

Total No. of Questions : 5]

SEAT No. :

[Total No. of Pages : 3

PD3676

[6487]-11

First Year M.Sc.

STATISTICS

**ST-11 : Basics of Real Analysis and Calculus
(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 all the questions given below:

[5×2=10]

- a) Prove or disprove: The union of the infinitely many closed sets in $\mathbb{R}$ need not be closed.
- b) Let Q be the set of rationals with its standard metric $d: d(x, y) = |x - y|$ for any $x \in Q, y \in Q$. Let $S = \{x \in Q : x > 0 \text{ and } 2 < x^2 < 3\}$. Prove that S is closed and bounded in Q .
- c) Let $\{x_n\}$ be a real sequence such that $x_n > 0 \forall n$. Show that if x_n is converges, then its limit is non-negative.
- d) Find the limit of following:

$$\lim_{x \rightarrow \infty} \left\{ \frac{x^3}{x^2 - 1} - \frac{x^3}{x^2 + 1} \right\}$$

- e) Define the following term with an example:
 - i) Refinement of a partition.
 - ii) Lower Riemann sum.

P.T.O.

Q2) Attempt any Three of the following:

[3×5=15]

- a) State and prove Bolzano-Weistrass theorem.
- b) Define compact set. Show that every finite subset of metric space is compact.
- c) Define neighbourhood of a real number in metric space. Show that every neighbourhood is an open set.
- d) Prove or disprove: set of real numbers in $[0,1]$ is uncountable.
- e) i) Let $\{P_n\}$ be a sequence in metric space. Give the definition of the convergence of this sequence with illustration. **[2]**
ii) Prove or disprove: $\sqrt{3}$ is irrational number. **[3]**

Q3) Attempt any Three of the following:

[3×5=15]

- a) State and prove ratio test for the convergence of series. Further examine for convergence of series $\sum \frac{4^n}{n!}$
- b) Define power series. Find the radius of convergence of the following power series.
 - i) $\sum_{n=0}^{\infty} \frac{(n_x)^n}{n!}$
 - ii) $\sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(n+1)(n+2)} x^n$
- c) State and prove Cauchy mean value theorem.
- d) Show that if $f:[a,b] \rightarrow \mathbb{R}$ be bounded on $[a, b]$ then f is integrable on $[a, b]$ if and only if for each $\epsilon > 0$ there exist a partition p of $[a, b]$ such that $U(p, f) - L(p, f) < \epsilon$.
- e) Examine whether $f(x)$ is Riemann integrable or not.

$$f(x) = \begin{cases} x^2 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Q4) Attempt any Three of the following:

[3×5=15]

- a) Show that sequence in $\mathbb{R}$ converges if and only if it is cauchy.
- b) State and prove fundamental theorem of calculus.
- c) If $f(x)$ is Riemann-Stieltjes integrable $\alpha(x)$ and $p_2 \subset p_1$ then show that
$$U(p_1, f, \alpha) - L(p_1, f, \alpha) \geq U(p_2, f, \alpha) - L(p_2, f, \alpha).$$
- d) Test the convergence of $\int_a^b \frac{dx}{(x-a)^p}$ $p \geq 0$.
- e) State the gamma function. Also test its convergence.

Q5) Attempt any one of the following:

[1×15=15]

- a) i) State and prove Archimedean property of real number. Also prove that if $x \in \mathbb{R}, y \in \mathbb{R}$ and $x < y$ then there exist $p \in \mathbb{Q}$ such that $x < p < y$.
- ii) Prove or disprove: e is irrational.
- iii) If $s_1 = \sqrt{2}$ and $s_{n+1} = \sqrt{2}\sqrt{s_n}$ for $n \geq 1$, prove that $\{s_n\}$ is monotonically increasing sequence, bounded above and $\lim_{n \rightarrow \infty} s_n = 2$.

[6+4+5]

- b) i) Show that, for every real $x > 0$ and every integer $n > 0$ there is one and only one positive real y such that $y^n = x$.
- ii) Evaluate $\int_0^1 x d(x^2)$.
- iii) Show that if $s_n \rightarrow s$ and $t_n \rightarrow t$ then $s_n t_n \rightarrow st$.

[7+5+3]



Total No. of Questions : 5]

SEAT No. :

PD3677

[Total No. of Pages : 3

[6487]-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 calculator is allowed.
- 4) Symbols and abbreviations have their usual meaning.

Q1) Attempt each of the following questions :

[2 each]

- a) Define dimension of a vector space with examples.
- b) Define geometric multiplicity with the help of one example.
- c) Let A and B are square matrices of order n , Is this true that $AB = 0$ implies $A = 0$ or $B = 0$? Justify.
- d) Is this true that row rank and column rank of a matrix are equal? Justify.
- e) If A and B are two idempotent matrices of order n . Check whether $(A-B)$ is idempotent matrix.

Q2) Attempt any three questions from the following questions :

[5 each]

- a) Define vector space and its basis with an example and prove that the number of members in any one basis of a subspace is the same in any other basis.
- b) Solve the following system of linear equation and obtain all possible solutions, if it is consistent.
$$5X_1 + X_2 - 3X_3 - 4X_4 = 0$$
$$7X_1 + 5X_2 - 2X_3 - 4X_4 = 0$$
- c) Prove or disprove :
 - i) If A is symmetric matrix then its left eigen vector is same as its right eigen vector.
 - ii) If t is eigen value of matrix A then $(t + c)$ is eigen value of $(A + cI)$.
- d) Evaluate $\int_0^2 \int_0^2 e^{x+y} dy dx$ using Trapezoidal Rule. Take $h = k = 0.5$.
- e) Prove or disprove : Let A^- is generalized inverse of a matrix A then AA^- is idempotent matrix.

P.T.O.

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

- a) Use Gram-Schmidt orthogonalisation procedure on the following vectors and get an orthogonal basis for the vector space spanned by these vectors.
 $\alpha_1 = [1, 0, 0, 1]$ $\alpha_2 = [1, 2, 0, 4]$ $\alpha_3 = [1, 2, 0, -4]$
- b) Prove that for a real symmetric matrix, eigen values are real.
- c) Is this true that superset of linearly independent set of vectors is linearly dependent? Justify.
- d) State and prove Cayley-Hamilton theorem.
- e) Derive the formula for Simpson's $\frac{1}{3}$ rule for bivariate integration

$$\int_a^{a+2h} \int_b^{b+2k} f(x, y) dy dx.$$

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

- a) Define a quadratic form and describe its usual classification. Examine the nature of the following quadratic form.
 $xy + 2yz + 3xz$
- b) Obtain a Moore-Penrose generalized inverse of matrix A, where A is

$$\text{given by } A = \begin{bmatrix} 2 & 0 & 6 \\ 1 & 9 & 1 \\ -1 & 4 & -7 \end{bmatrix}.$$

- c) Prove or disprove : if A is orthogonal matrix then quadratic form of A is positive semi definite.
- d) Solve the following equations using Newton Raphson method taking initial values as

$$X_0 = 1.3 \text{ and } y_0 = 1.3$$

$$x^2 + y = 3$$

$$y^2 + x = 3$$

- e) Find the maximum and minimum values of the quadratic form $5x^2 + 2y^2 - xy$ subject to the constraint $x^2 + y^2 = 1$ and determine the values x and y at which maximum and minimum values occur.

Q5) Attempt any one questions from the following questions :

[15]

- a) i) Using spectral decomposition of a matrix, Obtain A^7 if A is given by

$$A = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}$$

- ii) Use Newton's Bivariate interpolation formula to find $f(12, 7)$ for given table of values of $f(x, y)$.

$\begin{matrix} x \\ y \end{matrix}$	0	10	20
0	1	3	7
5	3	6	11
10	2	11	17

[10+5]

- b) i) State and prove that a necessary and sufficient condition for a quadratic form to be positive definite.
- ii) Show that columns of an orthogonal matrix are linearly independent. Is the converse true? Justify.
- iii) Prove that if A is a real symmetric matrix of order n then there exists a real orthogonal matrix P such that $P'AP$ is diagonal matrix with diagonal entries are $t_1, t_2, t_3, \dots, t_n$ where $t_1, t_2, t_3, \dots, t_n$ are eigen values of A.

