Total No. of Questions : 5]

SEAT No. :

P2664

[Total No. of Pages : 3

[6074]-312
S.Y.M.Sc.
STATISTICS
ST -32 (A) : BAYESIAN INFERENCE
(2019 Pattern) (Semester-III) (4 Credits)

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

Instructions to the candidates:

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

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

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

b) Define the Following.

i) Risk function.

ii) Bayes Risk.

c) Explain the application of Bayes theorem for performing hypothesis test.

d) 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.

e) Define improper prior with illustration.

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

a) Let X be a continuous random variable with pdf $f(x|\theta) = \dfrac{3\theta^3}{x^4}, x > \theta$.

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.

P.T.O.

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

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

a) Define the Jeffrey's prior. Let X ~ *Binomial (n,p),* find the Jeffrey prior of *p*.

b) Discuss Gibbs sampling in briefly.

c) Explain the terms: subjective priors and probability matching prior.

d) Let X_1,X_2,.....Xn be the random sample of size n are 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.

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

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

b) Discuss conjugate class of priors with an example.

c) Define Bayes estime Let X_1,X_2,........,X_n be a random sample from P(λ) suppose prior distribution of λ is given by the *p.d.f.* $\pi(\lambda)=e^{-\lambda}$ if λ>0, Find the bayes estimator for λ.

d) Describe the any two application of Bayesian inference.

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

a) i) Describe relative likelihood approach to determine prior subjectively. **[7]**

ii) Explain Bayesian information criterion (BIC) with illustration. **[8]**

b) Define MAP estimator, Let X be the random variable having geometric distribution wih parameter p. The prior distribution of p is

$$\Pi(p) = 2p \quad ;0 < p < 1$$

Then find the MAP estimator of p. **[6]**

i) Let $X_1, X_2, \ldots X_n$ be random sample of size n from N(μ, σ^2), σ^2 known. The prior distribution of μ is N(μ_0, σ_0^2). Find posterior distribution of μ based on sufficient statistic for μ. **[7]**

ii) Differentiable between maximum likelihood estimator and MAP estimator. **[2]**

● ● ●

Total No. of Questions : 5]

SEAT No. :

P2674

[Total No. of Pages : 3

[6074]-417

M.Sc.-II

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 calcualtor is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

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

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

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

c) LOOCV stands for...

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

e) 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 $f(x)$?

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

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

b) 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.

c) What are the three types of missing data? Explain in detail.

d) Explain the procedure of boosting for regression trees.

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

a) What is the quantile function corresponding to the distribution function,

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

b) Describe the connection between kernel density estimation and kernel regression estimation.

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 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.Ml algorithm? Provide all the intermediate steps for calculations.

b) Explain the differences between the histogram density estimator and the naive density estimator. What are the pros and cons of each of them?

c) What is bias-variance trade off? 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 dample of size n,X_1,X_2,.....X_n is available along with the vectore of allocations of the observations x_i to the first and second components of the mixture, z_1,z_2,....z_n.

i) Compute the pmf of Z.

ii) Construct all the steps need to implement EM algorithm for above model.

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

[5+10]

● ● ●

Total No. of Questions : 5]

SEAT No. :

P-2655

[Total No. of Pages : 3

[6074]-111

F.Y. M.A./M.Sc.

STATISTICS

ST-11 : Basics of Real Analysis and Calculus (4 Credits)

(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 all the following : **[5 × 2 = 10]**

a) Define Limit point, Boundary point.

b) With illustrations define the terms Improper integral of first kind.

c) Define Norm of partition.

d) State Bolzano-Weirstrass theorem.

e) Check whether the following set is countable set. Justify your answer.

A = {2, 4, 6, ...}

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

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

b) A set is closed if and only if its complement is open.

c) Prove that there is no rational number whose square is 7.

d) Let A be a nonempty set of real numbers which is bounded below. Let –A be the set of all numbers of the form $-x$, where $x \in A$, then prove that *inf* A = –*sup* (–A).

