Total No. of Questions : 5]

SEAT No. :

PC-3982

[Total No. of Pages : 2

[6347]-1

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 compulsoory.***
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 :** **[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) Define Cauchy sequence with an illustration.

d) Show that if $A \subset B$ and $B \subset C$ then $(A \cup B) \subset C$

e) Define Monotone sequence with an illustration.

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

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

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

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

c) State and prove Heine - Borel theorem.

d) State and prove the Archimedean principle of real numbers.

P.T.O.

Q3) Answer the following questions (Any 3) : **[3 × 5 = 15]**

a) Examine whether the folowing series $\Sigma\, a_n$ converges or diverges.

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

ii) $\sum_{n=0}^{\infty} \frac{1}{n!}$

b) State and prove Cauchy criteria for the convergence of series of real numbers.

c) Prove or disprove the set of rational numbers is dense in R.

d) Prove that if p > 0 then $\sqrt[n]{p} = 1$

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

a) Define : An open set, closed set in metric space. Show that finite union of closed set is closed.

b) Show that a convergent sequence in a metric space is bounded. Is the converse is true? Justify your answer.

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

d) State and prove finite intersection property of compact set.

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

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

i) Common refinement of two partitions.

ii) Reimann sum associated with a partition.

iii) Improper integral of first kind.

iv) Local maximum of a function.

B) State and prove Bolzano - Weierstrass theorem in R. **[7]**

b) A) State and prove Mean Value Theorem. **[9]**

B) Show that $f \in RS(\alpha)$ 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. :

PC3983

[Total No. of Pages : 3

[6347]-2
First Year M.Sc.
STATISTICS
ST-12 : Linear Algebra and Numerical Methods
(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 Eigen vector with Example.

b) Explain Rank One Matrices with an illustration.

c) Define Outer product of two vectors.

d) Define Singular value decomposition.

e) Define Eigen value with example.

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

a) Verify the Cayley-Hamilton Theorem for the given matrix.

$$A = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix}$$

b) Approximate the area beneath $y = \sin x$ on the interval $[0, \pi]$ using the Trapezoidal Rule with $n = 6$ subintervals.

c) Find the Jordan form for the matrix

$$A = \begin{bmatrix} -7 & 8 & 2 \\ -4 & 5 & 1 \\ -23 & 21 & 7 \end{bmatrix}$$

d) If A is any square matrix then prove that $|A^k| = |A|^k$

P.T.O.

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

a) Solve the following system of linear equation by using Gauss Jacobin method.

$8X_1 - 3X_2 + 2X_3 = 20$

$4X_1 + 11X_2 - X_3 = 33$

$6X_1 + 3X_2 + 12X_3 = 35$

b) Examine the nature of Quadratic form.

$5X^2 - Y^2 + 7Z^2 + 5XY - 3YZ$

Obtain non-singular transformation which reduces it to canonical form.

c) Find cofactor of given matrix A and check whether it satisfied properties of co-factor.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

d) Find inverse of square matrix (by Adjoin method) following matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$$

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

a) Solve the following.

i) If A and B are anti-symmetric matrices the prove that $(A + B)^2 = A^2 + B^2$

ii) If A and B are Idempotent matrices, then check whether AB is idempotent matrix.

b) Prove that if two rows or two columns of a matrix A is same then det (A) = 0.

c) Reduce the given matrix to upper triangular form and find its determinant.

$$P = \begin{bmatrix} 3 & 8 & 7 \\ 1 & 2 & 4 \\ -1 & 3 & 2 \end{bmatrix}$$

d) Prove that, if one row or column of determinant any multiple of any other column or Row does not affect the value of determinant.

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

a) i) Using following data find the Newton's interpolating polynomial and also find the value of y at x = 5. **[8]**

X	0	10	20	30	40
Y	7	18	32	48	85

ii) Apply Gram-Schmidt Orthogonalization to the following sequence of vector. **[7]**

$\{(1, -1,1), (1, 0, 1), (1, 1,2)\}$

b) i) Find the Singular Value Decomposition of the matrix A, where

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$ **[7]**

ii) Define idempotent matrix. Show that for an idempotent matrix A, rank (A) = trace (A). **[8]**

❋ ❋ ❋

Total No. of Questions : 5]

SEAT No. :

PC3984

[Total No. of Pages : 4

[6347]-3

First Year 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 scientific calculator and statistical tables is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

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

a) Define the following terms: **[2 marks each]**

i) Decomposition of P.d.f.

ii) Marshall-olkin model of bivariate exponential distribution.

iii) Mathematical expectation of random variable.

b) Attempt the following: **[2 marks each]**

i) A box contains tickets numbered 1 to N. Let X be the largest number drawn in n random drawings with replacement. Then show that,

$$P[X \le k] = \left(\frac{k}{N}\right)^n$$

ii) If the distribution of random variable X is symmetric about zero then prove that, $\phi_X(t)$ is even function of t.

($\phi_X(t)$ is characteristic function of t)

P.T.O.

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

a) If $X \rightarrow B(3, 1/2)$ and $Y \rightarrow B(2, 1/3)$. X and Y are independent then find the probability distribution of $U = X + Y$.

b) Define cumulative distribution function $F(x)$ of random variable X based on probability space $(\Omega, \mathbb{F}, \mathbb{P})$. Also show that $(F(x))^2$ is also a c.d.f.

c) Prove that for any two random variables X and Y .

$$\operatorname{var}(X) = E_Y\left[\operatorname{var}(X|Y)\right] + \operatorname{Var}_Y\left[E(X|Y)\right]$$

d) If x and y are two continuous i.i.d. random variables such that X and Y are symmetric around zero then show that U=X+Y is symmetric around zero.

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

a) If X is a continuous non-negative random variable with c.d.f. $F(x)$ then show that

$$E(X) = \int_0^\infty \left[1 - F(x)\right] dx$$

b) Find the moment generating function of random variable X having p.d.f.

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \qquad ; x \in \mathbb{R}$$

c) Examine whether the following p.d.f. is symmetric

$$f(x) = \begin{cases} 1+x & ; \text{ if } -1 \le x \le 0 \\ 1-x & ; \text{ if } \ 0 < x \le 1 \end{cases}$$

If it is symmetric then state the point around which it is symmetric.

d) Let X_1, X_2, X_3 be a random sample from the distribution with p.d.f:

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

Compute the probability that smallest of X_1, X_2, X_3 exceeds the median of the distribution.