[5+5+5]



Total No. of Questions : 5]

SEAT No. :

PD3678

[Total No. of Pages : 3

[6487]-13

M.Sc. (Part-I)

STATISTICS

ST-13 : Probability Distributions

(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 table and scientific calculator is allowed.
- 4) Symbols and abbreviations have their usual meanings.

Q1) Attempt each of the following:

[5×2=10]

- a) Define the following terms:
 - i) Probability generating function
 - ii) Degenerate random variable
- b) Define probability measure. Verify whether the following is probability measure or not.

$$\Omega = \{a, b, c\}, \mathcal{F} = \{\phi, \{a\}, \{b\}, \{c\}, \Omega\}$$

$$P\{a\} = \frac{1}{3}, P\{b\} = \frac{2}{3}, P\{c\} = 0.$$

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

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

- d) State non-central t-distribution and give its probability density function.
- e) Obtain the covariance between X_1 and $X_{1,23} \dots p$.

Q2) Attempt any three of the following:

[3×5=15]

- a) Let $X \sim \text{Bernoulli}(p)$ and $Y \sim \text{Bernoulli}(q)$ be independent; $0 < p, q < 1$, $p+q=1$. Find joint probability mass function and cumulative distribution function for X and Y . Also find its m.g.f.
- b) State and prove characteristic properties of univariate distribution function.

P.T.O.

- c) Define symmetric random variable about point 'a'. Let X be a random variable having probability density function,

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad -\infty < x < \infty$$

Check whether $f(x)$ is symmetric about zero or not.

- d) i) If X is a random variable with probability density function,

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & , x > 0, \alpha > 0 \\ 0 & , \text{otherwise} \end{cases}$$

Find probability distribution of $Y = \alpha x$.

- ii) Let x be a $N(\mu, \sigma^2)$ random variable. Find the probability density function of X^2 .

- e) Is the function, $f(x) = \begin{cases} e^{x/2} & , x < 0 \\ 1 - e^{-x/2} & , x \geq 0 \end{cases}$

distribution function. If yes, find its probability density function.

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

- a) Let X be a random variable with probability density function,

$$f(x) = \frac{c}{(1+x^2)^m}, \quad x \in \mathbb{R}, \quad m \geq 1 \quad \text{where } c = \frac{\sqrt{m}}{\sqrt{\frac{1}{2}} \sqrt{m - \frac{1}{2}}} \quad \text{show that, } E(X^{2r})$$

exists if and only if $2r < 2m-1$.

- b) If $EX^2 < \infty$, then prove that,

$$\text{var}(X) = \text{var}\{E(X|Y)\} + E\{\text{var}(X|Y)\}$$

- c) Let X, Y be independent random variables with common probability

$$\text{density function, } f(x) = \begin{cases} \alpha e^{-x} & , x > 0, \alpha > 0 \\ 0 & , \text{otherwise} \end{cases}$$

Find the probability distribution of $x + y$ using convolution method.

- d) If $\mathfrak{C}(\underline{X}) = \Sigma$ then show that $1 - \rho_{1.23\dots p}^2 = \frac{|\Sigma|}{\sigma_{11} |\Sigma_{22}|}$ in standard notations.

- e) Let X be a random variable with probability density function,

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

Show that, a random variable X may have no moments although its moment generating function exists.

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

- a) Define Marshall-Olkin's exponential distribution. Identify marginal distribution of variables involve in it.
- b) State bivariate poisson distribution and obtain its moment generating function.
- c) Explain the concept of probability integral transformation. Let X be a random variable with cumulative distribution function $F_X(\cdot)$, if $Y = F_X(X)$, then show that, $Y \sim U(0,1)$.
- d) Let X_1, X_2, X_3 be independent gamma random variable with shape parameter $\lambda_1, \lambda_2, \lambda_3$ 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) , where $Y_1 = \frac{X_1}{Y_3}$ and $Y_2 = \frac{X_2}{Y_3}$.
- e) Derive the probability density function of non-central F-distribution. Also, express the density function with non-centrality parameter $\lambda=0$.

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

- a) 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 $W_1 = X_{(r+1)} - X_{(r)}$? **[7]**
 - ii) Describe Kolmogorov - Smirnov test procedure for testing the hypothesis $H_0 : F(x) = F_0(x)$ against $H_1 : F(x) \neq F_0(x)$. **[4]**
 - iii) Let $\underline{X} \sim Nn(\underline{0}, I_n)$. For a symmetric matrix A , show that the quadratic form $\underline{X}' A \underline{X}$ has $\chi^2_{(r)}$ distribution if and only if A is idempotent with $\rho(A) = r$. **[4]**
- b) i) Define Wilcoxon's signed rank test for population median. Obtain the null probability distribution of its test statistic and hence obtain mean and variance. **[5]**
 - ii) State and prove Fisher - Chochran theorem. **[6]**
 - iii) Show that for a random sample of size 2 from $N(0, \sigma^2)$ population,

$$E\{X_{(1)}\} = -\frac{\sigma}{\sqrt{\pi}}. \quad [4]$$



Total No. of Questions : 5]

SEAT No. :

PD3679

[Total No. of Pages : 2

[6487]-14

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 questions are compulsory.*
- 2) *Figure 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) Attempt each of the following :

[2 each]

- a) Explain stratified random sampling with real life situation.
- b) Write a short note on two-Phase sampling.
- c) Explain Lahiri's method.
- d) Write a short note on circular systematic sampling.
- e) State any two real life situation where census method is impracticable. Also write two merits of sampling.

Q2) Attempt any THREE of the following :

[5 each]

- a) Define PPS sampling design. Under this design show that $\bar{Z}_n$ is an unbiased estimator of population mean and find its variance.
- b) Under the presence of linear trend prove that linear systematic sampling is better than SRSWOR.
- c) What is two phase sampling? How it is used in ratio and regression method of estimation of population mean?
- d) Prove that under SRS the approximate value of mean squared error of $\hat{R}$ is.

$$MSE(\hat{R}) = \frac{1}{\bar{X}^2} \left[v(\bar{y}) + R^2 v(\bar{x}) - 2R\rho\sigma_{\bar{x}}\sigma_{\bar{y}} \right]$$

- e) Explain Midzuno scheme of sampling. Obtain π_i and π_{ij} under it.

P.T.O.

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

- a) Describe the stratified random sampling for proportion. Obtain variance and 95% confidence interval of stratified sample mean.
- b) In cluster sampling with cluster of unequal sizes, propose two estimators of population mean. Which estimator will you prefer? Why?
- c) Derive the expression for expected value of the simple regression estimator and its mean squared error.
- d) Explain centered systematic sampling (CSS). Obtain mean squared error of sample mean under CSS.
- e) Explain in brief Warner's Randomized Response Technique.

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

- a) Determine the sample sizes when sampling for attributes under SRSWR and SRSWOR for pre specified value of relative error.
- b) Prove that in inverse sampling $(m-1)/(n-1)$ is an unbiased estimator of population proportion.
- c) In stratified random sampling with cost function of the form $c = c_0 + \sum_{i=1}^K n_i c_i$. Prove that $n_i \propto (W_i s_i)/(c_i)^{0.5}$ for minimizing variance subject to a fixed total cost. Also obtain expression for n_i .
- d) Derive unbiased ratio type estimator of population mean. Find variance of it.
- e) Write a brief note on Murthy's estimate. Derive the expression of an estimator of population mean under it.

Q5) Attempt any ONE of the following :

- a)
 - i) Explain the concept of two dimensional systematic sampling. Give an example to obtain aligned and unaligned sample.
 - ii) Write a note on PPS systematic sampling.
 - iii) Write a note on bootstrap method of estimation.
- b)
 - i) Discuss the problem of construction of strata in stratified random sampling.
 - ii) "Under SRSWOR sample variance is an unbiased estimator of population mean square". Check whether this statement is true or false. Give justification.
 - iii) Write a brief note on Hansen-Hurwitz technique.

[7+4+4]

[7+3+5]



Total No. of Questions : 5]

SEAT No. :

PD3680

[Total No. of Pages : 3

[6487]-21

M.Sc. - I

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.

[5×2=10]

- a) Define :
 - i) Classes of independent events
 - ii) Independent of classes
- b) If $\{X_n, n \geq 1\}$ is sequence of random variables then show the $\lim X_n$ is also a random variable.
- c) Show that $X_n \xrightarrow{P} 0$ if $E|X_n|^r \rightarrow 0, (r > 0)$.
- d) State Holder's Inequality.
- e) Define
 - i) Sigma field
 - ii) Borel sigma field

P.T.O.