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

a) State and prove Ratio test for convergence of series.

b) Show that the union of two countable sets is countable.

c) State and prove the relation between the Beta Function: and the Gamma function.

d) Show by an example how the sequence $\{f_n\}$ oscillates finitely.

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

a) Determine the convergence of the given series.

i) $\sum \frac{1}{2^n + 4}$

ii) $\sum \frac{3^n}{n!}$

b) Consider a function f defined on $[a, b]$. If f has a local maxima at a point $x \in (a,b)$, and $f'(x)$ exists, then show that $f'(x) = 0$. Also state analogous statement for local minima and prove it.

c) State and prove Heine-Borel theorem.

d) Define metric space. Examine whether $d(x - y) = (x - y)^2$ for x and y in $\mathbb{R}$ is a metric or not.

Q5) Attempt any one of the following :

a) Prove or disprove the following statements : **[5 × 3 = 15]**

i) Countable union of closed set is closed.

ii) Every convergent sequence is bounded sequence.

iii) If series Σa_n is convergent then series $\Sigma |a_n|$ is also convergent

iv) Every neighbourhood is an open set.

v) Set of irrational numbers is uncountable.

b) i) State & Prove fundamental theorem of calculus. **[5]**

ii) Define field. Check whether the set of integer is ordered field or not. **[7]**

iii) Give an example of a set satisfying the condition in each of the following : **[3]**

A) Set is neither Closed nor Open set.

B) Bounded set which is Closed set as well as Open set.

Total No. of Questions : 5]

SEAT No. :

P2656

[Total No. of Pages : 3

[6074]-112

F.Y. M.Sc.

STATISTICS

ST-12 : Linear Algebra and Numerical Methods

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

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

Instructions to the candidates:

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

Q1) Attempt all questions. (Each carries 2 marks). **[5×2=10]**

a) Show that the set $V = \{(x, y)\} \in R^2 \mid xy \geq 0\}$ is not a vector space of R^2.

b) Define basis of vector space.

c) Define an orthogonal matrix.

d) Define eigen value and eigen vector of matrix.

e) What is singular value decomposition?

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

a) Find the rank and the nullity of the following matrix.

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

b) Show that a set S with two or more vectors is linearly dependent if and only if at least one of the vectors is expressible as a linear combination of other vectors is S.

c) Define quadratic form and show that a real symmetric matrix is positive definite if all its eigen value are positive.

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

P.T.O.

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

a) Solve the following system of linear equations:

$$x + 2y - 4z = -4$$

$$2x + 5y - 9z = -10$$

$$3x - 2y + 3z = 11$$

b) Define an idempotent matrix and show that for an idempotent matrix A, rank (A) = trace(A).

c) Define a generalized inverse and Moore-Penrose generalized inverse with an example. Prove the uniquencess property of M-P generalized inverse.

d) Consider the quadratic from:

$$3x^2 + 2y^2 + 3z^2 + 2xz$$

Write the above Quadratic form in matrix form also find Eigen value of the matrix of Quadratic Form.

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

a) Determine whether or not S = {(1, 0, 0), (1, 1, 0), (1, 1, 1)} is basis for R^3.

b) Using spectral decomposition of a matrix, Obtain A^8 if is given by

$$A = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix}$$

c) Define Symmetric matrix and Skew Symmetric matrix with an illustration and show that every square matrix A can be expressed as sum of Symmetric matrix and Skew Symmetric matrix.

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

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

a) i) Find solution of equations using Newton Raphson iterative method

$x^2 + 3x - y - 1 = 0$, $xy + 3xy + 3y + 9 = 0$,

take initial value $x_0 = -4$, $y_0 = 6$ **[10]**

ii) Define trace of a matrix of order n. Let A and B be two square matrices then show that trace (A+B) = trace (A) + trace (B). **[5]**

b) i) Suppose the function $f(x, y)$ is known as $f(0, 0) = -1$, $f(0, 1) = 4$, $f(0, 2) = 2$, $f(1, 0) = 2$, $f(1, 1) = 0$, $f(1, 2) = -2$, $f(2, 0) = 3$, $f(2, 1) = 4$, $f(2, 2) = 3$ for these values construct the Newton's bivariate interpolation polynomial also find approximate value of $f(0.5, 1.5)$. **[10]**

ii) State and prove the necessary and sufficient condition for a quadratic form to be Positive definite. **[5]**

