) i) Prove that for the cluster sampling for proportion $\widehat{\text{P}_{cl}} = \sum_{i=1}^{n} \frac{\text{P}i}{n}$,

where $\text{P}i = \frac{ai}{\text{M}}$ = Proportion of units in the i^{th} cluster that belongs to characteristic of interest, is an unbiased estimator of population proportion. Also find the variance of the estimator. **[8]**

ii) Distinguish between sampling and non-sampling errors. Discuss in detail the mathematical model for measurement of observational errors. **[7]**

b) i) Define an estimator of the population mean in case of two-stage samplings with unequal first stage units. Obtain bias in estimation and obtain its variance. **[8]**

ii) Define regression estimator of population mean in double sampling. Show that it is biased estimator. Also find the mean squared error of this estimator. **[7]**

✈ ✈ ✈

Total No. of Questions : 5]

SEAT No. :

PA-3416

[Total No. of Pages : 2

[5920]-21

M.Sc.

STATISTICS

ST 21 : Probability theory

(2019 Pattern) (Semester-II)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt each of the following.

a) Define each of the following: **[1 each]**

i) Borel field

ii) Random variable.

b) Explain: **[1 each]**

i) Moments of random variable.

ii) Probability space.

c) Distinguish between field and sigma field. **[2]**

d) If X:No of heads turned up when 2 coins are tossed state the sigma field generated by X. **[2]**

e) State the additive property of probability measure. **[2]**

Q2) Attempt any three of the following: **[3×5=15]**

a) If $\{A_n = \left(a - \frac{1}{n}, b + \frac{1}{n}\right)$. Find $\lim_{n\to\infty}(A_n)$.

b) State and prove correspondence theorem of distribution function.

c) State and prove Jensen's inequality.

d) If $\{A_n\}_{n\geq 1}$ is increasing sequence then show that

$$\lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right).$$

P.T.O.

Q3) Attempt any three of the following: **[3×5=15]**

a) Define convergence of sequence of random variable. Explain with suitable example.

b) Compare convergence in probability and convergence in distribution.

c) If $|X_n| \leq Y$ and Y is integrable. $X_n \xrightarrow{a.s} X$ then prove that $E(X_n) \to E(X)$.

d) If $x_n \xrightarrow{P} x$, $y_n \xrightarrow{P} y$ then show that $\dfrac{x_n}{y_n} \xrightarrow{P} \dfrac{x}{y}$, provided $y_n, y > 0$.

Q4) Attempt any three of the following: **[3×5=15]**

a) State Borel 0-1 criterion and explain with suitable example.

b) State central limit theorem & explain its utility with suitable example.

c) State strong law of Large numbers. Explain with suitable example.

d) Write a note on independence of classes and independence of random variables.

Q5) Attempt any one of the following:

a) i) Prove the following. **[3×3=9]**

1) $(\lim\sup A_n)^c = \lim\inf (A_n^c)$

2) Intersection of two sigma fields is sigma field.

3) If $x_n \xrightarrow{P} x, y_n \xrightarrow{P} y$ then $x_n + y_n \xrightarrow{P} x + y$.

ii) State and prove C_r inequality, and give suitable example to explain it. **[6]**

b) i) Prove or disprove the following. **[3×3=9]**

1) If $0 \leq x_n \uparrow x$ then show that $E(x_n) \uparrow E(x)$

2) Distribution function is right continuous function.

3) $P(A \cap B) \leq P(A) + P(B) - 1$

ii) Define inverse mapping. If X is A measurable random variable then show that. **[6]**

1) $X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B)$

2) $X^{-1}(A^C) = (X^{-1}(A))^C$

❖ ❖ ❖

Total No. of Questions : 5]

SEAT No. :

PA-3417

[Total No. of Pages : 3

[5920]-22

M.Sc.-I

STATISTICS

ST-22 : Regression Analysis

(2019 Pattern) (Semester-II) (4 Credits)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt all questions. **[5×2=10]**

a) Prove that, the hat matrix H and (I–H) are idempotent.

b) For the multiple linear regression model, show that SSR (β)=y' Hy.

c) Obtain the least square estimators of the regression coefficients for centered model.

d) Give the inverse gaussian link function for the generalized linear model.

e) Explain any two criteria for the model adequacy.

Q2) Attempt any 3 questions out of 4 questions. **[3×5=15]**

a) Discuss the fitting of polynomial regression of y on single regressor variable X using orthogonal polynomials. Hence, give the first three polynomials in X.

b) Derive the distribution of error sum of squares and regression sum of squares.

c) Consider, the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$, with E($\varepsilon$)=0,Var ($\varepsilon$)=$\sigma^2$ and ε are uncorrelated. Show that, Cov $(\hat{\beta}_0, \hat{\beta}_1) = -\bar{x}\sigma^2 / S_{xx}$ and Cov$(\bar{y}, \hat{\beta}_1) = 0$

P.T.O.

d) Consider $E(Y_1)=\theta_1+\theta_2-\theta_3$, $E(Y_2) = E(Y_4)=\theta_2-\theta_4$, $E(Y_3) =\theta_1+\theta_2$ and covariance is $\sigma^2 I_n$

i) Check whether above model is full rank model or non-full rank model.

ii) Obtain rank of estimation space and rank of error space.

iii) Obtain one solution of normal equations and hence obtain BLUE of $\theta_1+\theta_2$ if it is estimable parametric function.

Q3) Attempt any 3 questions out of 4 questions. **[3×5=15]**

a) Assuming regression model with two regressor variables, discuss the effect of multicollinearity on X'X matrix and one least square estimates of regressor parameters.

b) What are outliers in regression models? Discuss any two techniques used in detecting the presence of outliers. If outlier is detected will discard it from data? Justify your answer.

c) Under what conditions ridge regression is used for fitting the model to given data ?Discuss the problem of selection of appropriate ridge estimators using graphical method.

d) How various residual plots are useful in checking validity of assumptions made in analysis of experimental data.

Q4) Attempt Any 3 questions out of 4 questions. **[3×5=15]**

a) Consider the maximum likelihood estimator $\tilde{\sigma}^2$ of σ^2 in simple linear regression model. Whether $\tilde{\sigma}^2$ is a biased estimator of σ^2? If yes, show the amount of bias in $\tilde{\sigma}^2$. What happen to the bias as the sample size n becomes large?

b) Discuss any two criteria for evaluating subset regression model.

c) Explain the development of the PRESS statistic.

d) Discuss Likelihood ratio test in logistic regression model.

Q5) Attempt any 1 questions out of 2 questions. **[1×15=15]**

a) i) Derive likelihood ratio test to test $H_0 : \underline{\beta} = \underline{\beta}_0$.

ii) For the multiple linear regression model, Explain the role of extra sum of squares principal in testing the relative importance of regressor variables. interpret the various components of regression sum of squares in case of model with three regressor variables.

iii) Discuss the Durbin-Watson test to detect the presence of autocorrelation in errors. **[7+4+4]**

b) Let $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ be the generalized linear model.

i) Obtain the normal equation for it and hence show that, its BLUE is unique almost surely.

ii) Show that $V(\hat{\underline{\beta}}) = S^{-}\sigma^2, V\left(\underline{\lambda}'\hat{\underline{\beta}}\right) = \underline{\lambda}'S^{-}\underline{\lambda}\sigma^2$ and

Cov $\left(\underline{\lambda}'_{(1)}\hat{\underline{\beta}}, \underline{\lambda}'_{(2)}\hat{\underline{\beta}}\right) = \underline{\lambda}'_{(1)} S^{-} \underline{\lambda}'_{(2)} \sigma^2$

iii) Show that, covariance between any function belongs to error space and any BLUE is zero.