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

- a) If X_n is sequence of RVs then show that $U = \sup\{X_n, n \geq 1\}$ and $V = \inf\{X_n, n \geq 1\}$ are also a random variables.
- b) Let $F(x)$ be the distribution function (DF) of RV X , where.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{4} & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Obtain decomposition of DF $F(x)$.

- c) Define measure. Show that every measure is finitely additive and monotone increasing.
- d) Show that :
 - i) If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ then $X_n Y_n \xrightarrow{P} XY$.
 - ii) If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ then $\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y}$, where $P(Y_n = 0) = 0 \forall n$ and $P(Y = 0) = 0$.

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

- a) Define inverse image of a set. Let $X: \Omega_1 \rightarrow \Omega_2$ and $\mathbb{F}$ be a class of subsets of Ω_2 . Let $\mathbb{A} = X^{-1}(\mathbb{F})$. If $\mathbb{A}$ is a σ -field then prove that $\mathbb{F}$ is also a σ -field and vice-versa.
- b) If X_1 and X_2 are random variables defined on a measurable space $(\Omega, \mathbb{A})$ to $(\mathbb{R}, \mathbb{B})$ then prove that, $X_1 + X_2$, $X_1 - X_2$, $X_1 * X_2$ and $\frac{X_1^2}{X_2^2}$ (provided it is defined) are also random variables.
- c) Define distribution function (DF) of a random variable X . Let F and G be two DF's then examine (with reason) which of the following are DF's :

$$\text{i) } \frac{1}{8}F(x) + \frac{7}{8}G(x)$$

$$\text{ii) } [F(x)]^{\frac{3}{2}}$$

$$\text{iii) } 2F(x) - G(x)$$

$$\text{iv) } \frac{[F(x) + G(x)]^2}{4}$$

- d) Define probability measure. prove that
- Probability measure is subtractive.
 - Probability measure is finite sub - additive.

Q4) Attempt any three of the following.

[3×5=15]

- Let $\mathbb{C}_1 = \{(-\infty, a); a \in \mathbb{R}\}$ and $\mathbb{C}_2 = \{(a, b); a, b \in \mathbb{R}\}$. Then show that $\sigma(\mathbb{C}_1) = \sigma(\mathbb{C}_2)$.
- State and prove Holder's inequality.
- Show that $X_n \xrightarrow{p} X$ if $E \left\{ \frac{|X_n - X|}{1 + |X_n - X|} \right\} \rightarrow 0$ as $n \rightarrow \infty$.
- Prove that field is closed under intersection of two sets.
 - The intersection of an arbitrary number of fields is a field.

Q5) Attempt any one of the following.

- Define the convergence in probability, convergence in distribution.
Let $X_n \xrightarrow{L} X$ and $Y_n \xrightarrow{p} c$. Then prove that, **[8]**
 - $X_n + Y_n \rightarrow X + c$
 - $X_n Y_n \rightarrow cX$
 - State and prove continuity property of probability measure. **[7]**
- Suppose X and Y are simple random variables defined on a probability space $(\Omega, \mathbb{A}, P)$. Then show that **[8]**
 - $E(X \pm Y) = E(X) \pm E(Y)$
 - $E(cX) = cE(X)$
 - If $X \geq 0$ on Ω then $E(X) \geq 0$
 - If $X \geq 0$ a.s. then $E(X) \geq 0$
 - State and prove Kolmogorov 0-1 law. **[7]**



Total No. of Questions : 5]

SEAT No. :

PD3681

[Total No. of Pages : 3

[6487]-22

First Year M.Sc.

STATISTICS

ST 22 : Regression Analysis

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

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) For simple linear regression model, show that, $SS_{reg} = \hat{\beta}_1 s_{xy}$.
- b) Obtain the relationship of analysis of variance test and coefficient of determination in case of multiple linear regression model.
- c) Determine whether the model, $y = \theta_1 e^{\theta_2 + \theta_3 x} + \varepsilon$ is a linear model, intrinsically linear model or a non-linear model. If the model is intrinsically linear, then linearize it by suitable transformation.
- d) Prove that the least square estimates for the slope and intercept minimize the sum of squared residuals.
- e) For the given data we obtain the eigenvalues of $X'X$ matrix as $\lambda_1 = 3.169, \lambda_2 = 1.006, \lambda_3 = 0.763, \lambda_4 = 0.553, \lambda_5 = 0.317, \lambda_6 = 0.192$. Calculate the condition number and state whether the problem of multicollinearity is harmful or not?

Q2) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Explain the reverse regression method with a suitable example. Also, obtain the least square estimators of the regression coefficients for it.
- b) State and prove Gauss-Markoff theorem.

P.T.O.

- c) Prove that the maximum value of R^2 is less than 1 if the data contain repeated observations on y at the same value of x .
- d) Suppose that we fit the straight-line regression model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$ but the response is affected by the second variable x_2 such that the true regression function is $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. Is the least-squares estimator of the slope in the original simple linear regression model unbiased? Also, show that the bias in $\hat{\beta}_1$.

Q3) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Discuss the method of principle component regression for dealing with the problem of multicollinearity.
- b) Explain the measure of influence Cook's D-Statistic.
- c) Obtain the bias, covariance and mean squared error of ridge regression estimator.
- d) For a regression model with p terms including the intercept, prove that
$$\sum_{i=1}^n \text{var}(\hat{y}_i) = p\sigma^2.$$

Q4) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Explain the Gauss-Newton iteration method of parameter estimation in nonlinear regression models.
- b) Discuss Wald test in logistic regression model.
- c) If Z is the $n \times k$ matrix of standardized regressors and T is the $k \times k$ upper triangular matrix, show that the transformed regressors $W = ZT^{-1}$ are orthogonal and have unit variance.
- d) Show that the least squares estimate of β with the i^{th} observation deleted can be written in terms of the estimate based on all n points as

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{e_i}{1 - h_{ii}} (X'X)^{-1} x_i.$$

Q5) Attempt any 1 question out of 2 questions.

[1×15=15]

- a) i) Describe the generalized linear model. Define link function and obtain it for poisson distribution. **[4]**
- ii) Describe the method of maximum likelihood estimation to estimate the parameters of the poisson regression model. **[6]**
- iii) Explain the testing procedure for poisson regression model based on a large sample test using the likelihood ratio test statistic. **[5]**
- b) Define the following terms:
- i) • Estimation space and estimability of parametric function
• Full rank model
• Best Linear Unbiased estimator (BLUE) **[3]**
- ii) Consider the model $y_i = \mu_i + \varepsilon_i, i = 1, 2, \dots, n$ where the parameters μ_i subject to condition $\sum_{i=1}^n \mu_i = 0$. Obtain normal equations and their solutions. Is μ_i is estimable? **[6]**
- iii) For the multiple linear regression models, obtain BLUE of vector of regression parameters and also obtain its variance. **[6]**



Total No. of Questions : 5]

SEAT No. :

PD3682

[Total No. of Pages : 3

[6487]-23

M.Sc. - I

STATISTICS

ST - 23 : Statistical Inference - I

(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) Define the following terms with an example:
 - i) Unbiased estimator
 - ii) Ancillary statistics
- b) Define conjugate prior with an illustration.
- c) State Rao-Blackwell Lehmann-Scheffe' theorem.
- d) State Chapman-Robin bounds.
- e) Prove or disprove: 'UMP test always exists'.

Q2) Attempt any THREE of the following:

[3×5=15]

- a) If $X_1, X_2, \dots, X_n$ be a random sample from Bernoulli (θ) and $T = \sum_{i=1}^n X_i$
then show that $\frac{T(n-T)}{n(n-1)}$ is an unbiased estimator of $\theta(1 - \theta)$.
- b) State the factorization criterion for a sufficient statistic. Using it show that the order statistic is a sufficient statistic for any parametric family of continuous distributions.

P.T.O.

- c) Define complete statistic. Let $X_1, X_2, \dots, X_n$ be *i.i.d.* $U(\theta, \theta + 1)$. Show that $(X_{(1)}, X_{(n)})$ is sufficient for θ but not complete for θ .
- d) Let $X_1, X_2, \dots, X_n$ be *i.i.d.* random variables with *pdf* $f(x; \theta), \theta \in \Theta \subseteq \mathbb{R}$. Let $T = T(X_1, X_2, \dots, X_n)$ be a statistic. Show that $I_T(\theta) \leq I_{(X_1, X_2, \dots, X_n)}(\theta)$ with the equality holds if and only if T is sufficient for θ .