● ● ●

Total No. of Questions : 5]

SEAT No. :

P-2657

[Total No. of Pages : 3

[6074]-113
F.Y. M.A/M.Sc.
STATISTICS
ST-13 : Probability Distribution (4 Credits)
(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 all questions. **[5 × 2 = 10]**

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

b) State Hogg and Craig Theorem.

c) Show that, $\sigma_{xy} = 0$, if one of the regression curve is constant.

d) Prove or disprove: If X is a random variable having symmetric probability distribution around 'a', then mean, median and mode coincide.

e) If $F(x)$ and $G(x)$ are distribution functions of two independent random variables then show that $F(x)^2G(x)^3$ is a distribution function.

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

a) Find the probability generating function of the random variable X~B(n, p) and hence it's mean and variance.

b) Let X is a random variable having pdf as, $f(x)=\begin{cases} x/2, & 0 \le x \le 1 \\ 1/2, & 1 < x \le 2 \\ \dfrac{3-x}{2}, & 2 < x \le 3 \end{cases}$.

Show that moments of all order exist. Find E(X) and V(X).

P.T.O.

c) Let X be a random variable having *df* $f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, -\infty<x<\infty$.

Obtain the distribution of X^2.

d) Consider a random variable X with *df* $f(x)=\begin{cases}\theta e^{-\theta(x-a)}, & x>a\\ 0, & otherwise\end{cases}$.

Find $\xi_{\frac{1}{2}}(X)$. Also find $E\left|X-\xi_{\frac{1}{2}}(X)\right|$.

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

a) Define bivariate exponential distribution. State the memory less property satisfied by this distribution.

b) Derive probability density function of Dirichlet distribution as a bivariate beta distribution.

c) Describe one-sample sign test procedure to test $H_0: M = M_0$ against $H_1: M \neq M_0 (M > M_0 \text{ or } M < M_0)$.

d) Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with continuous density flinction. Show that, $Y_1 = \min(X_1, X_2, \ldots, X_n)$ is exponential with parameter λ iff each X_i is exponential with parameter λ.

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

a) Let X be a random variable having the following pmf.

X	–2	0	1	2
P(X = x)	1/4	1/4	1/3	1/6

Find median $\left(\xi_{\frac{1}{2}}\right)$. Also find the quantile of order $p = 0.2$ of random variable X.

b) Define non-central Student's t-random variable and obtain its density function.

c) Prove that, two quadratic forms $Q_1 = X'AX$ and $Q_2 = X'BX$ are said to be independent iff $AB = 0$.

d) Let X be a random variable of the continuous type with probability density function $f(x)$. Let $y = g(x)$ be differentiable for every x and either $g'(x) > 0$ or $g'(x) < 0$ for every x. Show that the probability density function of continuous random variable $y = g(x)$ is,

$$h(y)=\begin{cases} f[g^{-1}(y)]\left|\dfrac{d}{dy}g^{-1}(y)\right|, & \alpha<y<\beta \\ 0 & otherwise \end{cases}$$

where, $\alpha = \min\{g(-\infty),g(\infty)\}$ and $\beta = max\{g(-\infty),g(\infty)\}$

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

a) i) Suppose (X_1, X_2) follows jointly bivariate poisson variate with parameters $\lambda_1 = 7$, $\lambda_2 = 5$ and $\lambda_{12} = 4$. Compute marginal distribution of X_1 and X_2 and hence find I) $E(X_1)$ and $E(X_1)$ II) $var(X_1)$ and $var(X_2)$ III) $Corr(X_1,X_2)$ and IV) $P(X_1 = 4|X_2 = 6)$. **[7]**

ii) Obtain the probability density function of $X(r)$ in a random sample of size n from the exponential distribution with parameter α. Show that $X_{(r)}$ and $W_{rs} = X_{(r)} - X_{(s)}$, $r < s$ are independently distributed. What is the distribution of $W_1 = X_{(r+1)} - X_{(r)}$? **[8]**

b) i) State and prove Fisher Cochran's theorem. **[8]**

ii) State and prove characteristic properties of bivariate distribution function. **[7]**