[7+5+3]

● ● ●

Total No. of Questions : 5]

SEAT No. :

PA-3418

[Total No. of Pages : 3

[5920]-23

M.Sc. (Part-I)

ST 23 : STATISTICAL INFERENCE-I (4-Credits)

(2019 Pattern) (Semester-II)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt all questions. **[5×2=10]**

a) Define the following terms with an example:

i) Unbiased estimator

ii) Ancillary statistics

b) Obtain power of the following test based on a single observation from B(3,p) for testing $H_0 : p = 1/4$ vs $H_1 : p = 3/4$. $\phi(x) = \begin{cases} 1 & if\ x = 3 \\ 0 & if\ x = 0,1,2. \end{cases}$

c) Define conjugate prior with an illustration.

d) State Bhattacharya bounds.

e) Define MP test of size α . prove or disprove : MP test is unique

Q2) Attempt any three of the following: **[3×5=15]**

a) State and prove Neyman's factorization theorem for discrete case.

b) Define sufficiency and Fisher criteria for sufficiency. Let X~N (μ,σ^2), obtain the Fisher information, $1_X(\sigma)$, μ is known

c) Let $X_1, X_2, \ldots X_n$ be a random sample from N $(\theta\ \sigma^2)$,σ^2 is known, Obtain level α uniformly most accurate lower bound onθ.

d) Define complete statistic. Let X be uniform random variable with mean $\frac{\theta}{2}$ based on a sample of size n, let T be complete for θ. show that T need not be complete for $g(\theta)$ even when $g(\cdot)$ is a one-to-one function.

P.T.O.

Q3) Attempt any Three of the following: **[3×5=15]**

a) State Rao-Blackwell theorem. Let X_1, X_2, ... X_n be *i.i.d.* B $(1,\theta)$ Define

$$T_1 = \begin{cases} 1, X_1 = 1, X_2 = 0 \\ 0, \text{otherwise} \end{cases}$$

Let $T = \sum_{i=1}^{n} X_i$. Find the Rao-Blackwellize T_1 with respect to T.

b) Explain likelihood equivalence. Let M be a statistic such that

$M(\underline{x}) = M(\underline{y}) \Rightarrow \underline{x} \sim \underline{y}$ (Likelihood equivalence). Then show that M is sufficient for θ.

c) Define pitman family. If a $(\theta)\downarrow$ and b $(\theta)\uparrow$ then prove that max $\{a^{-1}(X_{(1)}), b^{-1}(X_{(n)})\}$ is minimal sufficient statistic for θ.

d) Define squared error loss function and obtain the Bayes estimator under squared error loss function.

Q4) Attempt any three of the following: **[3×5=15]**

a) Define the term: Fisher information. Compute it in case of $N(\mu,1)$ and $N(0,\sigma^2)$ distribution

b) Let X be a random variable such that under H_0 $P(X=x)=\frac{1}{3}$; x=0,1,2: and

under H_1 $P(X=x)=\binom{2}{x}\left(\frac{1}{2}\right)^2$; x=0,1,2 Find two test for H_0 Versus H_1 at

level $\alpha=\frac{1}{3}$ and find their powers. which one of them is more powerful? **[7]**

c) Let X_1, X_2,.....X_n be a random sample from U $(0,\theta)$ Obtain shortest expected length of confidence interval for θ.

d) Define monotone likelihod ratio property of a distribution. Does $N(\theta,1)$ and cauchy $(0,\theta^2)$ Possess an MLR property?

Q5) Attempt any one of the following: **[1×15=15]**

a) i) Define shortest expected length confidence interval (SELCI). Let $X_1, X_2, \ldots, X_n$ be radom sample from $f(x,\theta) = \frac{\theta}{x^{\theta+1}}, x > 1, \theta > 0.$ Find $(1-\alpha)100\%$ SELCI for θ. **[7]**

ii) Define Bayes estimate. Let $X_1, X_2, \ldots X_n$ be a random sample from P (λ). For estimating, λ using the quadratic error loss function, and prior distribution over Θ, Given by the *pdf* $\pi(\lambda)=e^{-\lambda}$ if $\lambda > 0$ and zero otherwise is used. Find the Bayes estimator for λ. **[8]**

b) i) State and prove Neyman-Pearson Lemma. **[6]**

ii) What is the relation between problem of testing of hypotheses about a parameter and construction of confidence interval for that parameter? Illustrate with suitable example. **[5]**

iii) State and prove Basu's theorem. **[4]**

● ● ●

Total No. of Questions : 5]

SEAT No. :

PA-3419

[Total No. of Pages : 4

[5920]-24
M.Sc. - I
STATISTICS
ST - 24 : Multivariate Analysis
(2019 Pattern) (Semester - II)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *All questions are compulsory.*
2) *Figures to the right indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt each of the following. **[5×2=10]**

a) Define p-variate Normal Distribution. Write its m.g.f. and characteristic function.

b) Define canonical correlation with its application.

c) Suppose the random vector $\underline{X}$ of two components X_1 & X_2 having variance-covariance matrix $\sum = \begin{bmatrix} 2 & -2 \\ -2 & 5 \end{bmatrix}$. Obtain the proportion of variation explained by first principal component.

d) If $\underline{X} \sim N_3(0, \Sigma)$, where $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}$. Find the value of ρ for which $X_1 + X_2 + X_3$ and $X_1 - X_2 - X_3$ are independent.

e) Define the following:

i) Wishart matrix

ii) Wishart distribution

P.T.O.

Q2) Attempt any three of the following: **[3×5=15]**

a) The following data matrix contains data on test scores with

X_1= Score on first test

X_2= Score on second test

X_3= total score on the two test

$$X = \begin{bmatrix} 12 & 18 & 14 & 20 & 16 \\ 17 & 20 & 16 & 18 & 19 \\ 29 & 38 & 30 & 38 & 35 \end{bmatrix}$$

Obtain sample mean vector and sample correlation matrix.

b) Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$. Obtain the maximum likelihood estimators of $\underline{\mu}$ and Σ.

c) If $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ and $\underline{X}$ is partition as $\begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$. Find the conditional distribution of $\underline{X}^{(1)}$ given $\underline{X}^{(2)} = \underline{x}^{(2)}$.

d) Write a note on one way MANOVA.