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

a) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from density $f(x)$ and distribution function $F(x)$. Let $X_{(1)}, X_{(2)}, \ldots X_{(n)}$ be the corresponding order statistics. Find the distribution of sample range.

b) Let $X_1, X_2, \ldots, X_N$ are independent and identically distributed Bernoulli (p) random variables and N is $B(m,p_1)$ then find the probability distribution of $S_N = X_1 + X_2 + \ldots + X_N$. State $E(S_N)$ and $var(S_N)$.

c) Let Ω be the sample space given by $\Omega = \{a,b,c\}$ and $\mathbb{F}$ be the field defined on Ω as,

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

and $X(a) = 2$, $X(b) = 3$, $X(c) = 5$. Then check whether $X(\bullet)$ is a random variable.

d) For the p.d.f., $f(x) = \frac{1}{2\lambda} \cdot e^{-|x-\mu|/\lambda}$; $-\infty < x < \infty$

$\lambda > 0$

$-\infty < \mu < \infty$

Show that the moment generating function exists and equals.

$$M_X(t) = \frac{e^{\mu t}}{(1-\lambda^2 t^2)} \quad ; |t| < \frac{1}{\lambda}$$

Q5) Attempt any one of the following:

a) i) Two random variables X and Y have the following joint p.d.f. **[10]**

$$f(x,y) = \begin{cases} 2-x-y & ; 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & ; \text{otherwise} \end{cases}$$

Find:

1) Marginal densities of x and y
2) E(X), E(Y), var(X) and var(Y)
3) Covariance between X and Y

ii) If 't' is any positive real number then show that the function given below is a p.m.f.: **[5]**

$$P(X = x) = e^{-t}(1-e^{-t})^{x-1} \quad ; x = 1, 2, 3....$$

Also compute E(X)

b) i) For the c.d.f. given below: **[10]**

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \dfrac{x+1}{2} & \text{if} \quad 0 \le x < 1 \\ 1 & \text{if} \quad x \ge 1 \end{cases}$$

What kind of the random variable X is? Is it continuous or discrete or mixed r.v.? Justify.

ii) If $E(X^r)$ exists then $E(X^s)$ also exists for all $1 \le s \le r$ .i.e. prove that if the moments of a specified order exist then all the lower order moments also exist. **[5]**

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PC3985

[Total No. of Pages : 3

[6347]-4

First Year M.Sc.

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) ***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 : **[2 each]**

a) Define the following terms with illustration
   i) Sampling frame
   ii) Sampling unit

b) Explain the method of circular systematic sampling with illustration.

c) State unbiased and biased estimator of population mean in case of stratified random sampling.

d) Discuss the multistate sampling and also gives on real life example where this type of sampling is used.

e) State one advantage and one disadvantage of the Lahiri Method.

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

a) Explain in brief cumulative total method of selecting a sample of size n by SRSWR with Probability Proportional to Size (PPS) and state its limitations

b) Show that for the simple random sampling with replacement sample mean sum of square is an unbiased estimator of the population variance.

c) Determine the size of sample to be drawn from the population using SRRSWOR method for prespecified width of the confidence interval for the population mean with the confidence coefficient $(1-\alpha)$.

d) Prove that for any sampling design P(.), $E_p[I_i(s)] = \pi_i$ where $I_i(s)$ and π_i are inclusion indicator and first order inclusion probability.

P.T.O.

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

a) For the simple random sampling without replacement show that

$$\sum_{t-1}^{\binom{N}{n}}\left(\sum_{i=n}^{n} Yi\right)_t = \sum_{i-1}^{N}\binom{N-1}{n-1} Yi$$

b) Show that in case of varying probability scheme with replacement $\overline{z} = \frac{1}{n}\sum_{i=1}^{n} zi$, where $zi = \frac{yi}{NPi}$ i = 1, 2,,,,,, N is an unbiased estimator of population mean. Also find its variance.

c) Prove that in case of inverse sampling estimator $\frac{(m-1)}{n-1}$ is an unbiased estimator for the population proportion P, where m is predetermined number of units possessing given attribute in the sample of size n.

d) State an unbiased estimator of population means in case of post stratification. Hence find variance of the estimator.

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

a) Explain Modified systematic sampling with illustration.

b) Define regression estimator of the population mean. Obtain its expectation and mean square error.

c) Prove that sample mean of cluster means selected in sample is an unbiased estimator of population mean.

d) Find variance of the estimator of the population mean in case of two stage sampling having equal number of first stage units.

Q5) Attempt any one of the following :

a) i) Find the optimum value of sample size to be used in the first and second stage of the two-stage sampling. **[7]**

ii) Explain in brief 'Warner's Randomized Response Technique' (RRT) **[8]**

b) i) Derive the Yates corrected estimator of the population mean. **[8]**

ii) With two strata, sampler would like to have $n_1 = n_2$ for the administrative convenience, instead of using the values given by Neyman allocation. If V and V_{opt} denote the variances given by $n_1 = n_2$ and Neyman allocations, respectively, show that the $\frac{V - V_{opt}}{V_{opt}} = \left(\frac{r-1}{r+1}\right)^2$ where $r = \frac{n_1}{n_2}$ given by Neyman alloction. Assume N_1 and N_2 large. **[7]**

× × ×

Total No. of Questions : 5]

SEAT No. :

PC3986

[Total No. of Pages :3

[6347]-11

First Year M.Sc.

STATISTICS

ST-21 : Probability Theory

(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:

i) Convergence in probability

ii) Convergence in r^{th} mean

b) Show that $X_n \xrightarrow{P} 0$ if $E|X_n|^r \to 0, (r > 0)$.

c) Write the economical definition of random variable.

d) State the correspondence theorem.

e) Define positive and negative part of the random variable.

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

a) State and prove Holder's inequality.

b) Let F(x) be the distribution function (DF) of RV X, where

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{1}{4} & \text{if } 0 \le x < 1 \\ \dfrac{1}{2} + \dfrac{1 - e^{-(x-1)}}{2} & \text{if } x \ge 1 \end{cases}$$

Show that F is linear combination of distribution functions of discrete and continuous RV..

P.T.O.

c) Let $\mathbb{C}_1 = \{(a,b]; a,b \in \mathbb{R}\}$ and $\mathbb{C}_2 = \{[a,b]; a,b \in \mathbb{R}\}$. Then show that $\sigma(\mathbb{C}_1) = \sigma(\mathbb{C}_2)$.

d) If X and Y are random variables then show that Z = *max*(X, Y) and W = *min*(X, Y) are also a random variables.

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

a) Define probability measure. Prove that

i) Probability measure is subtractive.

ii) Probability measure is finite sub-additive.

b) Show that, $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$.

c) Let $\{A_n; n \geq 1\}$ and $\{B_n; n \geq 1\}$ be two sequence of sets then prove that:

i) $\overline{\lim}(A_n \cup B_n) = \overline{\lim}(A_n) \cup \overline{\lim}(B_n)$

ii) $\underline{\lim}(A_n \cap B_n) = \underline{\lim}(A_n) \cap \underline{\lim}(B_n)$

d) Show that:

i) If $X_n \xrightarrow{a.s.} X$ and $X_n \xrightarrow{a.s.} Y$ then $X = Y$ *a.s.*

ii) If $X_n \xrightarrow{a.s.} X$ and $X_n \xrightarrow{a.s.} Y$ then $X_n + Y_n \xrightarrow{a.s.} X + Y$.