Total No. of Questions : 5]

SEAT No. :

P-2658

[Total No. of Pages : 3

[6074]-114

F.Y. M.A./M.Sc. (Statistics)

ST - 14 : SAMPLING THEORY

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

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 :** **[2 each]**

a) Define the first and second order inclusion probability.

b) Explain the method of systematic sampling when population size is multiple of sample size.

c) Give two real life situations where cluster sampling is used.

d) State the assumptions used in ratio and regression method of estimation.

e) Define bias in the estimation and derive relation between variance and mean square error of the estimator.

Q2) **Attempt any Three of the following question.** **[5 each]**

a) For any sampling design P(.) prove that $E_p[I_i(s)I_j(s)] = \pi_{ij}$, where $I_i(s)$ and $I_j(s)$ are the first order inclusion indicator.

b) Let (X_i, Y_i) be the values with respect the two variables X and Y associated with the units having labels i, $i = 1, 2,..., n$. Let $\overline{x}_n, \overline{y}_n$ be the sample means and $S_{XY} = \frac{1}{N-1}\sum_{i=1}^{N-1}\left(X_i - \overline{X}_N\right)\left(Y_i - \overline{Y}_N\right)$ then prove that under simple random sampling without replacement $\text{Cov}\left(\overline{x}_n, \overline{y}_n\right) = \frac{N-n}{nN}S_{XY}$.

P.T.O.

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) In simple random sampling without replacement, obtain the sample size for the specified margin of error in the estimation of population mean $\overline{Y}_N$ using sample mean $\overline{y}_n$.

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

a) Discuss Balanced systematic sampling with an illustration.

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

c) Obtain the optimum sub-sample size for two stage sampling with equal cluster size.

d) Explain Cumulative total method of selecting random sample using PPSWR scheme. Also state limitations of this method.

Q4) Attempt any Three of the following question. [5 each]

a) In case of two stage sampling with equal number of second stage in each first stage unit, prove that mean of second stage units in the sample is an unbiased estimator of population mean.

b) What is the difference between the cluster sampling and stratified sampling? Show that weighed average of stratum sample mean with weight as size of the stratum is an unbiased estimator of the population mean.

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

$$Est(R.E) = \left[s_b^2 + \left(1 - M^{-1}\right)s_w^2\right]Ms_b^2$$

d) Obtain mean square error of the ratio estimator $\overline{y}_R = \frac{\overline{y}_n}{\overline{x}_n}\overline{X}_N$ of population mean.

Q5) **Attempt any one of the following question.**

a) 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's allocation. Assume N_1 and N_2 large. **[7]**

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

ii) Define regression estimator of population mean in double sampling. Show that it is biased estimator. Also find the mean squared error of this estimator. **[8]**

x x x

Total No. of Questions : 5]

SEAT No. :

P2659

[Total No. of Pages : 3

[6074]- 211
F.Y. 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 & scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

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

a) Define

i) Sigma field

ii) Increasing sequence of sets.

b) Prove of disprove: $\{x\}$ is a Borel set.

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

d) Suppose X and Y are simple random variables defined on a probability space $(\Omega, \mathbb{A}, P)$. If $X \geq Y$ on Ω then show that $E(X) \geq E(Y)$.

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

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

a) Prove that:

i) Countable union is distributive over finite union.

ii) Countable intersection is distributive over finite intersection.

P.T.O.

b) 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.

c) Show that, $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X$.

d) Let $\Omega = \{a, b, c, d\}$, $\mathbb{A} = \{\phi, \Omega, \{a, b\}, \{c, d\}\}$, $X(a) = X(b) = -1$, $X(c) = X(d) = 2$. Examine whether X is $\mathbb{A}$-measurable.