Q3) Attempt any THREE of the following:

[3×5=15]

- a) State and prove Rao Blackwell theorem.
- b) Let $X_1, X_2, \dots, X_n$ be independent uniformly distributed over $(-i\theta, i\theta), \theta > 0, i = 1, 2, \dots, n$. Obtain minimal sufficient statistic for θ .
- c) State and prove Cramér-Rao Lower Bound.
- d) Let $X_1, X_2, \dots, X_n$ be a random sample from Gamma $\left(1, \frac{1}{\lambda}\right)$. To estimate λ , let the prior *pdf* on λ be $\pi(\lambda) = e^{-\lambda}$ if $\lambda > 0$ and the loss function be squared error. Find the Bayes estimator for λ .

Q4) Attempt any THREE of the following:

[3×5=15]

- a) 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?
- b) Define ancillary statistic with illustration. Prove that, any ancillary statistic is independent of complete sufficient statistic.
- c) What is pivotal quantity? Explain its use in constructing of confidence intervals.
- d) Define monotone likelihood ratio property of a distribution. Show that the binomial distribution satisfies MLR property.

Q5) Attempt any ONE of the following:

[1×15=15]

- a) i) Define shortest expected length confidence interval (SEL CI). Let $X_1, X_2, \dots, X_n$ be random sample from $U(0, \theta), \theta > 0$. Find $(1 - \alpha)100\%$ SEL CI for θ . **[7]**
- ii) Define uniformly most accurate confidence interval. Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\theta, \sigma^2)$, σ^2 is known. Obtain level α uniformly most accurate lower bound on θ . **[8]**
- b) i) State and prove Neyman-Pearson Lemma. Let $X_1, X_2, \dots, X_n$ be a random sample from $f(x; \theta) = \frac{\theta}{x^2}$, where $0 < \theta \leq x < \infty$. Find an MP test of its size for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ ($\theta_0 < \theta_1$). **[9]**
- ii) Let $f(x, y, \lambda, p) = \binom{x}{y} p^y (1-p)^{(x-y)} \frac{e^{-\lambda} \lambda^x}{x!}, y = 0, 1, 2, \dots, x; x = 0, 1, 2, \dots, \lambda > 0$ and $0 < p < 1$ be joint pmf of (X, Y) . Show that it belongs to two parameter exponential family and obtain minimal sufficient statistic $(T_1, T_2)'$ for $(p, \lambda)'$. **[6]**



[6487]-24

First Year M.Sc.

STATISTICS

ST-24: Multivariate 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 each of the following.

[5×2=10]

- a) Define Multivariate Normal Distribution (MVND). Write characteristic function of MVND.
- b) Define the following:
 - i) Wishart matrix
 - ii) Wishart distribution
- c) Let $\underline{X}$ have variance-covariance matrix $\Sigma = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$ Find the correlation between X_1 and $\frac{X_2}{2} + \frac{X_3}{2}$.
- d) Distinguish between t-test and Hotelling T^2 statistic.
- e) State the additive property of Wishart distribution.

Q2) Attempt any three of the following.

[3×5=15]

- a) Let $\underline{X} \sim N_3(\underline{\mu}, \Sigma)$ where $\underline{\mu} = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$ Obtain the conditional distribution of $X_1 / \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$.
- b) Show that the principal components are uncorrelated and have variances equal to the eigen values of Σ .

P.T.O.

- c) Let $\underline{X} \sim N_p(\underline{\mu}, \underline{\Sigma})$ and $\underline{X}$ is partitioned as $\begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$, then obtain the marginal distributions of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$.
- d) Consider the following density function:

$$f(x, y) = \frac{1}{2\pi} \exp \left\{ \frac{-1}{2} (x^2 + y^2 + 4x - 6y + 13) \right\}$$
 Obtain the distribution of $\begin{bmatrix} X \\ Y \end{bmatrix}$.

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

- a) Derive the sampling distribution of the maximum likelihood estimators of the parameter of the multivariate normal distribution.
- b) If $\underline{X} \sim N_p(\underline{\mu}, \underline{\Sigma})$ and $\underline{X}$ is partitioned as $\begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$, then prove that $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are independent iff $\Sigma_{12} = 0$.
- c) What are principal components? What purpose do they serve? Determine the first principal component and its variance when $\Sigma = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$.
- d) Define the problem of classifying an observation into one of the two populations. Define linear discriminant function and probabilities of misclassification.

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

- a) If $D \sim W_p(n, \Sigma)$ and D is partitioned as $\begin{bmatrix} D_{11(m \times m)} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$. Then show that $D_{11} \sim W_p(n, \Sigma_{11})$.
- b) Evaluate Hotelling's T^2 for testing $H_0: \underline{\mu} = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$ using the data $\begin{bmatrix} 6 & 10 & 8 \\ 9 & 6 & 3 \end{bmatrix}$. Specify the distribution of T^2 for this situation. Test H_0 at $\alpha = 0.05$ level. What conclusion do you reach?
- c) Write a note on:
 i) K-means clustering
 ii) Single Linkage Method
- d) Define canonical correlations and canonical variables. Derive the first canonical correlation coefficient and the corresponding canonical variables. Also show that first canonical correlation explains maximum correlation.

Q5) Attempt any one of the following:

a) i) Let $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \sim N_4(\underline{\mu}, \Sigma)$ with $\underline{\mu} = \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 3 & 0 & 2 & 2 \\ & 1 & 1 & 0 \\ & & 9 & -2 \\ & & & 4 \end{bmatrix}$, then find

the distribution of: [8]

1) $[X_1 - 2X_2 + 3X_3 + X_4]$

2) $\begin{bmatrix} X_1 - 2X_2 + X_3 + X_4 \\ 7X_1 - 2X_2 + 3X_3 \end{bmatrix}$

ii) Let $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$ be a random sample from $N_p(\underline{0}, I_p)$. Obtain the distribution of Wishart matrix in canonical case. [7]

b) i) Analyse the following data by using One Way MANOVA and draw the conclusion: [8]

Treatment Observations

1) $\begin{bmatrix} 6 \\ 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

2) $\begin{bmatrix} 3 \\ 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \end{bmatrix} \begin{bmatrix} 6 \\ 7 \end{bmatrix}$

3) $\begin{bmatrix} 2 \\ 8 \end{bmatrix} \begin{bmatrix} 6 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 7 \end{bmatrix}$

ii) Show that LRT for testing $H_0 : \underline{\mu} = \underline{\mu}_0$ leads to Hotelling T^2 -Statistic for $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$. [7]



Total No. of Questions : 5]

SEAT No. :

PD3684

[Total No. of Pages : 3

[6487]-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 all questions.

[5×2=10]

- a) What is a Markov property?
- b) Explain stopping time variable with an example.
- c) Define:
 - i) Recurrent state
 - ii) Transient state
- d) Explain the term stochastic processes along with its applications.
- e) Define stationary distribution of a Markov chain and state the situation in which it is unique.

Q2) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Explain the terms with illustrations:
 - i) Markov chain
 - ii) Time homogeneous Markov chain
 - iii) Finite dimensional distributions of Markov chain
 - iv) Initial distribution of a Markov chain
 - v) One-dimensional random walk

P.T.O.

- b) Explain the compound Poisson process with its application. Obtain the mean and variance for compound Poisson process.
- c) Explain branching process and calculate the probability of extinction π_0 for

A) $P_0 = \frac{1}{6}, P_1 = \frac{1}{2}, P_3 = \frac{1}{3}$

B) $P_0 = \frac{1}{4}, P_1 = \frac{1}{2}, P_3 = \frac{1}{4}$

- d) Prove or disprove: Every aperiodic Markov chain is also irreducible.

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

- a) Prove or disprove: For a markov chain periodicity is a class property.
- b) State and Prove Chapman Kolmogorov Equation.
- c) Define branching processes $\{Z_n; n = 0, 1, 2, \dots\}$ and obtain its corresponding offspring distribution. Also, obtain its mean and variance.
- d) Let $\{X_n, n > 0\}$ be a markov chain with state space $S = \{0, 1, 2\}$ 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: Persistent null, Persistent non-null, transient.