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

a) Show that $X_n \xrightarrow{P} X$ iff $E\left\{\frac{|X_n - X|}{1+|X_n - X|}\right\} \to 0$ as $n \to \infty$.

b) Consider a function X defined as follows:

$$X(w) = \begin{cases} C_0 & \text{if } \omega \in A_0 \\ C_1 & \text{if } \omega \in A_1 \\ C_2 & \text{if } \omega \in A_2 \end{cases}$$

Where $\{A_0, A_1, A_2\}$ forms partition of Ω. Find the σ - field induced by function X.

c) Define simple random variable. If X_1 and X_2 are simple random variables then prove that, $X_1 + X_2$, $X_1 - X_2$, $X_1 * X_2$ and $\frac{X_1}{X_2}$ (provided it is defined) are also simple random variables.

d) If $f(X)$ is continuous real valued function and $X_n \xrightarrow{P} X$ then $f(X_n) \xrightarrow{P} f(X)$.

Q5) Attempt any one of the following:

a) i) If X and Y are independent random variables then show that:

$E(XY) = E(X)E(Y)$. **[8]**

ii) State and prove Kolmogorov 0 - 1 law. **[7]**

b) i) Suppose X and Y are simple random variables defined on a probability space $(\Omega, \mathbb{A}, P)$. Then show that **[8]**

1) $E(X \pm Y) = E(X) \pm E(Y)$

2) $E(cX) = cE(X)$

3) If $X \geq 0$ on Ω then $E(X) \geq 0$

4) If $X \geq 0$ a.s. then $E(X) \geq 0$

ii) State and prove continuity property of probability measure. **[7]**

Total No. of Questions : 5]

SEAT No. :

PC3987

[Total No. of Pages : 2

[6347]-12

First Year M.Sc.

STATISTICS

ST-22: Regression 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 all questions. **[5×2=10]**

a) For simple linear regression model, show that, $SS_{\text{reg}} = \hat{\beta}_1 s_{xy}$.

b) Obtain the relationship of analysis of variance test and coefficient of determination in case of multiple linear regression model.

c) 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?

d) Write the given model in general linear model set up.

$E(y_1) = \theta_1 + \theta_2$, $E(y_2) = \theta_1 - \theta_3$ and $E(y_3) = \theta_1 + \theta_2$.

e) Define link function for the generalized linear model.

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

a) Obtain $100(1-\alpha)\%$ confidence interval on regressor parameters β_0 and β_1 of the simple linear regression model for the known and unknown σ^2.

b) Explain the reverse regression method with a suitable example. Also, obtain the least square estimators of the regression coefficients for it.

c) Define Polynomial regression model. Estimate the parameters of orthogonal polynomial regression model and give its variance.

d) State and prove Gauss-Markoff theorem.

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

a) In case of near or high multicollinearity, explain the all possible consequences which can be encountered.

b) Discuss the method of principle component regression for dealing with the problem of multicollinearity.

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

d) Obtain the bias, covariance and mean squared error of ridge regression estimator.

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

a) For nonlinear regression model $y = \theta_1 e_2 x + \varepsilon$ obtain least square normal equation. Now, linearize the model and obtain least square estimates of parameters and hence compare the two methods.

b) Explain the Gauss-Newton iteration method of parameter estimation in nonlinear regression models.

c) Discuss Wald test in logistic regression model.

d) What is pure error? When one can have an estimate of it?

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) i) Explain the logistic regression model with single explanatory variable. Also, obtain the maximum likelihood equation for it. **[8]**

ii) Let E $(Y_1) = 2\mu$, E $(Y_2) = \mu$ and $Cov\ (\underline{y}) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \sigma^2$. Obtain an unbiased estimator of σ^2. **[7]**

ⓘⓘⓘⓘ

Total No. of Questions : 5]

SEAT No. :

PC3988

[Total No. of Pages : 3

[6347]-13
M.Sc. - I
STATISTICS
ST - 23 : Statistical Inference -I
(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 term simple and composite hypothesis with an illustration.

b) Define shortest expected length.

c) Explain uniformly most powerful unbiased test.

d) State Bhattacharya bounds.

e) Define prior and posterior distribution.

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

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

b) Define multiparameter exponential family. If $X \sim G(\alpha,\lambda)$ then show that the distribution of X belonging to multiparameter exponential family.

c) Define Joint sufficient statistic. Let $X_1, X_2, ... X_n$ be a random sample from $N(\mu, \sigma^2)$ both are unknown find sufficient statistic for (μ,σ^2).

d) Define complete sufficient statistic. Let $X_1, X_2, ... X_n$ be a random sample from Bernoulli (θ)obtain complete sufficient statistic for θ.

P.T.O.

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

a) Define Fisher information matrix and find information matrix I (μ,σ^2) if $X \sim N(\mu,\sigma^2)$.

b) Show that, in the regular case of point estimation, that $E\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)^2 = -E\left(\frac{\partial^2 \log f(x,\theta)}{\partial\theta^2}\right)$, stating all the conditions you need.

c) Define pitman family. Let $X_1, X_2, \ldots\ldots X_n$ be i.i.d. with pdf $f(x,\theta)$ which is a member of pitman family. Let $S_\theta = (a(\theta), b(\theta))$ be support of $f(x,\theta)$. If $b(\theta) = b$ constant then Obtain minimal sufficient statistic for θ.

d) Let $X_1, X_2, \ldots\ldots X_n$ be a random sample from exponential distribution with mean $= \theta$. Obtain 100 $(1-\alpha)$% shortest expected length confidence interval for θ.

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

a) Let the distribution of a random variable X under $H_0 : X \sim P_1(x)$ and $H_1 : X \sim P_2(x)$. be given as possible distributions:

X	1	2	3	4	5	6	7
$P_1(x)$	0.01	0.02	0.03	0.05	0.07	0.05	0.77
$P_2(x)$	0.03	0.09	0.1	0.1	0.2	0.18	0.3

Find the MP test of size 0.1.

b) Define biased and unbiased estimator of parameter. Let pdf of X is $f(x,\theta) = 2\theta x e^{-\theta x^2}, x > 0, \theta > 0$, obtain unbiased estimator of $\frac{1}{\sqrt{\theta}}$.

c) Suppose $X_1, X_2, \ldots\ldots X_n$ are iid Exp (λ) random variables, where λ is the rate parameter. UMP test of $H_0 : \lambda \le \lambda_0$ against $H_1 : \lambda > \lambda_0$ of size α.