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

a) 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 $\sigma(\mathbb{C}_1) = \sigma(\mathbb{C}_2)$.

b) State and prove C_r-inequality.

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

$$F(x) = \begin{cases} 0 & if\ x < -2 \\ \frac{1}{3} & if\ -2 \leq x < 0 \\ \frac{1}{2} & if\ 0 \leq x < 5 \\ \frac{1}{2} + \frac{(x-5)^2}{2} & if\ 5 \leq x < 6 \\ 1 & if\ x \geq 6 \end{cases}$$

Obtain decomposition of DF F(x).

d) Suppose X is a random variable defined on a probability space $(\Omega, \mathbb{A}, P)$ to $(\mathbb{R}, \mathbb{B})$. Show that a set function $P_x(S)$ defined on $(\mathbb{R}, \mathbb{B})$ as $P_x(S) = P(X^{-1}(S))$, for $S \in \mathbb{B}$ is a probability measure on the measurable space $(\mathbb{R}, \mathbb{B})$.

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

a) State and prove disjointification lemma.

b) 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.

c) Show that $X_n \xrightarrow{p} 0$ *iff* $E\left\{\frac{|X_n|}{1+|X_n|}\right\} \to 0$ as $n \to \infty$.

d) Define probability measure. Prove that

i) Probability measure is subtractive.

ii) Probability measure is finite sub-additive.

Q5) Attempt any one of the following.

a) i) Suppose X and Y are simple random variable 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 Borel 0-1 law. **[7]**

b) i) Define the convergence in probability, convergence in distribution.

Let $X_n \xrightarrow{L} X$ and $Y_n \xrightarrow{p} c$. Then prove that, **[8]**

1) $X_n + Y_n \to X + c$

2) $X_n Y_n \to cX$

3) $\frac{X_n}{Y_n} \to \frac{X}{c}, if\ c \neq 0$

ii) If X and Y are independent random variables then show that $E(XY) = E(X)\,E(Y)$. **[7]**

◀▶ ◀▶ ◀▶

Total No. of Questions : 5]

SEAT No. :

P-2660

[Total No. of Pages : 3

[6074]-212

F.Y. M.Sc./M.A.

STATISTICS

ST-22: Regression Analysis

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

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

Instructions to the candidates :

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

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

a) Explain the need of adjusted R^2 over the limitations of R^2.

b) If $e_2 = 2.0256$, $MS_{res} = 3.8625$ and $h_{22} = 0.0707$ then obtain the standardized residual (d_2) and state whether the related observation is outlier or not?

c) Give the inverse Gaussian link function for the generalized linear model.

d) In polynomial regression model, test the significance of the highest order term.

e) How condition number and condition index is useful in detection of multicollinearity?

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

a) Discuss the method of maximum likelihood estimation for estimating parameters of multiple linear regression model.

b) Obtain $100(1-\alpha)\%$ joint confidence region for both of the estimates of β_0 and β_1 of the centered version of the linear regression model.

c) Derive the likelihood ratio test for testing $H_0 : R\underline{\beta} = \underline{r}$.

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

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

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

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

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

d) Explain the measure of influence Cook's D-Statistic.

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

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

b) Explain the development of the PRESS statistic.

c) Discuss Likelihood ratio test in logistic regression model.

d) For the nonlinear regression model $y = \theta_1 e^{\theta_2 + \theta_3 x} + \varepsilon$, obtain least squares normal equation. Now, linearize the model and obtain least square estimates of parameters. Compare the two methods.