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

- a) State and prove the first entrance theorem.
- b) A rat is trapped in maze, initially he has to choose one of the two directions if he goes to the right then he will wonder around in the maze for 3 minutes and will then return to his initial position. If he goes to left then with probability $1/3$ he will depart from the maze after two minutes of traveling and with probability $2/3$ he will return to his initial position after 5 minutes of traveling. Assume that the rat is at all times equally likely to go to the left or the right. What is the expected no. of minutes that he will be trapped in the maze?
- c) Give the two definitions of Poisson process and establish their equivalence.
- d) Vehicle stopping at roadside restaurant form a Poisson process with rate $\lambda = 20/\text{hours}$. 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 1 hour.

Q5) Attempt any 1 question out of 2 questions.

[1×15=15]

- a) i) Define a continuous time Markov process with state space as the set of non-integers. State Kolmogorov's backward and forward equations. Using these obtain a system of equations for limiting probabilities of Birth-death process, under suitable condition to be stated. **[7]**
- ii) The mean value function $m(t)$ of the renewal process $\{N(t) \geq 0\}$ is given by $m(t) = t/4$. Using this information find $P\{N(5) = 3\}$ **[4]**
- iii) Define the following terms: **[4]**
- A) Counting process
- B) Regenerative process
- C) Stopping time random variable
- D) Yule process
- b) i) Let $P(S)$ be the pgf of the offspring distribution and $X_0 = 1$ and $P_n(S)$ be the pgf of branching process $\{X_n, n \geq 0\}$. Show that, $P_n(S) = P_{n-1}(P(s)) = P(P_{n-1}(S))$. **[4]**
- ii) Derive **[4]**
- A) Wald equation
- B) Kolmogorov forward equation
- iii) Define Yule process. For Yule process show that, **[7]**
- A) Probability that starting with a single individual, the population size at time t will have a geometric distribution with mean $e^{\lambda t}$.
- B) Probability that starting with i individuals, the population size at time t will have negative binomial distribution.



Total No. of Questions : 5]

SEAT No. :

[Total No. of Pages : 2

PD3685

[6487]-32

S.Y. M.Sc.

STATISTICS

ST-32(A) : Bayesian Inference
(2019 Pattern) (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.

[5×2=10]

- a) Define Prior and Posterior distribution. State the relationship between prior, posterior and likelihood, given by Bayes.
- b) Explain the application of Bayes theorem for performing hypothesis test.
- c) Explain the term subjective probability and frequentist probability.
- d) Let $X \sim B(n, \theta)$ and prior distribution of θ is $\beta_1(a, b)$. Find the posterior distribution of θ .
- e) What is credible interval? Is it unique?

Q2) Attempt any three of the following.

[3×5=15]

- a) Let X be a continuous random variable with pdf $f(x|\theta) = \frac{3\theta^3}{x^4}, x > \theta$.
Let the prior pdf of θ be given by $\pi(\theta) = e^{-(\theta-1)}, \theta > 1$. Compute the posterior distribution of θ .
- b) Write short note on informative prior and non-informative prior.
- c) Define zero-one (all-or-nothing) loss function and obtain the Bayes estimator under zero-one (all-or-nothing) loss function.
- d) Let $X_1, X_2, \dots, X_n$ be a random sample of size n from $N(\mu, \sigma^2), \sigma^2$ known. Suppose $\mu \sim N(\mu_0, \sigma_0^2)$. Obtain 100(1 - α)% highest posterior density (HPD) confidence interval for μ .

P.T.O.

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

- a) Define the Jeffrey's prior. Let $X \sim \text{Binomial}(n, p)$, Find the Jeffrey's prior of p .
- b) Discuss MCMC algorithm in briefly.
- c) Write note on EM algorithm.
- d) 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 likelihood (ML) estimate of X given $Y = y$.

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

- a) Describe the any three application of Bayesian inference.
- b) Discuss conjugate class of priors with an example.
- c) Define Bayes estimate. Let $X_1, X_2, \dots, X_n$ be a random sample from $P(\lambda)$. 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 .
- d) Explain the terms : Decision function and Loss function.

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

- a)
 - i) Define the highest posterior density credible interval (HPDCI) for a real valued parameter θ . Assuming that the posterior distribution of θ to be symmetric and unimodal, obtain HPDCI for θ . **[7]**
 - ii) Explain Bayesian information criterion (BIC) with illustration. **[8]**
- b)
 - i) Discuss Gibbs sampling in briefly. **[7]**
 - ii) Explain the terms : subjective priors and probability matching prior. **[8]**



Total No. of Questions : 5]

SEAT No. :

PD3686

[Total No. of Pages : 3

[6487]-33

M.Sc. - II

STATISTICS

ST-33 : Design and Analysis of Experiments

(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 table and scientific calculator is allowed.*
- 4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt each of the following:

[5×2=10]

- a) In a single factor ANOVA problem involving five treatments, with a random sample of four observations from each one, it is found that $SST_r = 16.1408$ and $SSE = 37.3801$. Then obtain the value of test statistic.
- b) Check whether following block design is connected:

Block-I Block-II

1	1
2	4
3	5

- c) Write the second order response surface model.
- d) Define the following:
 - i) BIBD
 - ii) Symmetric BIBD
- e) Explain the term confounding.

P.T.O.

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

- Explain One-way ANOVA with repeated measures.
- Give the statistical analysis of a 2^3 factorial experiment in which 4 replicates are used and AC is totally confounded.
- Construct one-quarter fraction of the 2^6 design with highest possible resolution. Write down its alias structure.
- What is Simplex Lattice Design. Explain a $[3, 3]$ Simplex Lattice Design.

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

- An teacher sets three question papers and distribute them to randomly to their students. After evaluation of answer book following marks are obtained to students:

Test	Test Marks
Test-I	63, 64, 95, 64, 60, 85
Test-II	58, 56, 51, 84, 77
Test-III	100, 79, 82, 80, 74, 97

Use Levene's test to check the assumption of homogeneity of variances.

- In BIBD, prove the Fishers inequality.
- Give the statistical analysis of 3^2 factorial experiment.
- Write a note on method of steepest ascent and steepest descent.

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

- Explain following method for comparing pairs of treatment means:
 - Dunnet's Test
 - Duncan's Multiple Range Test
- Obtain parameters of the following PBIBD:

Blocks	Treatments		
1	1	2	6
2	1	6	4
3	3	4	6
4	1	2	5
5	2	5	3
6	3	5	4

- Confound AB^2C^2 in 3^3 factorial experiment in 3 blocks.
- Write a note on Signal-to-Noise ratio.

Q5) Attempt any one of the following:

- a) i) Analyze the following data: [8]

Replicate-I		Replicate-II	
c = -1	'1' = -3	abc = 5	ac = 1
b = -1	ab = 2	'1' = -1	a = 1
abc = 6	bc = 1	c = 0	bc = 1
a = 0	ac = 2	ab = 3	b = 0

- ii) Show that BIBD is connected block design. [7]
- b) i) The incidence matrix of a block design is given below. Check whether the design is connected and/or orthogonal and/or variance balanced. : [8]

$$N = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

- ii) Explain Response Surface Methodology. Differentiate between Circumscribed(CCC), Face Centered(CCF) and Inscribed(CCI) designs of CCD for number of factors k=2. [7]



Total No. of Questions : 5]

SEAT No. :

PD3687

[Total No. of Pages : 2

[6487]-34

S.Y. 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) What are the limitations of the perceptron?
- b) Define Single Linkage in Agglomerative Hierarchical Clustering.
- c) What is ensemble learning?
- d) Define supervised machine learning with example.
- e) What is cross validation?

Q2) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Explain the k-means clustering algorithm. How does it work, and what are its advantages and limitations?
- b) What are basic steps in Apriori algorithm?
- c) Describe the backpropagation algorithm and its role in training neural networks.
- d) Consider a football game between two rival teams: Team A and Team B. Suppose Team A wins 65% of the time and Team B wins the remaining matches. Among the games won by Team A, only 35% of them come from playing on Team B 's football field. On the other hand, 75% of the victories for Team B are obtained while playing at home. If Team B is to host the next match between the two teams, which team will most likely emerge as the winner?

P.T.O.

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

- What is K-nearest neighbor (KNN) algorithm and explain how does it works?
- What is agglomerative hierarchical clustering, and how does it differ from k-means clustering?
- What are hard-margin and soft margin support vector machines?
- Explain Biological Neural Network and Artificial Neural Network.