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

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

a) i) Suppose the conditional pdf of a random variable X given θ is $f(x,\theta)=\frac{2x}{\theta^2};0<x<\theta$ where the prior disrtribution of θ is unifrom (0,1). Based on single observation *x* form *X* find Bayes estimator for θ under squared error loss function and absolute error loss function. **[8]**

ii) State and prove Basu's theorem. Let X_1 and X_2 be random sample from (μ,σ^2), σ^2 is known then show that $M = X_1 + X_2$ and $T = X_1 - X_2$ are independent. **[7]**

b) i) State and prove Neyman-Pearson Lemma. A sample of size 1 is taken from pdf $f(x)=\frac{2}{\theta^2}(\theta-x)$, where $0\le x<\theta$. Find an MP test of its size for testing $H_0:\theta = \theta_0$ Vs $H_1: \theta = \theta_1$ $(\theta_1 < \theta_0)$. **[10]**

ii) Consider hypotheses $H_0:\theta\in\Theta_0\ Vs H_0:\theta\in\Theta_1$. Suppose that, for every *T*, the power function is continuous in θ. If *T** is uniformly most powerful among all similar tests and has size α, then show that T* is a UMPU test. **[5]**

❋ ❋ ❋

Total No. of Questions : 5]

SEAT No. :

PC3989

[Total No. of Pages : 4

[6347]-14
First Year M.Sc.
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) If $\underline{X} \sim N_2\left(\underline{0}, \Sigma\right)$, where $\Sigma = \begin{bmatrix} 1 & 0.7 \\ & 1 \end{bmatrix}$ then find $P(X_1 < X_2)$.

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

c) Define the following:

i) Factor loadings

ii) Factor scores

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

e) Suppose $\underline{X} \sim N_2\left(\underline{\mu}, \Sigma\right)$ with $\underline{\mu} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\Sigma = I_2$. Consider $A=[1 \;\; 1]$ and $B = [1 \; -1]$. Verify whether $A\underline{X}$ and $B\underline{X}$ are independent.

P.T.O.

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

a) Let $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ be a random sample of size n from p variate normal distribution with mean vector $\underline{\mu}$ and variance covariance matrix Σ. Derive the maximum likelihood estimators of the parameters $\underline{\mu}$ and Σ.

b) 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.

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 $\underline{X}^{(2)}$ given $\underline{X}^{(1)} = \underline{x}^{(1)}$.

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

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

a) Write a note on One Way MANOVA.

b) If $D \sim W_p(n, \Sigma)$, then show that

i) $\dfrac{|D|}{|\Sigma|}$ is distributed as the product of p independent χ^2 variates with n(n–1).....n –(p –1) d.f. respectively.

ii) $\dfrac{\sigma^{pp}}{d^{pp}} \sim \chi^2_{n-(p-1)}$, where d^{pp} and σ^{pp} are the last elements of D^{-1} and Σ^{-1} respectively.

c) If $\underline{X}$ is a vector of p components $X_1, X_2, \ldots, X_p$ having variance-covariance matrix Σ with eigen values $\lambda_1, \lambda_2, \ldots, \lambda_p$ and corresponding eigen vectors $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_p$. 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) The variance-covariance matrix is $\Sigma = Cov \begin{bmatrix} X_1 \\ X_2 \\ Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 1 & 0.95 & 0 \\ 0 & 0.95 & 1 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$.

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.

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) Let the random vector $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \sim N_4(\underline{\mu}, \Sigma)$ with $\underline{\mu} = \begin{bmatrix} 3 \\ 2 \\ -2 \\ 0 \end{bmatrix}$ and

$\Sigma = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$. Let $A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix}$, find

i) $E(A\underline{X})$

ii) $V(A\underline{X})$

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

$D_{11} \sim W_p(n,\Sigma_{11})$.

d) Let $\underline{X} \sim N_p(\underline{\mu},\Sigma)$

i) Obtain the characteristic function of $\underline{X}$

ii) Find the distribution of $(\underline{a}'\underline{X}+b)$, where b is known constant and $\underline{a}$ is PX_1 vector of constants.

Q5) Attempt any one of the following:

a) i) Identify the sampling distribution of maximum likelihood estimators of multivariate normal distribution. Show that these maximum likelihood estimators are independently distributed. **[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) Evaluate Hotelling's T^2 for testing $H_0 : \underline{\mu} = \begin{bmatrix} 7 \\ 11 \end{bmatrix}$ using the data

$\begin{bmatrix} 2 & 8 & 6 & 8 \\ 12 & 9 & 9 & 10 \end{bmatrix}$ Specify the distribution of T^2 for this situation.

Test H_0 at $\alpha = 0.05$ level. What conclusion do you reach? **[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. **[7]**

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PC-3990

[Total No. of Pages : 4

[6347]-21
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 the questions :** **[5 × 2 = 10]**

a) State postulates of Poisson process.

b) Give the Transition Probability Matrix (TPM) for random walk with absorbing barriers and birth-death chain.

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

d) Define stochastic processes with their classification.

e) State any two properties of Brownian motion.

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

a) Consider a finite Markov Chain $\{X_n, n > 0\}$ with state space $s = \{0, 1, 2\}$

and TPM is $P = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.4 & 0.6 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix}$. Calculate:

i) $P[X_2 = 1, X_0 = 2]$

ii) $P[X_2 = 1]$

iii) $P[X_2 = 1, X_1 = 1, X_0 = 2]$ where initial probability distribution is $\alpha_0 = 0.3$, $\alpha_1 = 0.3$ and $\alpha_2 = 0.4$.

P.T.O.

b) Suppose that every man in a certain society has exactly 2 children, which independently have probability 1/2 of being a boy and 1/2 of being girl. Suppose also that the number of males in the n^{th} generation $\{X_n, n > 0\}$ forms 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 one boy having probability 2/3 and one girl having probability 1/3, what is probability that his male line will continue forever?

c) Consider a finite Markov Chain $\{X_n, n > 0\}$ with state spaces $s = \{0, 1, 2\}$

and TPM is $P = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.4 & 0.6 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix}$. Calculate :

i) $P[X_2 = 1, X_0 = 2]$

ii) $P[X_2 = 1]$

iii) $P[X_2 = 1, X_1 = 1\ X_0 = 2]$ where initial probability distribution is $\alpha_0 = 0.3$, $\alpha_1 = 0.3$ and $\alpha_2 = 0.4$.

d) Define persistent and transient states of a Markov chain. Show that being transient is a class property.

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

a) Show that a state k is recurrent iff $\sum_{n=1}^{\infty} P_{kk}^{(n)} = \infty$.

b) Define Birth and death process and obtain its stationary distribution.

c) Define a non-null persistent state. Show that in a finite Markov chain there exists a non-null persistent state.

d) 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 ergodic and non-ergodic states.