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

a) i) Explain the terms: leverage points and influential points. Also, describe the importance of detecting influential observations. **[6]**

ii) For the multiple linear regression model, explain the role of extra sum of squares principal in testing the relative importance of regressor variables. Interpret the various components of regression sum of squares in case of model with three regressor variables. **[6]**

iii) Show that solution of normal equation actually minimizes the residual sum of squares. **[3]**

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

i) Obtain the normal equation for it and hence show that, its BLUE is unique almost surely. **[7]**

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

$$\mathrm{Cov}(\underline{\lambda}'_{(1)}\hat{\underline{\beta}}, \underline{\lambda}'_{(2)}\hat{\underline{\beta}}) = \underline{\lambda}'_{(1)} S - \underline{\lambda}'_{(2)}\sigma^2.$$ **[5]**

iii) Show that, co-variance between any function belongs to error space and any BLUE is zero. **[3]**

✷ ✷ ✷

Total No. of Questions : 5]

SEAT No. :

P2661

[Total No. of Pages : 3

[6074]-213

F.Y.M.Sc.

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) Explain the term likelihood equivalence.

b) Distinguish between Most powerful and uniformly most powerful test.

c) Define Ancillary statistics with an illustration.

d) State Chapman-Robin bounds.

e) Define the term uniformly most accurate confidence bound.

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

a) State and prove Cramer-Rao Inequality.

b) Define Pitman family. If $a(\theta)\downarrow$ and $b(\theta)\uparrow$ then find the minimal sufficient statistic for θ.

c) Define complete sufficiency statistic. Let $X_1, X_2, \ldots\ldots, X_n$ be random sample from Bernoulli (p), find complete sufficient statistic for p.

d) Define Fisher information matrix and find information matrix $I(\alpha, \lambda)$ if $X \sim G(\alpha, \lambda)$.

P.T.O.

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

a) State and prove Basu's theorem.

b) If $X_1, X_2, \ldots\ldots., X_n$ is random sample from $f(x,\theta)$, $\theta \in \Omega$ which is member of one parameter exponential family and joint pmf is given by

$L(x, \theta) = u(\theta)T(x) + nV(\theta) + w(x)$ where, $T(x) = \sum_{i=1}^{n} k(x_i)$ then show that T is minimal sufficient statistic for θ.

c) Let $X_1, X_2, \ldots\ldots., X_n$ be a random sample from $N(\mu, 1)$. Obtain unbiased estimator of μ^2 which is function of $\bar{X}$.

d) Let $X_1, X_2, \ldots\ldots\ldots\ldots\ldots\ldots, X_n$ be a random sample from $N(\theta, 1)$. 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 and H_1 be given as possible distributions :

X	1	2	3	4	5	6
$P_1(x)$	0.01	0.01	0.01	0.01	0.01	0.95
$P_2(x)$	0.05	0.04	0.03	0.02	0.01	0.85

Find the MP test of size 0.03 and also its power for testing $H_0 : X \sim P_1(x)$ against $H_1 : X \sim P_2(x)$.

b) State and prove Rao-Blackwell Theorem.

c) Let $X_1, X_2, \ldots\ldots., X_n$ be a random sample from $N(\mu, \sigma^2)\mu$ is known find $I_n(\sigma^2)$.

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) State and prove Neyman-Pearson Lemma. **[7]**

ii) Define uniformly most accurate confidence interval. Let $X_1, X_2, \ldots\ldots, X_n$ 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) Let X has binomial distribution with parameter n and p. Suppose that n is given and the unknown parameter p has prior distribution, which is uniform on the interval [0, 1]. By considering the squared error loss function and observation $X = n$. Find Bayes estimator of p and the median of posterior distribution of p. **[8]**

ii) Let $f(x,y,\lambda,p)=\binom{x}{y}p^y(1-p)^{(x-y)}\dfrac{e^{-\lambda}\lambda^x}{x!}$, $y = 0, 1, 2, \ldots\ldots, x$; $x = 0, 1, 2, \ldots..$; $\lambda > 0$ 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 $(p, \lambda)'$. **[7]**

✦✦✦✦

Total No. of Questions :5]

SEAT No. :

P2662

[Total No. of Pages : 3

[6074]-214

M.Sc.-I

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 all of the following: **[5×2=10]**

a) If $\underline{X} \sim N_p\left(\underline{\mu}, \Sigma\right)$ then state the distribution $A\underline{X}-B$ where A is any *qxp* matrix and B is any *qx*1 vector.

b) Define the non-singular multivariate normal distribution and state its probability density function.

c) State the orthogonal factor model with all the assumptions.

d) Define the term 'Wishart matrix'. Why Wishart distribution is called as generalization of chi-square distribution?

e) Define the term 'canonical correlation'.