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

- Explain association analysis and describe a real-world application of it.
- What are impurity measures in Decision Tree Learning, and how are they used to build a decision tree?
- What is a confusion matrix explain all the term use in it.
- Explain the basic components of learning process.

Q5) Attempt any 1 question out of 2 questions. [15]

- Explain the different linkage methods used in the Hierarchical clustering Algorithm. [6]
 - A company wants to classify emails as either Spam or Not Spam based on certain features. The company uses two features: Contains “Offer” (whether the word “offer” appears in the email) and Contains “Click” (whether the word “click” appears in the email). For the following data from past emails. Classify a new email that contains the word “Offer” and does not contain the word “Click” using the Naive Bayes classifier. [9]

Email	1	2	3	4	5	6
Contains “Offer”	Yes	Yes	No	Yes	Yes	No
Contains “Click”	Yes	No	Yes	Yes	Yes	No
Spam/Not Spam	Spam	Spam	Not Spam	Not Spam	Spam	Not Spam

- What do you mean by random Forest Algorithm and explain how it works? [6]
 - For the following dataset of 8 points, with two features X_1 and X_2 . Classify the of new point ($X_1 = 4, X_2 = 4$) by using K- Nearest Neighbor algorithm with $K = 3$. [9]

Point	P1	P2	P3	P4	P5	P6	P7	P8
X_1	1	2	3	6	7	8	3	2
X_2	1	2	3	5	7	6	5	4
Class	1	0	0	1	0	0	1	1



Total No. of Questions : 5]

SEAT No. :

PD3688

[Total No. of Pages : 2

[6487]-35

M.Sc. - II

STATISTICS

ST-32(B) : Statistical Quality Control
(2019 Pattern) (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.

[5×2=10]

- a) What are the disadvantages of implementing V-mask procedure?
- b) Write a short note on Six-sigma methodology.
- c) State the relationship between capability index C_p and performance index C_{pk} .
- d) Find sample size required for p -chart, so that there will be at least 95% samples have one or more defectives. Give that process fraction defective is 0.05.
- e) Describe Nelson control chart for low defect counts.

Q2) Attempt any 3 questions out of 4 questions :

[3×5=15]

- a) Define the term quality and give different dimensions of quality.
- b) 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.
- c) The following is $\bar{X}$ and S chart based on $n = 4$ and 3-sigma control limits :

$\bar{X}$ -chart	S-chart
UCL = 710	UCL = 18.08
CL = 700	CL = 7.979
LCL = 690	LCL = 0

- i) Estimate parameters μ and σ .
- ii) If the specifications are at 705 ± 15 and process output is normally distributed, estimate the fraction non-confirming.
- iii) For the $\bar{X}$ control chart, find the probability of type-I error, assuming σ is constant. (Take $C_4 = 0.9213$ and $n = 4$).
- d) Obtain the Average Run Length (ARL) for :
 - i) Shewart control chart
 - ii) CUSUM chart

P.T.O.

Q3) Attempt any 3 questions out of 4 questions :

[3×5=15]

- a) Give the comparison between Shewhart chart and CUSUM chart V-mask procedure.
- b) Discuss EWMA control chart for monitoring process mean and variance. Explain how it is better than Shewhart $\bar{X} - R$ control chart.
- c) Explain the construction and working of Confirming Run Length (CRL) chart for process fraction defective.
- d) A process is in control with $\bar{\bar{X}} = 100$ and $\bar{S} = 1.05, n = 5$. The process specifications are at 95 ± 10 . The quality characteristics has normal distribution. Compute C_p , C_{pk} and C_{pm} . Interpret these ratios.

Q4) 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 probability of chart statistic and find the expression for ARL(0).
- b) Explain the criteria for detecting lack of control process.
- c) Explain the constructions of control chart for residuals after fitting first order autocorrelated model.
- d) Define double sampling plan. Obtain its OC and ASN functions.

Q5) Attempt any 1 question out of 2 questions :

[1×15=15]

- a) Using CUSUM chart, check whether the following process is under control or not, if the target value is 175, $k = 2$, $h = 4.77$, $\hat{\sigma} = 20$? The observations are as follows. 160, 186, 190, 250, 158, 195, 135, 285, 215, 150. **[15]**
- b)
 - 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]



Total No. of Questions : 5]

SEAT No. :

PD3689

[Total No. of Pages : 2

[6487]-41

S.Y. 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) Attempt all questions.

[5×2=10]

- a) Define CAN estimator with illustration.
- b) Define joint consistency and marginal consistency.
- c) State one parameter Canonical Form.
- d) How are asymptotic confidence intervals constructed?
- e) Let $X_1, X_2, \dots, X_n$ is random sample from $N(\theta, 1)$. Suppose $\theta \in \{0, 1\}$, check whether sample mean is a consistent estimator of θ .

Q2) Attempt any THREE of the following:

[3×5=15]

- a) Let $X_1, X_2, \dots, X_n$ is random sample from truncated Poisson distribution truncated at 0. Derive likelihood ratio test and Wald test to test $H_0 : \theta = \theta_0$ Vs $H_1 : \theta \neq \theta_0$
- b) Explain method of scoring with example.
- c) Find maximum likelihood estimate of σ based on one observation when $X \sim \text{Lognormal}(0, \sigma^2)$.
- d) Explain in detail locally most powerful test with example.

P.T.O.

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

- a) Show that joint consistency is equivalent to marginal consistency.
- b) Let $X_1, X_2, \dots, X_n$ is random sample from Binomial (1, θ). What is the asymptotic distribution of $\bar{x}(1 - \bar{x})$ at $\theta = \frac{1}{2}$? Justify it.
- c) Let $X_1, X_2, \dots, X_n$ is random sample from $U(\theta - 1, \theta + 1)$. Find Maximum likelihood estimate of θ .
- d) Under crammer regularity conditions show that likelihood equation estimator is CAN for θ .

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

- a) Consider density function $f(x, \theta) = \theta x^\theta e^{-x^\theta}; x > 0$ and $\theta > 0$. Obtain CAN estimator for θ based on percentile method. Also obtain its asymptotic variance.
- b) Define most powerful test. Describe it in detail.
- c) Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\theta, 1)$. Obtain $100(1 - \alpha)\%$ shortest expected length confidence interval for θ .
- d) Let $X_1, X_2, \dots, X_n$ is random sample from Cauchy(μ, λ). Find consistent estimator for (μ, λ).

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

- a) i) Let $y = \alpha + \beta x + \epsilon$ be a linear model with $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2$
then show that $\hat{\beta}_n = \frac{S_{xy}}{S_{xx}} \xrightarrow{P} \beta$ and $\hat{\alpha}_n = \bar{y}_n - \hat{\beta}_n \bar{x}_n \xrightarrow{P} \alpha$. [8]
- ii) Write a note on Newton Raphson Method to obtain MLE with one example. [7]
- b) i) Explain Bartlett's test for homogeneity of variance. [8]
- ii) What is need of Variance Stabilizing Transformation (VST)? Construct 95% VST confidence interval for Exponential (mean = θ). [7]



Total No. of Questions : 5]

SEAT No. :

PD3690

[Total No. of Pages : 2

[6487]-42

M.Sc. - II

STATISTICS

**ST-44 (B): Analysis of Clinical Trials
(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) What is patient compliance?
- b) Define:
 - i) randomization model
 - ii) titration model
- c) Write a note on patient selection.
- d) What are the primary safety variables?
- e) Explain the term run-in periods.

Q2) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Explain advantages and disadvantages of Treatment and Response Adaptive randomization.
- b) Describe all phases of clinical development in clinical trials.
- c) What is Interim analysis and safety report in clinical trial?
- d) Explain the role of sampling distributions for the valid and unbiased assessment of true efficacy and safety of the study medication.

Q3) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Explain the method of permuted block randomization with example and its advantages over complete randomization.
- b) What is population model and invoke population model? State the difference between them.
- c) Write a note on:
 - i) Target population
 - ii) Patient selection
- d) What is patient compliance? What is the difference between missing values and dropouts?

P.T.O.