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

a) Define $P_{ij}^{(n)}$ and n-step transition probability for a Markov chain with state space S. Show that $P_{ij}^{(n)} = P_{ik} P_{kj}^{(n-1)}, \forall n \geq 1$.

b) Define a weakly stationary and strictly stationary stochastic process. Show that if the process is strictly stationary, then it is also weakly stationary.

c) Let $\{X_n, n > 0)$ be a two state Markov chain with $P_{00} = \frac{1}{3}, P_{01} = \frac{2}{3}, P_{10} = \frac{3}{4}$ and $P_{11} = 1/4$. Show that the vector valued process $\{Y_n, n > 0\}$, $Y_n = \{X_{n-1}, X_n\}$ is also a Markov chain. Find TPM of the process.

d) Show that the ultimate extinction probability of a Galton-Watson branching process is the smallest non-negative root of the equation $s = f(s)$, where f is the probability generating function of the offspring distribution. Hence obtain the extinction probability of a Branching process whose off-spring probabilities are $P_0 = q$, $P_1 = p$, $p + q = 1$, $0 < p < 1$.

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) Suppose that the number of accidents per week at industry is Poisson process with rate 4. Also, the numbers of workers injured in each accident are independent random variables with a Poisson mean of 2. Assume that the number of workers injured in each accident is independent of number of accidents that occur. What are the expected number of injuries and variance of number of injuries of during a month of February? **[4]**

iii) Obtain the mean of the offspring distribution $E(X_n)$ and the probability of ultimate estimation of Goltan-watson process with offspring pgf $f(s) = 0.2 + 0.2s + 0.3s^2 + 0.3s^3$. **[4]**

b) i) Define stationary distribution of Markov chain. Show that if π_1 and π_2 are two distinct stationary distributions then $\lambda\pi_1 + (1 - \lambda)\pi_2$ $0 \leq \lambda \leq 1$ is also stationary distribution. State a condition under which stationary distribution is unique. **[7]**

ii) If $\{N_1(t), t > 0\}$ and $\{N_2(t), t > 0\}$ are two independent Poisson processes with parameters λ_1 and λ_2 respectively. Then, show that:

$$P[N_1(t)=k]N_1(t)+N_2(t)=n]=\binom{n}{k}p^k\ q^{n-k}\text{, where } p=\frac{\lambda_1}{\lambda_1+\lambda_2}.$$

[8]

⌘⌘⌘

Total No. of Questions : 5]

SEAT No. :

PC3991

[Total No. of Pages : 3

[6347]-22
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 the term Prior distribution and Posterior distribution.

b) Represent the squared error loss function, absolute error loss function and zero-one (all-or-nothing) loss function graphically.

c) Define non-informative prior with an illustration.

d) Define loss function with illustration.

e) Let $X \sim N(\theta,4)$ and assume that we are using squared error loss. Let $\hat{\theta} = X$. Find the expected loss (risk) for this estimator.

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

a) Suppose $X_1, X_2,......, X_n$ are i.i.d. from a Bernoulli distribution with unknown parameter p. By using conjugate prior show that the posterior distribution is a Beta $(k + \alpha, n - k + \beta)$ distribution, and find the MAP estimator.

b) Define absolute error loss function and obtain the bayes estimator under ab solute error coss function.

c) Let $X_1, X_2,......, X_n$ 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.

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.

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

a) What is Gibbs Sampling technique? Describe briefly.

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

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) Explain Bayes rule. A man is known to speak the truth two out of three times. He throws a die and reports that the number obtained is four. Find the probability that the number obtained is actually a four.

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

a) 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.

b) If X follows poisson distribution with parameter λ and the prior distribution of λ is Gamma with parameters α and β. Find posterior distribution of λ based on $Y = \sum_{i=1}^{n} X_i$ which is sufficient statistics for λ where $X_1, X_2, \ldots, X_n$ is random sample of size n from *Poisson* distribution.

c) What are the two key steps in EM algorithm?

d) i) Let $X \sim N(\theta, 4)$ and assume that we are using squared error loss. Let $\hat{\theta} = x$. Find the expected loss (risk) for this estimator.

ii) Let $X \sim B(n, \theta)$ and prior distribution of θ is $\beta_1(a, b)$. Find the posterior distribution of θ. Is given prior conjugate or not?

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

a) i) Suppose that we know $Y|X = x \sim N(x,1)$. Show that the posterior density of X given $Y = y$, $f_{x|y}(x|y)$ is given by $X|Y = y \sim N\left(\frac{y}{2},\frac{1}{2}\right)$. Also find bayes estimator under squared loss. **[9]**

ii) Explain Robust Bayesian analysis in briefly. **[6]**

b) i) Let X ~ N (0,1). Suppose that we know $Y|X = x \sim N(x,1)$. Show that the posterior density of X given $Y = y$, $f_{X|Y}(x|y)$ is given by $X|Y = y \sim N\left(\frac{y}{2},\frac{1}{2}\right)$. Also find bayes estimator under squared loss. **[7]**

ii) Explain the terms with illustration: Subjective priors and probability matching prior. **[8]**

❋ ❋ ❋

Total No. of Questions : 5]

SEAT No. :

PC3992

[Total No. of Pages : 3

[6347]-23
S.Y.M.Sc.
STATISTICS
ST-33 : Design and Analysis of Experiments
(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 each of the following. **[5×2=10]**

a) What are the assumption of ANOVA.

b) Check whether following block design is connected:

Block - I	Block - II	Block - III	Block - IV
1	1	1	1
2	3	2	2
		4	3
			4

c) Define BIBD, with usual notations prove that: $vr = bk$

d) Explain the linear and quadratic effect of factor.

e) Define the following:

i) Resolvable BIBD

ii) Affine Resolvable BIBD

P.T.O.

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

a) Give the statistical analysis of PBIBD(2).

b) What is factorial experiment? Give an illustration. Explain advantages and disadvantages of factorial experiment.

c) If [(1), ae, abc, bce, acd, cde, bd, abde] is the key block of a 2^5 factorial experiment, then obtain the confounded interactions.

d) What is Simplex Lattice Design. Explain a [3, 2] Simplex Lattice Design.

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

a) Explain the procedure of Kruskal-Wallis Test.

b) Explain Yates method to compute factorial effect totals in a 3^2 factorial experiment.

c) Confound I(AB) = AB^2 and J(AB) = AB in 3^2 factorial experiment in 3 blocks.

d) Write a note on Box-Behnken Design (BBD).

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

a) A test is given to students taken at random from the fifth class of 3 schools of the town. The individual scores are:

School I:	9	7	6	5	8
School II:	7	4	5	4	8
School III:	6	5	6	7	6

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

b) Explain Friedman Test.

c) Construct one-half fraction of a 2^5 design with highest possible resolution. Write down its alias structure.

d) Write a note on Taguchi design.

Q5) Attempt any one of the following:

a) i) Check whether following design is connected and/or orthogonal and/or variance balanced. **[8]**

	Blocks			
Treatments	A	B	C	D
	C	D	A	B
	A	C	C	A
	B	B	D	D

ii) In BIBD, prove that **[7]**

1) $\lambda (v-1) = r (k-1)$

2) $b \geq v$.

b) Analyze the following data: **[15]**

Replicate - I		Replicate - II		Replicate - III		Replicate - IV	
ab=3	b=1	'1'=4	a=3	'1'=6	b=3	b=3	a=4
'1'=2	a=4	c=1	ac=0	a=3	c=4	'1'=4	c=3
ac=0	abc=0	ab=3	b=0	bc=2	ab=5	ac=3	ab=5
bc=4	c=5	abc=2	bc=2	abc=1	ac=2	abc=4	bc=0

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PC3993

[Total No. of Pages : 3

[6347]-24

S.Y. M.Sc.