Q3) Attempt any three of the following. **[3×5=15]**

a) Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$. If A be any qXp matrix then prove that $A\underline{X} \sim N_q(A\underline{\mu},\ A\Sigma A')$. If $\underline{X} \sim N_3(\underline{\mu}, \Sigma)$ where $\underline{\mu} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ and

$\Sigma = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$ Obtain the distribution of $\begin{bmatrix} X_1 - X_2 \\ X_1 + 2X_2 - X_3 \end{bmatrix}$.

b) Define orthogonal factor model with an assumption. Also explain the terms factor loading and specific variance.

c) If $\underline{X}$ is a vector of p components $X_1, X_2, ..., X_p$ having variance-covariance matrix Σ with eigen values $\lambda_1, \lambda_2, ..., \lambda_p$ and corresponding eigen vectors $\underline{e}_1, \underline{e}_2, ..., \underline{e}_p$. Let $Y_i = \underline{e}_i' \underline{X}$ is the i^{th} PC. Show that

$$\sum_{i=1}^{p} V(X_i) = \sum_{i=1}^{p} V(Y_i) = \sum_{i=1}^{p} \lambda_i.$$

d) Write a note on partial correlation coefficient.

Q4) Attempt any three of the following. **[3×5=15]**

a) State and prove a necessary and sufficient condition for the two multivariate normal vectors to be independent.

b) The variance-covariance matrix is $\sum = Cov \begin{bmatrix} X_1 \\ X_2 \\ Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 1 & 0.95 & 0 \\ 0 & 0.95 & 1 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$

i) Find the first canonical correlation between $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and $\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$.

ii) Find the first pair of canonical variables.

c) If $D \sim W_p(n, \Sigma)$ and D is partition as $\begin{bmatrix} D_{11(mXm)} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$. Then show that $D_{11} \sim W_p(n, \Sigma_{11})$.

d) Consider the following density function:

$$f(x, y) = \frac{1}{2\pi} \exp\left\{\frac{-1}{2}\left(x^2 + y^2 + 4x - 6y + 13\right)\right\}$$

Obtain the distribution of $\begin{bmatrix} X \\ Y \end{bmatrix}$

Q5) Attempt any one of the following. **[1×15=15]**

a) i) Derive the density of Wishart distribution in canonical case. **[8]**

ii) Consider the matrix of distances. $\begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 10 & 3 & 0 & \\ 6 & 5 & 4 & 0 \end{bmatrix}$ do the cluster analysis by using single and complete linkage method. Also draw the dendogram and commend on it. **[7]**

b) i) Show that LRT for testing $H_0 : \underline{\mu} = \underline{\mu}_0$ leads to Hotelling T^2 - Statistic for $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$. **[8]**

ii) Obtain first two population principal components for the following variance-covariance matrix and also calculate the proportion of the total population variance explained by the first two principal component. $\sum = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. **[7]**

Total No. of Questions : 5]

SEAT No. :

PA-3420

[Total No. of Pages : 4

[5920]-31

S.Y. M.Sc.

STATISTICS

ST 31 : Applied Stochastic Processes

(2019 Pattern) (Semester - III) (4 Credits)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt the following questions. **[2 each]**

a) Show that for a finite Markov chain at least one state is Persistent non-null.

b) Define: i) Regenerative process, ii) Semi-Markov process.

c) Show that a state i is recurrent if and only if $\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$.

d) Define a stochastic process and give an example of a stochastic process.

e) Explain Stochastic process with their classification.

Q2) Attempt any 3 questions out of 4 questions. **[5 each]**

a) A virus may be present in one of N different forms and it may switch from one to another, which is called a mutation. At every (discrete) time point, the virus undergoes a mutation and when it does, it chooses uniformly from among the N - 1 other forms. Find the transition probability matrix.

P.T.O.

b) Define and give an example of each, (i) One-dimensional random walk, (ii) Period of a state of markov chain, (iii) Non-null recurrent state, (iv) Stationary distribution of Markov chain.

c) An organization has N employees where N is a large number. Each employee has one of three possible job classifications and changes classification independently according to a markov chain with TPM

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

What % of employees is in each classification?

d) Let $\{B(t),\ t \geq 0\}$ is a standard Brownian motion then compute the conditional distribution of $B(s)$ provided $B(t_1) = A$ and $B(t_2) = B$, where $0 < t_1 < s < t_2$.

Q3) Attempt any 3 questions out of 4 questions. **[5 each]**

a) Suppose that every man in a certain society has exactly 3 children, which independently have probability 0.5 of being a boy and 0.5 of being girl. Suppose also that the number of males in the n^{th} generation $\{X_n, n > 0\}$ form a branching process $X_0 = 1$. Find the probability that the male line of a given man eventually become extinct. If a given man has two boys and one girl, what is probability that his male line will continue forever.

b) Consider a Markov process for which the embedded Markov chain is a birth- death chain with transition probabilities $P_{i,i+1} = 2/5$ for all $i \geq 0$, $P_{i,i-1} = 3/5$ for all $i \geq 1$, $P_{01} = 1$ and $P_{ij} = 0$ otherwise. Find the steady-state probabilities $\pi_i; i \geq 0$ for the embedded chain.

c) Let $\phi(s)$ be the probability generating function of the branching process $\{X_n, n > 0\}$. Show that $\phi_n(s) = \phi_{n-1}(\phi(s)) = \phi(\phi_{n-1}(s))$ where ϕ is the probability generating function of the offspring distribution and $X_0 = 1$.

d) If $\{N(t), t \geq 0\}$ is a renewal process with inter arrival times $X_1, X_2, \ldots$ then prove that $\frac{n(t)}{t} \to \frac{1}{\mu}$. with probability 1 as $t \to \infty$, where $\mu = E(X_i)$.

Q4) Attempt any 3 questions out of 4 questions. **[5 each]**

a) Customers arrive at a service station according to a Poisson process of rate $\lambda = 3$ customer per hour. Suppose 3 customers arrived during the 1st 30 minutes. What is the probability that only 4 customers arrived during the hour? What is the probability that there is no customer in 1st 10 minutes.

b) Vehicle stopping at road side restaurant form a Poisson process with rate $\lambda = 20$/hour. Vehicle has 1, 2, 3, 4 and 5 people in it with respective probabilities 0.3,0.3, 0.2, 0.1, and 0.1. Find the expected number of persons arrive at the restaurant within 1st hour.

c) For a Poisson process show, for $s < t$, that

$$P\{N(s)=k \mid N(t)=n\}=\binom{n}{k}\left(\frac{s}{t}\right)^{k}\left(1-\frac{s}{t}\right)^{n-k} \quad k=0,1,\ldots,n$$

d) Let $\{X_1(t), t \geq 0\}$ and $\{X_2(t), t \geq 0\}$ be two independent Poisson process with rates λ_1 and λ_2 respectively. What is the probability that $X_1(t) = 1$ before $X_2(t) = 1$?

Q5) Attempt any 1 questions out of 2 questions.

a) i) Let $\{X_n, n > 0\}$ be a Markov chain with state space S = {1, 2,3, 4} and transition probability matrix P as,

$$P=\begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$. Classify the states

into i) Persistent null (ii) Persistent non-null (iii) transient (iv) Ergodic (v) Non-ergodic (vi) Periodic. **[8]**

ii) Give the two definition of Poisson Process and establish their equivalence. **[7]**

b) i) A job consists of 3 machines and 2 repairmen. The amount of time a machine works before breaking down is exponentially distributed with mean 8, then **[8]**

1) What is the average number of machines not in use?