Q4) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) The plasma log concentration and time points of observation are related as $\log(C_i) = 0.038 - 2.13 t_i$ and $AUC_{(0-t)} = 132.13$, $C_{last} = 2.8$ then calculate PK-parameters: $AUC_{(0-\infty)}$, $AUC_{(t-\infty)}$ and K_e .
- b) Suggest the model and explain test procedure for assessment of overall average drug effect across time in repeated measures.
- c) Discuss: Parallel designs useful in Clinical Trials with appropriate layout.
- d) Define the following terms:
 - i) Intersubject variability
 - ii) Treatment IND

Q5) Attempt any 1 question out of 2 questions.

[1×15=15]

- a)
 - i) Explain the role of sampling distributions for the valid and unbiased assessment of true efficacy and safety of the study medication. **[6]**
 - ii) Classify the Clinical Trials depending upon their functioning. Explain their respective functions in brief. **[9]**
- b)
 - i) A single dose of a drug was given to a 50 kg person at a dose level of 10mg/kg. Blood samples were collected periodically. Estimate all possible pharmacokinetic parameters using the following data. **[9]**

	1	3	5	7	10
Drug level in Blood sample (Drug A)	20.00	11.30	7.00	4.30	2.00
Drug level in Blood sample (Drug A)	19.55	12.65	9.15	5.14	4.32

- ii) Explain classical confidence interval for testing the average bioequivalence. **[6]**



Total No. of Questions : 5]

SEAT No. :

PD3691

[Total No. of Pages : 2

[6487]-43

M.Sc.-II

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) Symbols and abbreviations have their usual meaning.

Q1) Attempt each of the following:

[5×2=10]

- a) Define the following:
 - i) Sample ACVF
 - ii) Sample ACF
- b) Define AR(1) process. Give its ACVF.
- c) Define SARIMA (0, 1, 1) X (0, 1, 1)₁₂ model.
- d) Define AICc and BIC.
- e) Write the model of ARIMA (p,d,q) process.

Q2) Attempt any three of the following.

[3×5=15]

- a) Discuss the following methods for eliminating trend in absence of seasonality:
 - i) Moving-Average Method
 - ii) Exponential smoothing method
- b) Obtain ACF of AR(2) process.
- c) Obtain ACF of the following time series:
 - i) $X_t = Z_t + 0.3Z_{t-1} - 0.4Z_{t-2}$ where $\{Z_t\} \sim WN(0,1)$
 - ii) $X_t = Z_t - 1.2Z_{t-1} - 1.6Z_{t-2}$ Where $\{Z_t\} \sim WN(0,0.25)$
- d) Let $\{Z_t\}$ be a sequence of independent normal random variables with mean 0 and variance σ^2 . Let a, b and c are constants then check whether the following process is stationary, specify its mean and ACF:

$$X_t = a + bZ_{t-1} + cZ_{t-2}$$

P.T.O.

Q3) Attempt any three of the following.

[3×5=15]

- a) Draw correlogram for the following data.

t	1	2	3	4
X_t	6	8	5	12

- b) Discuss the differencing technique in time series analysis. When does it is used? Hence define the differencing operators.

- c) Let $\{Y_t\}$ be a stationary time series with mean 0 and covariance $v_y(h)$. If

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \text{ then show that the time series } X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} \text{ is stationary}$$

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.

- d) What is spectral density function? State and prove its properties.

Q4) Attempt any three of the following.

[3×5=15]

- a) Explain the following tests for testing the estimated noise sequence.

- The portmanteau test
- The rank test

- b) 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 $\{Z_t\} \sim \text{WN}(0, \sigma^2)$

ii) $X_t = 0.7X_{t-1} - 0.1X_{t-2} + Z_t$, Where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$

- c) Explain the Innovation Algorithm.

- d) Discuss the simultaneous equation models: endogenous and exogenous models.

Q5) Attempt any one of the following.

- a) i) Discuss the following smoothing techniques in time series analysis.

[8]

- Holt-Winters' smoothing technique
- Adaptive smoothing

- ii) Determine the ψ_j and π_j coefficients of the ARMA(p, q) Process.

[7]

- b) i) Obtain h-step best linear predictor of AR(1) process. Also obtain its mean square error.

[8]

- ii) Explain forecasting in time series. Obtain best linear predictor X_{n+m} ; $m=1, 2, \dots$ of a stationary time series with mean μ and ACVF $v_x(h)$.

[7]



Total No. of Questions : 5]

SEAT No. :

PD3692

[Total No. of Pages : 2

[6487]-44

M.Sc.-II

STATISTICS

ST-42(B) : Operation Research

(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) Attempt all questions.

[5×2=10]

- a) What is meant by transshipment problem?
- b) Give any two applications of dual simplex method.
- c) Define dynamic programming problem (DPP).
- d) What is restricted assignment problem?
- e) Explain the terms balking and reneging of queuing theory.

Q2) Attempt any three of the following.

[3×5=15]

- a) Show that if a primal has infeasible solution then dual need not have unbounded solution.
- b) Define mixed integer programming problem (IPP). Derive Gomory's cut for it .
- c) Write a note on activity on nodes and activity on arrows.
- d) Obtain the value of given problem by using dynamic programming:

$$\text{Min } Z = y_1^2 + y_2^2 + y_3^2$$

$$\text{Subject to the constraint } y_1 + y_2 + y_3 = 10$$

$$\text{and } y_1, y_2, y_3 \geq 0.$$

P.T.O.

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

- a) Discuss any three types of simulation.
- b) Formulate the maximal flow mode to the linear programming.
- c) State and prove reduction theorem of assignment problem.
- d) Explain the additive separable return function and single additive constraint model of dynamic programming.

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

- a) State and prove modified slackness property of duality.
- b) What is the use of queuing theory. State the terminologies of M/M/1: FCFS/a/b (a & b are finite) of queuing theory.
- c) Discuss the Monte Carlo simulation method.
- d) Write a note on travelling salesman problem.

Q5) Attempt any one of the following.

- a) i) Discuss the revised simplex method. [5]
ii) In a railway marshalling yard, goods trains arrive at a rate of 30 trains per day. Assuming that the inter arrival time follows an exponential distribution and the service time (the time taken to hump a train) distribution is also exponential with an average of 36 minutes. Calculate:
 - 1) Expected queue size (line length)
 - 2) Probability that the queue size exceeds 10If the input of trains increases to an average of 33 per day, what will be the change in (A) and (B)? [7]
iii) Explain the looping method in transportation problem. [3]
- b) i) A bakery keeps stock of a popular brand of cake. Previous experience shows the daily demand pattern for the item with associated probabilities, as given below:

Daily demand :	0	10	20	30	40	50
Probability:	0.01	0.20	0.15	0.50	0.12	0.02

Use the following sequence of random numbers to simulate the demand for next 10 days.
Random numbers: 25, 39, 65, 76, 12, 05, 73, 89, 19, 49.
Also estimate the daily average demand for the cakes on the basis of the simulated data [5]
ii) Explain the queuing process. [5]
iii) Write the procedure of Hungarian method of assignment problem. [5]



Total No. of Questions : 5]

SEAT No. :

PD3693

[Total No. of Pages : 2

[6487]-45

M.Sc. - II

STATISTICS

ST 43(A) : Survival Analysis

(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 calculator is allowed.*
- 4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt the questions.

[10]

- a) Obtain hazard rate for the Makeham family.
- b) Define the following terms
 - i) Mean Residual life Function
 - ii) Cumulative hazard rate
- c) State one sample U-statistic theorem
- d) Suppose 25 ceramic capacitors from exponential distribution are put on life test. In order to reduce the test time, the test is terminated after 15 ceramic capacitors fails. Following is lifetime in years of ceramic capacitors. 3.4, 3.7, 4.1, 4.8, 5.1, 5.2, 6.2, 7.3, 7.9, 8.5, 9.2, 10.1, 11.2, 12, 13.
- e) Show that empirical distribution function is an unbiased estimator of distribution function.

Q2) Attempt the following (any three).

[15]

- a) Define the concept of Censoring. Hence, discuss Type I and Type II censoring with an illustration.
- b) Prove that hazard rate is constant if and only if underlying distribution is exponential.
- c) Obtain the hazard rate for the Weibull distribution with scale parameter λ and shape parameter γ . Also, discuss the nature of the hazard rate for different values of the shape parameter γ .
- d) Obtain asymptotic and exact confidence interval for the parameter of exponential distribution in case of complete data.

P.T.O.