STATISTICS

ST-34 : Machine Learning

(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) Which are the main layers in artificial neural network?

b) What is the Fl-score, and why is it important?

c) What is k-fold cross-validation?

d) What is the difference between a classification and a regression Decision Tree?

e) Define Single Linkage in Agglomerative Hierarchical Clustering.

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

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

b) What are the applications of Machine Learning in various fields?

c) What is tree pruning in Decision Tree learning, and why is it important? Explain the types of pruning.

d) What are Generative and Discriminative approaches to the classification problem? Provide examples.

P.T.O.

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

a) How does statistics play a role in Machine Learning, and why is it critical?

b) Explain the concepts of Ensemble Learning. How do Bagging and Boosting differ in terms of methodology and objectives?

c) Explain threshold, relu and sigmoid activation function.

d) What is a Decision Tree in Machine Learning? Explain how it works and the key components of a Decision Tree.

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

a) What is the relationship between data mining and machine learning?

b) What is a random forest, and how does it improve upon a single decision tree?

c) Differentiate between hard margin and soft margin in SVM. Why is the soft margin needed?

d) Given a set of 5 objects {A, B, C, D, E} and their distance matrix, create clusters using Agglomerative Hierarchical Clustering.

	A	B	C	D	E
A	0	2	6	10	9
B		0	5	9	8
C			0	4	5
D				0	3
E					0

Q5) Attempt any 1 question out of 2 questions (Carries 15 marks).

a) i) Explain the concept of Support Vector Machine (SVM) learning and its application in classification tasks. **[8]**

ii) What are Artificial Neural Networks (ANNs), and how do they function in Machine Learning? **[7]**

b) i) What is association rule and explain How to find the support and confidence of an association rule? **[7]**

ii) A company wants to predict whether customers will purchase a product (Buy) based on three features: Age (Young, Middle-aged, Senior) Income (Low, Medium, High), Student (Yes, No). Find the root node by using ID3 Algorithm. **[8]**

Customer	Age	Income	Student	Buy(Target)
1	Young	High	No	No
2	Young	High	Yes	Yes
3	Middle-aged	High	No	Yes
4	Senior	Medium	No	Yes
5	Senior	Low	No	No
6	Senior	Low	Yes	Yes
7	Middle-aged	Low	Yes	Yes
8	Young	Medium	No	No
9	Young	Low	Yes	Yes
10	Senior	Medium	Yes	Yes
11	Young	Medium	Yes	Yes
12	Middle-aged	Medium	No	Yes

x x x

Total No. of Questions : 5]

SEAT No. :

PC3994

[Total No. of Pages : 2

[6347]-25
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) 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.

b) Describe Nelson control chart for low defect counts.

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

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

e) State the relationship between capability index C_p and performance index C_{pk}.

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

a) Derive the control limits for $\overline{X}$ and s chart when parameters are known and unknown.

b) A process is in control with $\overline{\overline{X}} = 150$ and $\overline{S} = 2.5$, $n = 5$. The process specifications are at 145 ± 10. The quality characteristic has normal distribution. Compute C_p, C_{pk} and C_{pm} and interpret it

c) Write a note on synthetic control chart.

d) Give various sensitizing rules for Shewhart control charts.

P.T.O.

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

a) Explain the working of p-chart for fixed sample size when standards are not given.

b) Write a note on accepting sampling with rectification.

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

d) Distinguish between CUSUM and EWMA control charts.

e) Discuss the process capability ratio for an off - center process.

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

a) Write a short note on DMAIC procedure.

b) Give the comparison between Shewhart chart and CUSUM chart V-mask procedure.

c) Explain the construction and working of confirming run length chart for process fraction defective.

d) Explain how Average Run Length (ARL) plays an important role to compare two control charts fairly.

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

a) i) Define and explain the working of double sampling plan. Also find expression ASN under double sampling plan. **[8]**

ii) Describe of MIL STD sampling plan. **[5]**

iii) Distinguish between a defect and defective. **[2]**

b) i) What is the importance of OC-curve in the selection of sampling plans? Describe the impact of the sample size and acceptance number on the OC-curve. What is the disadvantage of having an acceptance number as zero? **[8]**

ii) 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) **[7]**

❋ ❋ ❋

Total No. of Questions : 5]

SEAT No. :

PC3995

[Total No. of Pages :2

[6347]-31

S.Y.M.A./M.Sc.

STATISTICS

ST-41 : Asymptotic Inference

(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 consistent estimator with illustration.

b) State invariance property of consistent estimator?

c) State any four properties of MLE.

d) Describe score test.

e) Prove or disprove: every unbiased estimator is consistent.

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

a) State and prove invariance property of CAN estimator.

b) Show that joint consistency is equivalent to marginal consistency.

c) Explain method of scoring with example.

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

e) Consider density function $f(x, y, \lambda, p) = \binom{x}{y} p^y (1-p)^{(x-y)} \frac{e^{-\lambda}\lambda^x}{x!}$, $y = 0,$ 1, 2, ...,x ; x = 0, 1, 2, ...; $\lambda > 0$ and $0 < p < 1$. Obtain CAN estimator for λ & p using method of moments.

P.T.O.

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

a) Describe Delta method with one example.

b) Explain method of percentile to obtain CAN estimator.

c) Let $X_1, X_2, ..., X_n$ is random sample from $Exp(\theta_1)$ and $Y_1, Y_2, ..., Y_n$ is random sample from $Exp(\theta_2)$ and X_i and Y_j are independent. Find consistent estimator for $P(X < Y)$.

d) Explain in detail locally most powerful test with example.

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

a) Let $X_1, X_2, ... , X_n$ is random sample from *Poisson(θ)*. Check whether $T_n = \bar{X}_n$ is BAN estimator for θ.

b) Write a note on Newton Raphson Method to obtain MLE.

c) Let $X_1, X_2, ..., X_n$ is random sample from $Gamma(\alpha, \lambda)$. Find consistent estimator for (α, λ).

d) Define most powerful test. Describe it in detail.

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

a) i) Describe in detail Likelihood Ratio Test for Multinomials Pearsons Chi-square test for goodness of fit. **[8]**

ii) Define Super-efficient estimator. Let $X_1, X_2, ... , X_n$ is random sample from $N(\theta, 1)$. Obtain super-efficient estimator for θ. **[7]**

b) i) Write a note on Newton Raphson Method to obtain MLE with one example. **[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. :

PC3996

[Total No. of Pages : 2

[6347]-32

S.Y.M.Sc.