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

a) For the data matrix given below, obtain the sample correlation matrix:

$$X = \underline{\begin{bmatrix} \dfrac{X_1}{X_2} \end{bmatrix}} \text{ where } \underline{X_1} = [4 \;\; 8] \text{ and } \underline{X_2} = [5 \;\; 9]$$

b) Let $\underline{X} \sim N_3\left(\underline{\mu}, \Sigma\right)$ with $\underline{\mu} = [1 \;\; -1 \;\; 2]'$ and $\sum = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$

Check whether X_1 and $U=X_1+3X_2-2X_3$ are independent.

P.T.O.

c) Explain the term 'Discriminant analysis. Derive Fisher's Linear discriminant function for two populations.

d) For the dispersion matrix Σ given below, calculate the percentage of variation explained by first and second principal components,

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

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

a) What is scree plot? Explain its application in principal component analysis with one illustration.

b) Let $\underline{X} \sim N_p\left(\underline{\mu}, \Sigma\right)$ then find the moment generating function of $\underline{X}$.

c) Let $\underline{X} \sim N_4\left(\underline{\mu}, \Sigma\right)$ with $\underline{\mu} = [4\ 3\ 2\ -1]'$ and $\Sigma \begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{bmatrix}$

Let $\underline{X}^{(1)} = [X_1\ X_2\ ]'$ and $\underline{X}^{(2)} = [X_3\ \ X_4]'$. Then obtain the distribution of $\underline{X}^{(1)}|\underline{X}^{(2)} = [3\ \ -4]'$

d) What is MANOVA? Explain the test procedure for 'One way MANOVA'.

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

a) Let $\underline{X} = \left[\underline{X}^{(1)}_{qx1}\ \underline{X}^{(2)}_{p-qx1}\right]' \sim N_p(\underline{\mu}, \Sigma)$, then prove that $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are independent iff $\Sigma_{12} = 0$

b) Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$. then find the maximum likelihood estimator for mean vector $\underline{\mu}$ (with Σ known).

c) Explain the term 'Dendogram'. why it is used in cluster analysis?

d) if D is any symmetric matrix of order 'p' with distribution $W_p(n,\Sigma)$ and it is partitioned as $D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$ then show that $D \sim W_m(n, \Sigma_{11})$.

Q5) Answer any one of the following:

a) i) Let $\underline{X} = \left[\underline{X}^{(1)}_{q\times 1}\ \underline{X}^{(2)}_{(p-q)\times 1}\right]' \sim N_p(\mu, \Sigma)$. Derive the conditional distribution of $\underline{X}^{(2)}|\underline{X}^{(1)} = x^{(1)}$

ii) Test the hyposthesis $H_0 : \underline{\mu} = \begin{bmatrix} 7 \\ 11 \end{bmatrix} \text{ Vs } H_1 : \underline{\mu} \neq \begin{bmatrix} 7 \\ 11 \end{bmatrix}$ at 5% l.o.s using the dataset $X = \begin{bmatrix} 2 & 8 & 6 & 8 \\ 12 & 9 & 9 & 10 \end{bmatrix}$.

[7+8=15]

b) i) For a random vector $\underline{X}= [X_1 X_2 X_3]'$ with dispersion matrix

$\Sigma = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$. find the population principal components.

ii) Explain the procedure of Hierarchical Agglomerative Clustering with the help of different types of linkages.

[10+5=15]

● ● ●

Total No. of Questions : 5]

SEAT No. :

P-2663

[Total No. of Pages : 3

[6074]-311
M.Sc. (Part-II)
STATISTICS
ST-31 : Applied Stochastic Processes
(2019 Pattern) (4 Credits) (Semester - III)

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

Instructions to the candidates :

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

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

a) What is a Markov property?

b) State postulates of Poisson process.