2) What proportion of time are both repairmen busy?

ii) Define Brownian motion. Write it as a function of Standard Brownian motion. Define Geometric and integrated Brownian motion. Show that Brownian motion process can be obtained as the limit of a random walk. State the assumptions and results which you have used. **[7]**

Total No. of Questions : 5]

SEAT No. :

PA-3421

[Total No. of Pages : 3

[5920]-32
M.Sc.
STATISTICS
ST - 32(A) : Bayesian Inference
(2019 Pattern) (Semester - III) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) **Answer the following :** **[5 × 2 = 10]**

a) Explain the term subjective probability and frequentist probability.

b) State MCMC algorithm.

c) Define non-informative prior with an illustration.

d) Outline the difference between frequentist approach and Bayesian approach for hypothesis testing.

e) Define Prior and Posterior distribution. State the relationship between prior, posterior and likelihood, given by Bayes.

Q2) **Attempt any Three of the following :** **[3 × 5 = 15]**

a) Define absolute error loss function and obtain the Bayes estimator under absolute error loss function.

b) Suppose $\underline{X} = (X_1, X_2, \_\_\_\_\_\_, X_{10})$ is a random sample of size 10 from $N(\mu, 4)$ with $\overline{X} = 2.1$. Suppose we assume a normal prior $N(100,4)$ for μ. Find the Bayesian estimator of μ under quadratic error loss.

c) Suppose that the signal $X \sim N(0, \sigma_x^2)$ is transmitted over a communication channel. Assume that the received signal is given by $Y = X + W$, where $W \sim N(0, \sigma_w^2)$ is independent of X. Find the maximum aposterior (MAP) estimate of X given $Y = y$.

d) Let $X_1, X_2, \ldots\ldots, X_n$ be an independent and identically distributed with *Binomial* (k, p). The prior distribution of p is beta distribution with parameters α and β. Obtain 95% highest posterior density (HPD) confidence interval for p.

Q3) **Attempt any Three of the following : [3 × 5 = 15]**

a) Describe the any three application of Bayesian inference.

b) Define maximum aposterior estimator. Let $(X_1 = 2, X_2 = 3, X_3 = 1)$ be a i.i.d. observations from *Binomial* $(5, p)$ and let the prior distribution of $p \sim U(0,1)$. Find the maximum aposterior estimator of p based on the given data.

c) Write note on EM algorithm.

d) A new telephone company predicts to handle an average of 1000 calls per hour. During 10 randomly selected hours of operation, it handled 7269 calls. Obtain 95% credible interval set. Telephone calls are placed according to Poisson process. Exponential prior distribution of the hourly rate of call is applicable.

Q4) **Attempt any Three of the following : [3 × 5 = 15]**

a) A blood test is considered for determining the sugar level of person with diabetes two hours after he had his breakfast. It is of interest to see if medication has controlled his blood sugar level. Assume that test result X has $N(\theta, 100)$ distribution. where θ is the true level. In the appropriate population (diabetic but under this treatment), θ is distributed according to $N(100, 900)$. Suppose we want to test $H_0 : \theta \leq 130$ against $H_1 : \theta > 130$. If a blood test shows a sugar level of 130, what can be concluded?

b) Discuss conjugate class of priors with an example.

c) Let $X_1, X_2, \ldots\ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$, σ^2 known. Suppose $\mu \sim N(\mu_0, \sigma_0^2)$. Obtain $100(1 - \alpha)$ %highest posterior density (HPD) confidence interval for μ.

d) Let $X \sim N(\mu, \sigma^2), \mu$ is known. Find the Jeffrey's prior for σ^2.

Q5) Attempt any One of the following : [1 × 15 = 15]

a) i) Define prior and posterior density. Let $X \sim N(0,1)$. Suppose that we know $Y \mid X = x \sim N(x, 1)$. Show that the posterior density of X given $Y = y$, $f_{X|Y}(x \mid y)$ is given by $X \mid Y = y \sim \mathrm{N}\left(\frac{y}{2}, \frac{1}{2}\right)$. Also find bayes estimator under squared loss. **[8]**

ii) Write note on Gibbs sampling. **[7]**

b) i) Let $Y \mid \theta \sim N(\theta, 4)$ and prior distribution of θ is $N(0,4)$. Suppose random sample $Y_1, Y_2, \ldots\ldots, Y_{25}$ drawn from $(\theta, 4)$. Find 95% credible interval for θ given that we have observed $\bar{Y} = \frac{1}{25}\sum_{i=1}^{25} Y_i = 0.56.$ **[9]**

ii) Explain the terms with illustration : subjective priors and probability matching prior. **[6]**

❄ ❄ ❄

Total No. of Questions : 5]

SEAT No. :

PA-3422

[Total No. of Pages : 3

[5920]-33
M.Sc.
STATISTICS
ST - 33 : Design and Analysis of Experiments
(2019 Pattern) (4 Credits) (Semester - III)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt all questions : **[2 × 5 = 10]**

a) Define the terms :

i) Alias

ii) Resolution

b) Write the layout of CCD with $k = 2$ and $\alpha = \sqrt{2}$ and five center points.

c) For BIBD, prove that $b \geq a + r - k$.

d) What is confounding? Why one has to use the technique of confounding?

e) State the BLUE of elementary contrast in treatment effects for one way classification model with single covariate.

Q2) Attempt any 3 questions out of 4 questions : **[3 × 5 = 15]**

a) Consider two-way classification model with r-observations per cell and interaction present. Give one complete set of linearly independent estimable parametric functions and their BLUEs. What is rank of error space and estimation space?

b) Show that for symmetric BIBD any two blocks have λ treatments common.

c) For the analysis of covariance model obtain the least square estimates of the parameters and hence obtain an expression for residual sum of squares.

d) Let $\overline{X}_1 = 9.8, \overline{X}_2 = 15.4, \overline{X}_3 = 17.6, \overline{X}_4 = 21.6, \overline{X}_5 = 10.8, \text{MSE} = 8.06$ and $n = 5$. Apply Newman-Keuls test to the above data and write the conclusion.

Q3) Attempt any 3 questions out of 4 questions : **[3 × 5 = 15]**

a) Generate a 2^{7-4} fractional factorial experiment with design generators D = AB, E = –AC, F = BC and G = ABC. Identify resolution of this design and give the alias structure for all main effects for the generated design.

b) Confound the 2^4 factorial design in two blocks where the effects to be confounded is ABCD.

c) Prove that, in 2^3 factorial experiment main effects and interaction effects are mutually orthogonal treatment contrast.

d) Define PBIBD. Explain all the parameters of PBIBD with conditions.

Q4) Attempt any 3 questions out of 4 questions : **[3 × 5 = 15]**

a) What are center points in 2^k factorial experiments? Discuss the purpose of addition of center points to factorial design.

b) Consider the second order response surface model with k explanatory variables fitted using n observations. Describe the method of canonical analysis to determine the nature of the stationary point. State the assumptions made.

c) What are rotatable designs? State the condition of rotatability for CCD.

d) Discuss briefly on the Taguchi designs and its analysis methods.