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) Differentiate between inter-subject and intra-subject variability.

b) What is safety report in clinical trials.

c) Write a note on target population.

d) What is patient compliance?

e) Define:

i) randomization model

ii) titration model

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

a) Explain the role of sampling distributions for the valid and unbiased assessment of true efficacy and safety of the study medication.

b) Explain the treatment IND and withdraw and termination of IND.

c) Describe all phases of clinical development in clinical trials.

d) Write a note on protocol in clinical trials.

P.T.O.

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

a) Explain Placebo concurrent control. Also give some important conditions for the Ethical Inclusion of Placebo concurrent control.

b) Explain the patient selection procedure involved in Clinical Trials.

c) Explain the following:

i) Complete

ii) Permuted Block

iii) Adaptive Randomization.

d) Explain advantages and disadvantages of Treatment and Response Adaptive randomization.

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-parametrs: $AUC_{(0-\infty)}$, $AUC_{(t-\infty)}$ 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 standard titration design? List them all.

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

a) i) Classify the Clinical Trials depending upon their functioning. Explain their respective functions in brief. **[9]**

ii) List out the kinds of uncertainty and biases involved in Clinical Trials. **[6]**

b) i) Explain the following terms related with clinical trials: **[9]**

A) Subject B) Clinical end points

C) Placebo D) Evaluation

E) Treatment

ii) Discuss: Bioavailability and bioequivalence studies. **[6]**

✪ ✪ ✪

Total No. of Questions : 5]

SEAT No. :

PC3997

[Total No. of Pages : 3

[6347]-33

Second Year M.Sc.

STATISTICS

ST-42(A): Econometrics and Time Series/Analysis

(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: **[5×2=10]**

a) Define the following:

i) White Noise

ii) IID Noise

b) Define Mixed Seasonal ARMA model.

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 are constants.

d) Define ACF and PACF.

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

P.T.O.

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

a) 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 are constants.

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

i) The difference sign test

ii) The rank test

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

i) $X_t + 1.9\, X_{t-1} + 0.88\, X_{t-2} = Z_t + 0.2\, Z_{t-1} + 0.7Z_{t-2}$, *where* $\{Z_t\} \sim WN\,(0,\sigma^2)$

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

d) Prove that if $\{X_t\}$ and $\{Y_t\}$ be two stationary processes with spectral density functions $f_x\,(\lambda)\; f_y\,(\lambda)$ respectively. Show that the process $X_t + Y_t$ has spectral density function $fx\,(\lambda) + f_y\,(\lambda)$.

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

a) Write a note on moving - average method.

b) Draw correlogram for the following data.

t	1	2	3	4
X_t	2	6	5	8

c) Let $\{Y_t\}$ be stationary process with $E\,(Y_t) = 0$ for some constants a and b. Define $X_t = a + bt + S_t + Y_t$, where S_t is a seasonal component with period $d = 12$. Show that $\nabla\nabla_d X_t$ is stationary process. Express its ACVF in terms of $V_y\,(h)$.

d) Explain Indirect Least Squares Method of Econometrics.

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

a) Define *MA (q)* process. Obtain its ACVF.

b) Let $\{X_t\}$ be a time series

$X_t = Z_t + 0.8Z_{t-2}$, where $\{Z_t\} \sim WN(0, 1)$

i) Find ACVF.

ii) Compute $V\left(\frac{1}{4}\sum_{i=1}^{4} X_i\right)$

c) Discuss the AICC and BIC criteria for order selection in time series modeling.

d) Explain Dickey-Fuller test.

Q5) Attempt any one of the following.

a) i) Obtain two step best linear predictor of *AR (1)* process. Also obtain its mean square error. **[8]**

ii) Compute ψ_j & π_j coefficients for $j = 1, 2,...,5$ for the following processes: **[7]**

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

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

b) i) Explain the following algorithms: **[8]**

A) Durbin-Levinson Algorithm

B) Innovation Algorithm

ii) Obtain stationary solution of ARM A (1, 1) process. At what condition process is invertible and non-invertible. **[7]**

■■■■

Total No. of Questions : 5]

SEAT No. :

PC3998

[Total No. of Pages : 3

[6347]-34

M.A./M.Sc. - II

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 all questions. **[5×2=10]**

a) Write the concept of network model.

b) Write the any two uses of the non-linear programming problem.

c) Write any two advantages of duality.

d) Write the applications of zero-one integer programming problem.

e) Write the rules for entering variable in case of tie for the simplex method of linear programming problem.

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

a) Explain the concept of Karmarkar's polynomial time algorithm.

b) Show that dual of a dual is dual.

c) Write the procedure for Gomory's all integer cutting plane method.

d) Define the following terms:

i) Queue length

ii) Spanning tree

iii) Idle time

iv) Merge Events

v) Independent flow

P.T.O.

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

a) Write the steps in Monte Carlo simulation approach.

b) Write Dijkstra's algorithm shortest route algorithm.

c) Explain the concept of quadratic problem and its general structure.

d) Explain the method of solution to the problem in case of unbounded solution and infeasible solution in transportation problem.

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

a) Write the procedure of random number generation by linear congruential method.

b) What is the travelling salesman problem in assignment problem? How to formulate it.

c) The mathematical model for daily data of two products of manufacturing firm has the quadratic form as given below:

Max $Z = 12x + 21y + 2xy - 2x^2 - 2y^2$

Subject to constraints

$8 - y \geq 0$

$10 - x - y \geq 0$

$X, y \geq 0$

Find the solution to the problem by Beale's method when $\lambda_1 = 0$ & $\lambda_2 \neq 0$.

d) Explain the procedure of conversion of transportation problem to linear programming problem (LPP).

Q5) Attempt any one of the following.

a) i) Describe the time cost trade of procedure of network analysis. **[5]**

ii) Narrate the assumption of M/M/l :∞/FCFS. **[5]**

iii) Write the procedure to obtain the solution of dynamic programming problem by LPP. **[5]**

b) i) The total profit y, in rupees, of a drug company from the manufacturing and sale of X drug bottles is given by, $y = -(x^2/400) + 2x - 80$ **[6]**

1) How many drug bottles must the company sell in order to achieve the maximum profit?

2) What is the profit per drug bottle when this maximum is achieved?

ii) Solve the following LPP by Big-M method: **[9]**

Max $Z = 3x_1 + 5x_2$

Subject to constraints

$x_1 \leq 4$

$2x_2 \leq 12$

$3x_1 + 2x_2 = 18$

$x_1, x_2 \geq 0$

ⓘⓘⓘⓘ

Total No. of Questions : 5]

SEAT No. :

PC3999

[Total No. of Pages : 3

[6347]-35
S.Y.M.Sc.
STATISTICS
ST - 43 (A) : Survival Analysis
(2019 Pattern) (Semester - IV) (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 calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt the following. **[10]**

a) Obtain the hazard rate for the Pareto family.

b) If $T \to Exp(\lambda)$ then obtain cumulative hazard rate for random variable T.

c) Give any two distinguishing point between type I and type II censoring.

d) Suppose 20 items from an exponential distribution are put on life test and observed for 130 hours. During this period, 14 items fail with the following life times, measured in hours:

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

e) Define the following terms:

i) DMRL class of life distribution.

ii) Residual survival function.