Q3) Attempt the following. (any three). [15]

- a) State two definitions of IFR class of life distributions and show that both the definitions are equivalent.
- b) Obtain the confidence band for the survival function.
- c) Prove the following implications.
 - i) $IFR \Rightarrow DMRL$
 - ii) $DMRL \Rightarrow NBUE$
- d) Explain the procedure to obtain the maximum likelihood estimators of parameters of Gamma distribution for type II censored data.

Q4) Attempt the following. (any three). [15]

- a) Prove that if $F \in NBUE$ then $(\psi_F(t)) \geq t, 0 \leq t \leq 1$, where $\psi_F(t)$ is scale total time on test transform.
- b) Show that Deshpande's test statistic lies between 0.5 to 1.
- c) If T is continuous non - negative random variable having distribution function $F(t)$ and cumulative hazard rate $R(t)$ then show that distribution of $R(t)$ is standard exponential.
- d) Obtain expression for actuarial estimator of survival function in case of censored data.

Q5) Attempt the following. (any one). [15]

- a)
 - i) Explain Hollander and Proschan test for testing exponentiality against NBU class of life distribution. [9]
 - ii) State and prove characteristic property of IFRA class of life distribution. [6]
- b)
 - i) The following failure and censor time (in operating hours) were recorded on 12 turbine veins : [6]
142, 149, 320, 345 +, 560, 805, 1130 +, 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.
 - ii) Discuss Mann Whitney U-test for testing whether two samples come from the population having same distribution function. [9]



Total No. of Questions : 5]

SEAT No. :

PD3694

[Total No. of Pages : 3

[6487]-46

S.Y. M.Sc.

STATISTICS

**ST - 43(B) : Categorical Data Analysis
(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) Attempt all questions.

[5×2=10]

- a) Define Mixed Logit Model.
- b) Define the test statistic for testing the significance of individual regression coefficient in Binary Logistic regression.
- c) State the Wald's statistic for testing a specified value of Poisson parameter.
- d) Define relative risk and explain with an example how it is more insightful than difference proportions.
- e) Explain 'Baseline Category Logits' for nominal response variables.

Q2) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Explain the applications of binomial and multinomial distribution in detail with examples.
- b) Give the framework of
 - i) Likelihood Ratio Test
 - ii) Score Test
- c) The following 2×2 contingency table is from a report on Aspirin use and Myocardial Infarction. Determine odds ratio and relative risk. Interpret the results.

Group	Myocardial Infarction	
	Yes	No
Placebo	189	10845
Aspirin	104	10933
Total	293	21778

- d) Discuss goodness of model fit measures with example for the following:
 - i) Pearson's Chi-square
 - ii) Deviance

P.T.O.

Q3) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Define categorical variable. Explain the different types of categorical variable with examples.
- b) Write a short note on Poisson regression model.
- c) Explain log-linear analysis for analyzing dependency in contingency table.
- d) Give the testing procedure for Cochran Mantel Haenszel test.

Q4) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Discuss the usefulness of General Linear Model in the context of categorical data analysis.
- b) For the multinomial distribution $(n, \{\pi_i\})$ with $c > 2$, obtain the confidence limit for π_j .
- c) Estimate the parameters for multiple logistic regression model. Consider the statement, "Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children." For the 1996 General Social Survey, conducted by the National Opinion Research Center, 842 replied "yes" and 982 replied "no." Let π denote the population proportion who would reply "yes." Find the p-value for testing $H_0: \pi = 0.5$ using the score test and construct a 95% confidence interval for π . Interpret the results.
- d) Perform the McNemar test on the following data and interpret the result.

MI Controls	MI Cases	
	Diabetes	No Diabetes
Diabetes	9	16
No Diabetes	37	82

Q5) Attempt any 1 question out of 2 questions.

[1×15=15]

- a) Explain the uses of the following:
 - i) Adjacent category model
 - ii) Conditional Logit Model
 - iii) Mixed logit model
 - iv) Negative Binomial regression
 - v) Poisson regression

- b) 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 that are “labeled.” Software reports following Table for a logistic regression model using LI to predict $\pi = P(Y = 1)$.

Computer Output for Problem						
Parameter	Estimate	Standard Error	Likelihood Ratio 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	
	Source	DF	LR Statistic 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	
....						

- State the prediction equation and interpret it. [3]
- Conduct a Wald test for the LI effect. Interpret. [3]
- Construct a Wald confidence interval for the odds ratio corresponding to a 1-unit increase in LI . Interpret. [3]
- Conduct a likelihood-ratio test for the LI effect. Interpret. [3]
- Construct the likelihood-ratio confidence interval for the odds ratio. Interpret. [3]



Total No. of Questions : 5]

SEAT No. :

PD3695

[Total No. of Pages : 3

[6487]-47

S.Y.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) Attempt all questions.

[5×2=10]

- a) What is the difference between stochastic EM and usual EM?
- b) State the uses of importance sampling method.
- c) A random sample from a probability density function, $f(x)$ is to be used to compute the integral, $\int_0^1 x^3 dx$. State the most appropriate choice of distribution for $f(x)$?
- d) LOOCV stands for what?
- e) For the importance sampling, what can be said about the tails of the proposal and target distribution?

Q2) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) The following table gives a small dataset.

Sr. No.	X	Y
1	5	y_1
2	10	35
3	15	55
4	20	y_2
5	25	y_3

If we use regression imputation, what will be the values of y_1, y_2, y_3 ?
Provide all the calculations.

P.T.O.

- b) Explain the procedure of boosting for regression trees.
- c) What will be the quantile function corresponding to the following probability density function, $f(x)$?

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1/2 & \text{if } 0 < x \leq 1 \\ \frac{1}{2x^2} & \text{if } x > 1 \end{cases}$$

- d) When does one recommend acceptance-rejection sampling? What are the limitations of it.

Q3) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Suppose that we want to estimate the probability of $P[Z > 3]$ for a standard normal random variable Z . Roughly how many observations would be required if we try to get a reasonable estimate of this via Monte Carlo sampling? What would be the steps if we decide to use importance with $N(4,1)$ as the proposal density?
- b) Describe the connection between kernel density estimation and kernel regression estimation.
- c) Identify the target distribution, the proposal distribution and the algorithm being used in the following R code.

```
n = 500
x = c()
y = c()
a = 4
b = 6
x[1] = 0.5
for(i in 1:(n-1))
{
  y[i] = rbeta(1,a,b)
  r =
  min([(dbeta(y[i],3.5,7.5)*dbeta(x[i],a,b))]/[(dbeta(x[i],3.5,7.5)
  *dbeta(y[i],a,b))],1)
  u = runif(1,0,1)
  x[i+1] = ifelse(u≤r,y[i],x[i])
}
```

- d) What are the two key steps in EM algorithm? Explain in detail.

Q4) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) In a binary classification problem, after fitting an initial model with equal weights to all the observations, it is found that out of three observations, one observation is misclassified. What is the revised weight for the misclassified observation as per the Adaboost.M1 algorithm? Provide all the intermediate steps for calculations.
- b) What is single imputation and multiple imputation? What are the advantages of multiple imputation?
- c) What is bias-variance tradeoff? Explain its role in non-parametric density estimation and also in k-fold cross-validation.
- d) Describe how bootstrap samples can be used to construct t intervals? What will be the standard error?

Q5) Attempt any 1 question out of 2 questions.

[1×15=15]

- a) Consider a mixture exponential density as

$$f(x) = pf_1(x) + (1-p)f_2(x), x > 0$$

where $f_1(x)$ and $f_2(x)$ are exponential density functions with rates λ and μ respectively. Assume that a random sample of size n , $X_1, X_2, \dots, X_n$ is available along with the vector of allocations of the observations x_i to the first and second components of the mixture, $z_1, z_2, \dots, z_n$.

- i) Compute the pmf of Z .
- ii) Construct all the steps needed to implement EM algorithm for above model.

[5+10]

- b) Suppose $Y = (y_1, y_2, y_3, y_4)$ has a multinomial distribution with cell probabilities $\left(\frac{1}{2} + \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}, \frac{\theta}{4}\right)$. Define the complete data as $X = (x_0, x_1, y_2, y_3, y_4)$ to have multinomial distribution with probabilities $\left(\frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}\right)$ and to satisfy $x_0 + x_1 = y_1$. If $Y = (125, 18, 20, 34)$, obtain the MLE of θ using EM algorithm.