Q5) Attempt any 1 question out of 2 questions : **[1 × 15 = 15]**

a) i) Define matrix experiments. Discuss on the flexibility of orthogonal designs with respect to $L_8(2^7)$ orthogonal array through linear graph.

ii) Define adjusted average quality loss function. Show that to minimize it one has to maximize the signal to noise ratio.

iii) Discuss the statistical test for testing the hypothesis of linearity in factor effects in two-level factorial design.

[7 + 4 + 4]

b) i) What is random effect model? Give an unbiased estimator of variance component in one way classification model assuming treatment effects random.

ii) Explain Friedman test for one-way ANOVA with repeated measure.

iii) Obtain the incidence matrix of RBD. Hence check its orthogonality, connectedness balancedness.

[6 + 5 + 4]

Total No. of Questions : 5]

SEAT No. :

PA-3423

[Total No. of Pages : 2

[5920]-34

M.Sc.

STATISTICS

ST 34 : MACHINE LEARNING

(2019 Pattern) (Semester - III) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) **Attempt all questions:** **[5 × 2 = 10]**

a) Write down any two real life application of KNN algorithms?

b) What is a dendrogram in Hierarchical clustering algorithm?

c) Write down the different types of nodes in decision tree.

d) Define unsupervised machine learning with a suitable example

e) What is soft margin in support vector machine?

Q2) **Attempt any 3 questions out of 4 questions:** **[3 × 5 = 15]**

a) Explain why should we not use the KNN algorithm for large datasets?

b) Explain how does back propagation work?

c) What do you understand about information gain? Also explain the mathematical formulation associated with it.

d) Explain the Agglomerative Hierarchical Clustering algorithm with the help of an example.

P.T.O.

Q3) **Attempt any 3 questions out of 4 questions:** **[3 × 5 = 15]**

a) What is support vector machine and how does it work?

b) Explain advantages of artificial Neural Network?

c) Explain the different linkage methods used in the Hierarchical clustering Algorithm.

d) What is decision tree and explain how does it work?

Q4) **Attempt any 3 questions out of 4 questions:** **[3 × 5 = 15]**

a) What are the applications of Machine Learning?

b) Explain threshold, relu and sigmoid activation function.

c) Define confusion matrix and explain the term precision and recall.

d) What are the advantages and disadvantages of KNN algorithm?

Q5) **Attempt any 1 question out of 2 questions:**

a) i) What is decision tree explain how does it work? **[8]**

ii) What are the various types of Hierarchical clustering? **[7]**

b) i) What are the advantages and disadvantages of Random Forest algorithm? **[8]**

ii) Explain false negative, false positive, true negative and true positive with a suitable example. **[7]**

❍❍❍❍

Total No. of Questions : 5]

SEAT No. :

PA-3424

[Total No. of Pages : 2

[5920]-35

M.Sc.

STATISTICS

ST 32 (B) : Statistical Quality Control (SPC)

(2019 Pattern) (Semester - III) (4 Credits)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt all questions. **[2×5=10]**

a) Explain the construction of $\overline{X}$ and S chart.

b) Define the capability index Cp and performance index C_{pk}. Also state interpretation of C_p and C_{Pk}.

c) What are the disadvantages of implementing V-mask procedure?

d) Write a short note on Six-sigma methodology.

e) Define oc curve with reterence to control chart.

Q2) Attempt any 3 questions out of 4 questions. **[3×5=15]**

a) Explain the working of tabular cumulative sum (CUSUM) chart for monitoring the process mean, when standard deviation is known.

b) Define the term quality and give different dimensions of quality.

c) Explain the construction and working of Hotelling T^2 control chart for process mean vector when dispersion matrix is known. State the real-life example, where Hotelling T^2 is required.

d) A process is in control with $\overline{\overline{X}}$ = 100 and $\overline{S}$ = 1.05, n = 5. The process specifications are at 95 ± 10. The quality characteristic has normal distribution. Compute C_p, C_{pk} and C_{pm}. Interpret these ratios.

P.T.O.

Q3) Attempt any 3 questions out of 4 questions. **[3×5=15]**

a) Explain the construction and working of non-parametric control chart based on sign test. State the probability of chart statistic and find the expression for ARL(0).

b) Explain the criteria for detecting lack of control process.

c) For the EWMA chart, obtain chart statistic, its expectation and variance.

d) Obtain the $100(1-\alpha)\%$ confidence interval for capability index C_p and give the testing procedure of it.

Q4) Attempt any 3 questions out of 4 questions. **[3×5=15]**

a) Give the construction of control chart for dispersion matrix when mean vector is known and unknown.

b) Describe Dodge Roming sampling plans

c) Explain any two criteria to determining the sample size for *p*-chart.

d) State the equivalence between sampling plans and testing of hypothesis problem.

Q5) Attempt any 1 question out of 2 questions.

a) i) Explain the need of sampling plans in Industry.

ii) Describe double sampling plan for attributes. Obtain AOQ and ASN for the same. **(7+8)**

b) Write a short note on: **[15]**

i) an item-by-item sequential sampling plan

ii) Total quality management

❍❍❍❍

Total No. of Questions : 5]

SEAT No. :

PA-3425

[Total No. of Pages : 2

[5920]-41
M.Sc.
STATISTICS
ST 41 : Asymptotic Inference
(2019 Pattern) (Semester - IV) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) **Answer the following :** **[5 × 2 = 10]**

a) Define CAN estimator. Give one example.

b) State any four properties of MLE.

c) Define BAN estimator.

d) Define Joint Consistency.

e) Describe Wald Test.

Q2) **Attempt any 3 questions :** **[3 × 5 = 15]**

a) Discuss the method of moments of obtaining consistent estimator for real valued parameter θ. Using the method of moments show that r^{th} sample raw moment is consistent for r^{th} population raw moment.

b) Let $X_1, X_2, \ldots\ldots, X_n$ be random sample from $U(0, \theta)$. Examine the consistency of $X_{(n)}$ for θ. Justify you answer.

c) State and prove invariance property of CAN estimator in one parameter setup.

d) Find maximum likelihood estimate of σ based on one observation when $X \rightarrow N(0, \sigma^2)$.

P.T.O.

Q3) Attempt any 3 questions : **[3 × 5 = 15]**

a) Let $X_1, X_2, \ldots\ldots\ldots, X_n$ be random sample from Ber (θ). What is the asymptotic distribution of $\bar{X}(1-\bar{X})$? Justify it.

b) Describe Newton Raphson method to obtain MLE with one example.

c) Let $X_1, X_2, \ldots\ldots\ldots, X_n$ be random sample from $f(x) = e^{-(x-\theta)}$ $x \geq \theta$. If $T_1 = X_{(1)}$ and $T_2 = \bar{X} - 1$ be two consistent estimator for θ. Which is more efficient?

d) Let $X_1, X_2, \ldots\ldots\ldots, X_n$ be random sample from Gamma (α, λ). Find consistent estimator for (α, λ).