Q2) Attempt the following. (Any THREE) **[15]**

a) Write a short note on Type I censoring, Type II censoring and Right random censoring.

b) Define normalized spacing and hence find distribution of normalized spacing.

c) Describe the procedure to obtain maximum likelihood estimator of parameters of Gamma distribution in case of complete data.

d) Obtain maximum likelihood estimator of parameter of exponential distribution in case of right random censored data.

Q3) Attempt the following. (Any THREE) **[15]**

a) Prove that hazard rate is constant if and if underlying distribution is exponential.

b) Prove the following implications:

i) NBU $\Rightarrow$ NBUE

ii) NBUE $\Rightarrow$ HNBUE

c) Obtain Kaplan-Meire estimator of survival function.

d) Define total time on test (TTT) transform $H_{F^{-1}}(t)$. Hence show that if F is IFR then $H_{F^{-1}}(t)$ is concave function of t.

Q4) Attempt the following. (Any THREE) **[15]**

a) Let $T_1, T_2, \ldots\ldots T_n$ be the random sample of size n from the exponential distribution with the parameter λ. In order to reduce the time of the experimentation experimenter has decided to terminate the experiment as soon as $m(< n)$ failures occurs. Find 100 $(1 - \alpha)$% confidence interval for the parameter λ.

b) Show that Deshpande's test statistic J_b lies within 0.5 to 1.

c) If $T_1, T_2, \ldots\ldots T_n$ are independent random variables and $Z = min \{T_1, T_2, \ldots T_n\}$ then show that Z has hazard rate $r(t) = \sum_{i=1}^{n} r_i(t)$, where $r_i(t),\ i = 1, 2, \ldots., n$ is the hazard rate of T_i.

d) Apply Gehan's test for the following hypothetical data and compute the test statistic

R_X A	3	5	7	9+	18
R_X A	13	19	20	20+	32+

Q5) Attempt the following. (Any ONE) **[15]**

a) i) Obtain confidence band for the survival function in case of complete data. **[6]**

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

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

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. :

PC4000

[Total No. of Pages : 3

[6347]-36

S.Y.M.A.M.Sc.

STATISTICS

ST - 43 (B) : Categorical Data Analysis

(2019 Pattern) (Semester - IV) (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) Define relative risk and explain with an example how it is more insightful than difference proportions.

b) Explain 'Baseline Category Logits' for nominal response variables.

c) State any two differences between Logit and Probit model.

d) Define Pearson residual for determining outliers.

e) State the Score's statistic for testing a specified value of Poisson parameter.

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

a) Write a short note on cumulative logistic regression model.

b) 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

P.T.O.

c) Give the framework of

i) Likelihood Ratio Test

ii) Score Test

d) The following table shows the frequencies of outcomes Y (success, failure) of treatments *X* (*A*,*B*) given in two clinics *Z* (1,2) to 300 patients. Investigate marginal and conditional independence of X & Y by computing appropriate measures. Interpret your findings.

Clinic	Treatment	Outcome Success	Outcome Failure
1	A	54	36
1	B	36	24
2	A	6	24
2	B	24	96

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

a) Explain log-linear analysis for analyzing dependency in contingency table.

b) Discuss the usefulness of General Linear Model in the context of categorical data analysis.

c) What is link function for general linear model? Explain the different purposes of the link function.

d) Write a short note Fisher's Exact test for 2×2 contingency table.

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

a) For given π_1 and π_2, show that the relative risk cannot be farther than the odds ratio from their independence value of 1.0.

Parameter	Estimate	Standard error	Likelihood ratio 95% confidence	limits
Intercept	0.00255	0.0003	0.0019	0.0032
Alcohol	0.00109	0.0007	–0.0001	0.0027

b) Write a short note on multiple logistic regression model.

c) Using delta method, derive the Wald confidence interval for the logit of a binomial parameter. Also, explain the use of interval to obtain one for π itself.

d) Derive the test statistic for testing the independence in two-way contingency table.

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

a) Discuss the following models for matched pair with example:

i) Comparing dependent proportions

ii) Logistic regression for matched pair

iii) Comparing margins of square contingency table

b) For the horseshoe crab data following table shows the output for the fitting of logistic regression model for π = probability of a satellite, using weight as the predictor.

i) Write the prediction equation. Interpret **[2]**

ii) Find $\hat{\pi}$ at the weight values 1.20, 2.44, and 5.20 kg, which are the sample minimum, mean, and maximum. **[3]**

iii) Find the weight at which $\hat{\pi}$ = 0.50. **[2]**

iv) Construct a 95% confidence interval to describe the effect of weight on the odds of a satellite. Interpret. **[4]**

v) Conduct the Wald or likelihood-ratio test of the hypothesis that weight has no effect. Interpret. **[4]**

❋ ❋ ❋

Total No. of Questions : 5]

SEAT No. :

PC4001

[Total No. of Pages : 2

[6347]-37

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. **[2×5=10]**

a) MCMC stands for...

b) For acceptance-rejection sampling, which one needs to be large, the support of the proposal distribution or the support of the target distribution? Justify.

c) State the properties of a kernel function.

d) Can jackknife be considered as a special case of k-fold cross validation? Justify.

e) Can every PDF be a kernel function? Justify.

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

a) Explain the similarities and differences between the bagging of trees and random forest models. In which way does the random forest algorithm improve the model based on bagging and how?

b) 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?

c) Compute the acceptance probability for Metropolis-Hastings algorithm with a target distribution $f(x)$ and a proposal distribution $q(\cdot)$.

d) Explain the procedure for multiple imputation via chained equations.

P.T.O.

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

a) Derive the Nadaraya-Watson regression estimator of the regression function $m(X) = E(Y/X)$.

b) Explain the procedure for k-nearest neighbour regression. Enlist the situations (with justification) under which k-nearest neighbour regression is expected to give better results than the linear regression and vice-versa.

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

d) What is the quantile function corresponding to the distribution function.

$$F(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1/3 & \text{if } 0 \le x < 1 \\ x/3 & \text{if } 1 \le x < 2 \\ 2/3 & \text{if } 2 \le x < 3 \\ 1 & \text{if } x \ge 3 \end{cases}$$

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

a) Explain the working of acceptance-rejection sampling.

b) What is Gibbs sampling technique? Describe brief.

c) Describe how the notion of bootstrap resampling can be used in the context of regression models.

d) Explain the working of importance sampling.

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

a) 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 $H_1 : \mu_X \neq \mu_Y$.

b) Describe the Metropolis Hastings algorithm and justify its working by providing all the appropriate formulae.

✪ ✪ ✪