Q4) Attempt any 3 questions : **[3 × 5 = 15]**

a) Let $X_1, X_2, \ldots\ldots\ldots, X_n$ be random sample from P (λ). What is the distribution of $\frac{\bar{X}}{S^2}$?

b) Let $X_1, X_2, \ldots\ldots\ldots, X_n$ be random sample from cauchy (μ, λ). Obtain asymptotic distribution of CAN estimator of (μ, λ).

c) Explain Bartlett's test for homogeneity of variance.

d) Explain Rao's Score test.

Q5) Attempt any one of the following :

a) i) Define superefficient estimator with example. **[8]**

ii) Describe likelihood test (LRT). Obtain the asymptotic distribution of LRT statistic. **[7]**

b) i) State Crammer regularity conditions in one parameter setup. Give an example which satisfied regularity conditions. Justify you answer. **[8]**

ii) What is the variance stabilizing transformation? Based on random sample of size n from exponetial (mean (θ)), obtain 100 (1-α)% confidence interval for θ based on variance stabilizing transformation. **[7]**

❄ ❄ ❄

Total No. of Questions : 5]

SEAT No. :

PA-3426

[Total No. of Pages : 3

[5920]-42
M.Sc. - II
STATISTICS
ST 44(B): Analysis of Clinical Trials (4 Credits)
(2019 Pattern) (Semester - IV)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates :

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) **Attempt all questions :** **[5 × 2 = 10]**

a) Define NDA, FDA or US-FDA with importance.

b) Define Placebo and Treatment with example.

c) Explain Four sequence design.

d) Write the difference between Longitudinal design and Cross sectional design with example.

e) Explain Interim analysis.

Q2) **Attempt any three out of Four questions :** **[3 × 5 = 15]**

a) Define Clinical Trial. Explain Placebo and Treatment with examples.

b) Explain Four Phases of Clinical Trial.

c) Explain Good clinical practice.

d) Write and Explain Protocols of Clinical trial.

Q3) **Attempt any three out of Four questions :** **[3 × 5 = 15]**

a) Explain Classical Shortest confidence interval approach for claiming bioequivalence.

b) For standard 2×2 cross over design, write down model and assumptions. Also develop a test for testing direct drug effects when there is no carry over effects.

c) What are the benefits and risks of clinical trial?

d) Define Blinding and explain Types of Blinding's.

Q4) **Attempt any three out of Four questions :** **[3 × 5 = 15]**

a) Write an algorithm of Williams design. Explain it with four formulation.

b) Derive the formula for sample size for two sample test for mean in case of two sided alternative.

c) Derive the relation between K_e and $t_{1/2}$.

d) Explain Kaplan mier nonparametric tests in clinical trial.

Q5) **Attempt any one out of Two questions :** **[1 × 15 = 15]**

a) i) The list of Primidone concentration (μg/mlversus time point (hrs) from subject over 32 hours period after administrated a 250 gm tablet of drugs. The blood sample were drawn immediately before and after at time points are as follows. **[9]**

Sr.No.	1	2	3	4	5	6	7	8	9	10	11	12	13
Ti	0	0.5	1	1.5	2	3	4	6	8	12	16	24	32
Ci	0	0	2.8	4.4	4.4	4.7	4.1	4	3.6	3.0	2.5	2	1.6

Find T_{max}, C_{max}, $AUC_{(0\text{-}32)}$, $AUC_{(0\text{-}\infty)}$, $\log_{10}Ci = 0.6713 - 0.01518t_i$.

ii) Define types of Decision rules and Describe them. **[6]**

b) i) Write the algorithm of Design A. Explain it in brief. **[6]**

ii) The study was conducted with 24 healthy volunteers with the standard 2 × 2 cross over design. During each dosing period each subject was administrated either 5, 50gm tablet (test formulation) for 5 ml of an oral suspension (50 milligram/ml) (Reference Formulation). Blood sample were obtained 0 hours prior to dosing and various times after dosing the total of a sequence means, and period means as follows, With $n_1 = n_2 = 12$ $\sigma_u^2 = 1473.7$, and $\sigma_d^2 = 83.623$ then find carry over effect and Direct drug effect. **[9]**

Sequence	Period I	Period II	Sequence Mean
I	85.82	81.8	83.81
II	78.74	79.30	79.02
Period means	82.28	88.55	81.42

✷ ✷ ✷

Total No. of Questions : 5]

SEAT No. :

PA-3427

[Total No. of Pages : 3

[5920]-43
M.Sc.
STATISTICS
ST - 42 (A) : Econometrics and Time Series
(2019 Pattern) (Semester - IV)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symblos and abbreviations have their usual meaning.***

Q1) Attempt each of the following : **[5 × 2 = 10]**

a) Define the following :

i) White Noise

ii) IID Noise

b) Explain the general approach for time series modelling.

c) If $\{X_t\}$ and $\{Y_t\}$ be independent stationary time series then show that the time series $aX_t + bY_t$ is also a stationary time series. Where a and b ar constants.

d) Define MA(1) process. Give its ACVF.

e) Define SARIMA (p, d, q) X (P, D, Q)s model.

Q2) Attempt any three of the following : **[3 × 5 = 15]**

a) Determine which of the following processes are causal and which of them are invertible :

i) $X_t + 0.2X_{t-1} - 0.48X_{t-2} = Z_t$, where $\{Z_t\} \sim WN(0, \sigma^2)$

ii) $X_t - 0.75X_{t-1} + 0.5625X_{t-2} = Z_t + 1.25Z_{t-1}$, where $\{Z_t\} \sim WN(0, \sigma^2)$

b) Obtain ACF of AR(2) process.

P.T.O.

c) Let $\{X_t\}$ be stationary time series with mean μ and ACF $p(h)$. Prove that the best linear predictor of the X_{n+h} in the form $aX_n + b$ can be obtained by $a = p(h)$ and $b = \mu(1 - p(h))$.

d) Define ARIM A(p, d, q) process. Show that ARIM A(0, 1, 1) is a random walk model.

Q3) Attempt any three of the following : **[3 × 5 = 15]**

a) Explain the following tests for testing the estimated noise sequence :

i) The turning point test

ii) The difference sign test

b) Define MA(q) process. Obtain its ACVF.

c) Explain Indirect Least Squares Method of Econometrics.

d) Compute ψ_j & π_j coefficients for $j = 1, 2, \ldots\ldots\ldots\ldots 5$ for the following processes :

i) $X_t - 0.5X_{t-1} = Z_t + 0.4Z_{t-1}$, where $\{Z_t\} \sim WN(0, \sigma^2)$

ii) $X_t + 0.6X_{t-1} = Z_t + 1.2Z_{t-1}$, where $\{Z_t\} \sim WN(0, \sigma^2)$

Q4) Attempt any three of the following : **[3 × 5 = 15]**

a) Calculate ACVF of the following stationary time series.

i) $Y_t = \mu + Z_t + \theta_1 Z_{t-1} + \theta_{12} Z_{t-12;}$ where $\{Z_t\} \sim WN(0, \sigma^2)$

ii) $X_t = Z_t + \frac{1}{\theta} Z_{t-1,}$ where $\{Z_t\} \sim WN\left(0, \sigma^2\theta^2\right)$

b) Obtain stationary solution of ARMA(1, 1) process. At what condition process is invertible and non-invertible.

c) Prove that $E\left(X_{n+h} - P_n X_{n+h}\right)^2 = v_0 - a_n^1 v_n(h)$

d) Explain exponential-Smoothing method. Apply it to estimate the trend for the following data.

Take $\alpha = 0.6$.

t	1	2	3	4	5	6	7	8
X_t	6	7	5	8	10	11	9	13

Q5) Attempt any one of the following :

a) i) Discuss the following methods for eliminating trend in absence of seasonality : **[8]**

A) Graphical Method

B) Moving-Average Method

ii) Let $\{Y_t\}$ be a stationary time series with mean 0 and covariance $v_y(h)$. If $\sum_{j=-\infty}^{\infty}|\psi_j|<\infty$ then show that the time series $X_t=\sum_{j=-\infty}^{\infty}\psi_j Y_{t-j}$ is stationary with mean 0 and ACVF $v_y(h)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\psi_j\psi_k v_y(h+k-j)$. Use this theorem to obtain ACVF of linear process. **[7]**

b) i) Determine the ψ_j and π_j coefficients of the ARMA(p, q) Process. **[8]**

ii) Define sample ACVF and Sample ACF. Obtain it for the following data. **[7]**

t	1	2	3	4
X_t	6	8	5	12

□□□

Total No. of Questions : 5] **SEAT No. :**

PA-3428 **[Total No. of Pages : 3**

[5920]-44

S.Y. M.Sc. (Statistics)

ST 42 (B) : OPERATION RESEARCH

(2019 Pattern) (Semester - IV)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt each of the following. **[5×2=10]**

a) State assumption of linear programming problem.

b) What is restricted assignment problem? How much problem can be solved?

c) State Khun-Tucker necessary and sufficient conditions.

d) Formulate mathematically a transportation problem having rn-origins and n-conditions.

e) Prove that extreme points are finite in number, provided number of decision variables is finite in number.

Q2) Attempt any three of the following. **[3×5=15]**

a) Explain the Hungarian method of solving an assignment problem.

b) State and prove modified Slacksness property of duality.

c) Discuss the effect in optimal solution to an LPP due to variations in cost element of basic and non-basic variable.

P.T.O.

d) Solve following LPP problem by using two phase method.

Maximize $Z = 2X_1 + 5X_2$

Subject to constraints

$-2x_1 x_2 \leq 0$

$x_1 + 3x_2 \leq 14$

$x_1 + x_2 \leq 8$

$x_1 \geq 0 \ x_2 \geq 0$

Q3) Attempt any three of the following. **[3×5=15]**

a) Prove that $x_1 = 4 \ x_2 = 0, \ X_3 = 1$ is a feasible solution but not a basic feasible solution for

$2x_1 - x_2 - x_3 = 7$

$4x_1 + 3x_2 - 2x_3 = 14$

b) Describe Gomory's cutting plan method to solve Pure Integer Programming Problem (PIPP).

c) What do you mean by simulation? Describe the simulation process and discuss its application to M/M/1 queue system.

d) State and prove fundamental theorem of duality.

Q4) Attempt any three of the following. **[3×5=15]**

a) Use dynamic programming problem to solve the following

LPP, Maximize $Z = 3x_1 + 4x_2$

Subject to constraints

$2x_1 + x_2 \leq 40$

$2x_1 + 5x_2 \leq 180$

$x_1 \geq 0 \ x_2 \geq 0$

b) State and prove condition for the existence of an unbounded solution to the LPP.

c) Define Quadratic Programming Problem(QPP). Explain the procedure of Beale's method to solve QPP.

d) State and prove necessary and sufficient condition for Transportation problem to have a feasible solution.

Q5) Attempt any one of the following.

a) i) Show with the help of an illustration that if a primal has infeasible solution then dual need not have unbounded solution. **[7]**

ii) Sole the following QPP by using Wolfe's method. **[8]**

$$z = 4x_1^2 + 5x_2^2 - 10x_1x_2 - 6x_1 - x_2$$

Maximize

subject to constraint

$3x_1 + 11x_2 \leq 16$

$4x_1 + 5x_2 \leq 5$

$x_1 \geq 0\ x_2 \geq 0$

b) i) Discuss the Charne's Big M-Method of solving LPP. **[6]**

ii) Sole the following pure integer programming problem by using Branch and Bound method. **[9]**

Maximize $z = 3x_1 + 10x_2$

subject to constraint

$x_1 + 5x_2 \leq 12$

$x_1 \leq 3$

$x_1 \geq 0\ x_2 \geq 0$

❍❍❍❍

Total No. of Questions : 5]

SEAT No. :

PA-3429

[Total No. of Pages : 3

[5920]-45

M.Sc.

STATISTICS

ST - 43 (A) : Survival Analysis

(2019 Pattern) (Semester - IV) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *All the questions are compulsory.*
2) *Figures to right indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) **Attempt the following :** **[2 Marks Each]**

a) Explain Right random censoring with an illustration.

b) Define the following

i) Mean residual life function.

ii) IFR class of life distribution.

c) Explain the concept of bathtub failure rate.

d) Suppose 20 items from an exponential distribution are put on life test and observed for 150 hours. During this period 15 items fail with the following life times, measured in hours:

3,19,23,26,27 37,38,41,45,58,84,90,99,109,138.

Obtain maximum likelihood estimator of on an average lifetime of items.

e) Show that empirical survival function is consistent estimator of survival function.

P.T.O.

Q2) **Attempt any three questions from the following : [5 Marks Each]**

a) Prove the following implications.

IMRL $\rightarrow$ NWUE and NBUE $\rightarrow$ HNBUE

b) Show that the hazard rate is constant if and only if underlying distribution is the exponential distribution.

c) Show that no-ageing is characterize by exponential equilibrium distribution function.

d) Explain the procedure to obtain the maximum likelihood estimators of parameters of Gamma distribution for complete data.

Q3) **Attempt any three questions from the following : [5 Marks Each]**

a) Derive the maximum likelihood estimators of parameter of exponential distribution if type II censoring is used also find the confidence interval for the parameter.

b) State and prove Cauchy functional equation.

c) Prove that F is IFR if and only if equilibrium distribution function is concave function of t.

d) Find the hazard rate for the Lehman's family.

Q4) **Attempt any three questions of the following : [5 Marks Each]**

a) The following failure and censor time (in operating hours) were recorded on 12 turbine veins : 142, 149, 320, 345^+, 560, 805, 1 130^+, 1720, 2480^+, 4210^+, 5280, 6890. (+ indicates censored observation). Censoring was result of failure mode other than wear out. Find the Kaplan- Meier estimator of the survival function.

b) Find estimator of variance of the actuarial estimator of the survival function.

c) Show that Deshpande's test statistic lies between 0.5 to 1.

d) Explain what is meant by covariates with illustration? Also explain model formulation in covariate analysis.

***Q5)* Attempt any one of the following :**

a) i) Explain Hollander and Proschan test for testing exponentiality against NBU class of life distribution. **[9]**

ii) Discuss two graphical methods to check the exponentiality of the data. **[6]**

b) i) Define spacing and normalize spacing. Hence find the distribution of spacing. **[7]**

ii) Discuss Mann Whitney U-test for testing whether two samples come from the population having same distribution function. **[8]**

Total No. of Questions : 5]

SEAT No. :

PA-3430

[Total No. of Pages : 3

[5920]-46
M.Sc.
STATISTICS
ST 43(B): Categorical Data Analysis (4 Credits)
(2019 Pattern) (Semester - IV)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates :

1) *All questions are compulsory.*
2) *Figures to the right indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) **Attempt all of the following questions :** **[5 × 2 = 10]**

a) Provide the general layout of a 2 × 2 contingency table and define joint and marginal probability.

b) The following 2 × 2 contingency table is from a report on Aspirin use and Myocardial Infarction. Determine odds ratio and provide your interpretation.

		Myocardial Infraction		
		Yes	No	Total
Group	Placebo	189	10845	11034
	Aspirin	104	10933	11037
	Total	293	21778	22071

c) Define Relative risk for a 2 × 2 contingencytable.

d) Define deviance term for GLM Model.

e) Define Hierachical model.

P.T.O.

Q2) **Attempt any three questions from the following questions :[3 × 5 = 15]**

a) i) Discuss the need for Poisson regression model and its use.

ii) State the assumptions of Poisson Regression Model.

iii) Provide the tests for determining Statistical significance of regression coefficients.

b) Discuss Cochran –Mantel –Haenszel test for 2*2 contigency tables.

c) What is Multinomial Distribution in Detail

d) Discuss the following goodness of model fit measures

i) Pearson Chisquare

ii) Akaike Information Criteria

Q3) **Attempt any three questions from the following questions :[3 × 5 = 15]**

a) Explain I × J Contingency table and explain the inference procedure for testing the dependency between two categorical variables with I × J contingencytable.

b) Explain General Linear Model in detail.

c) Explain any two association measures based on 2 × 2 contingency table.

d) Which scale of measurement is most appropriate for the following variables– Nominal, or Ordinal, Interval?

i) Political party affiliation (Democrat, Republican, unaffiliated).

ii) Highest degree obtained (none, high school, bachelor’s, master’s, doctorate).

iii) Patient condition (good, fair, serious, critical).

iv) Hospital location (London, Boston, Madison, Rochester, Toronto).

v) Favorite beverage (beer, juice, milk, soft drink, wine, other)

Q4) **Attempt any three questions from the following questions :[3 × 5 = 15]**

a) Explain Binary Logistic regression in detail.

b) Explain Cumulative Logistic Regression in detail.

c) Explain Log-Linear Analysis for analyzing dependency incontingency table.

d) Explain Probit model in detail.

Q5) **Attempt any one questions from the following questions :[1 × 15 = 15]**

a) A study used logistic regression to determine characteristics associated with Y = whether a cancer patient achieved remission (1 = yes). The most important explanatory variable was a labeling index (LI) that measures proliferative activity of cells after a patient receives an injection of tritiated thymidine. It represents the percentage of cells that are "labeled." Software reports Table for a logistic regression model using LI to predict $\pi = P(Y = 1)$.

		Standard	Likelihood Ratio		
Parameter	Estimate	Error	95% Conf.	Limits	Chi-Square
Intercept	–3.7771	1.3786	–6.9946	–1.4097	7.51
li	0.1449	0.0593	0.0425	0.2846	5.96

		LR Statistic	
Source	DF	Chi-Square	Pr>ChiSq
li	1	8.30	0.0040

Obs	li	remiss	n	pi-hat	lower	upper
1	8	0	2	0.06797	0.01121	0.31925
2	10	0	2	0.08879	0.01809	0.34010

i) Show how software obtained $\hat{\pi} = 0.068$ when LI = 8.

ii) Show that $\hat{\pi} = 0.50$ when LI = 26.0.

iii) Show that the rate of change in $\hat{\pi}$ is 0.009 when LI = 8 and is 0.036 when LI = 26.

iv) The lower quartile and upper quartile for LI are 14 and 28. Show that $\hat{\pi}$ increases by 0.42, from 0.15 to 0.57, between those values.

v) When LI increases by I, show the estimated odds of remission multiply by l.16.

b) Explain model for matched pairs also explain the types of matched pair.

✷✷✷

Total No. of Questions : 5]

SEAT No. :

PA-3431

[Total No. of Pages : 2

[5920]-47

M.Sc.

STATISTICS

ST - 44(A) : Computer Intensive Statistical Methods

(2019 Pattern) (Semester - IV) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Answer the following questions in 1-2 lines each. **[10]**

a) How many distinct jackknife samples can be drawn from a sample of size n assuming that all n observations are distinct? Justify your answer.

b) LOOCV stands for .....

c) State the properties of a kernel function.

d) For the importance sampling, what can be said about the tails of the proposal and target distribution?

e) What is the difference between stochastic EM and usual EM?

Q2) Attempt any three subquestions from the following:

a) Suppose that the data consists of regressor (X) values as 1:10 and the corresponding response (Y) values as the fourth powers of the respective regressor values. What is the predicted value for X = 3.5, by four nearest neighbours regression? **[5]**

b) What are the three types of missing data? Explain in detail. **[5]**

c) Derive the Nadaraya - Watson regression estimator of the regression function m(X) = E (Y/X). **[5]**

d) Describe the procedure to construct a confidence interval for the median of exponential distribution via parametric bootstrap. **[5]**

P.T.O.

Q3) Attempt any three subquestions from the following:

a) Explain how bagging can be carried out in a simple linear regression model and how it can be used for getting prediction intervals. **[5]**

b) Explain the procedure for multiple imputation via. chained equations. **[5]**

c) Explain the concept of complete and incomplete data in the context of EM algorithm. Is the mapping between incomplete and complete data one to one, many to one or one to many? Provide an illustration to justify your answer. **[5]**

d) Explain the procedure to estimate the value of by π Monte Carlo simulation. **[5]**

Q4) Attempt any three subquestions from the following:

a) Explain the differences between the histogram density estimator and the naive density estimator. What are the pros and cons of each of them? **[5]**

b) Describe the procedure for boosting in a classification tree. **[5]**

c) Explain the working of importance sampling. **[5]**

d) What is bias-variance tradeoff? Explain its role in non-parametric density estimation and also in k-fold cross-validation. **[5]**

Q5) Attempt any one of subquestion: **[15]**

a) What is Gibbs sampling? Explain its working in hierarchical models.

b) Describe the mechanism of a permutation test. Consider a two sample problem with $(X_1, X_2, Y_1) = (1, 9, 3)$. Compute the p-value for a permutation test for a null hypothesis $H_0 : \mu_X = \mu_Y$, versus the alternative $H_1 : \mu_X \neq \mu_Y$. **[15]**

◆ ◆ ◆