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

d) Explain stopping time variable with an example.

e) Define: i) Recurrent state ii) Transient state

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

a) Explain the compound Poisson process with its application. Obtain the mean and variance for compound Poisson process.

b) Explain branching process and calculate the probability of extinction π_0 for

i) $p_0 = \frac{1}{6}, p_1 = \frac{1}{2}, p_3 = \frac{1}{3}$ ii) $p_0 = \frac{1}{4}, p_1 = \frac{1}{2}, p_3 = \frac{1}{4}$

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

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

P.T.O.

Calculate : a) $P[X_2=1|x_0 = 2]$

b) $P[X_2=1]$

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

d) Define ergodic Markov chain and prove for an irreducible Markov chain there exist unique stationary distribution.

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

a) Define a renewal process. Also, state the elementary renewal theorem with its application.

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

c) Prove or disprove: For a markov chain periodicity is a class property.

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

b) 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<l.

c) Give the two definitions of Poisson process and establish their equivalence.

d) Vehicle stopping at roadside restaurant form a Poisson process with rate $\lambda = 20$/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 vith 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 $\{N(t)\geq 0\}$ is a Poisson process. Given that $N(t) = n$, prove that n arrival times $T_1, T_2, ..., T_n$, have the same distribution as the order statistic corresponding to n independent random variables uniformly distributed on (0,t). **[4]**

iii) Suppose $\{N(t)\geq 0\}$ is renewal process and $m(t)$ be renewal function then show that the renewal process satisfies the equation $m(t)=F(t)+\int_0^t m(t-x)\,f(x)\,dx$, where $F(t)$ and $f(x)$ be the respective distribution function and density function of inter-arrival time. **[4]**

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 1) Wald equation 2) Kolmogorov forward equation **[4]**

iii) Define Yule process. For Yule process show that,

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. **[7]**

❍❍❍❍

Total No. of Questions : 5]

SEAT No. : [ ]

P-2665

[Total No. of Pages : 4

[6074]-313
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 table & scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meanings.*

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) Define BIBD, with usual notations prove that:

i) $r \geq k$

ii) $b \geq v + r - k$

c) Check whether following block design is connected :

Block-I	Block-II
1	1
2	4
3	5

d) Define the following :

i) Treatment contrast

ii) Elementary treatment contrast

e) Distinguish between total and partial confounding

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

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

b) 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.

c) Give the statistical analysis of 2^3 factorial experiment through RBD.

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

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

a) To investigate the effect of 6 month training programme on blood pressure, blood pressure measured of 6 peoples (subjects) measured at 3 separate time points: pri-exercise, 3-month and post exercise. The data shown below :

	Blood Pressure		
Subjects	pri-exercise	3-month	post exercise
1	45	50	55
2	42	42	45
3	36	41	45
4	39	35	40
5	51	55	59
6	44	49	56

Analyse the data and draw the conclusion.

b) Explain Two-way ANOVA.

c) Define BIBD, with usual notations prove that:

i) $vr = bk$

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

d) Show that BIBD is connected block design.

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

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

b) Check whether following design is connected and/or orthogonal and/or variance balanced.

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

c) Explain Circumscribed (CCC), Face Centered (CCF) and Inscribed (CCI) designs of CCD for number of factors k = 2.

d) 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

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

a) i) Explain following method for comparing pairs of treatment means : **[8]**

A) Duncan's Multiple Range Test

B) Newman Keuls Test

ii) A router is used to cut locating notches on a printed circuit board. The vibration level at the surface of the board as it is cut is considered to be a major source of dimensional variation in the notches. Two factors are thought to influence vibration: bit size (A) and cutting speed (B). Two hit sizes(1/16 and 1/8 inch) and two speeds (40 and 90 rpm) are selected, and four hoards are cut at each set. of conditions shown below. The response variable measured as the resultant vector of three accelerometers (x, y and z) on each test circuit board. **[7]**

	Replicate			
Treatment Combination	I	II	III	IV
'1'	18	18	12	14
a	27	24	22	22
b	15	14	15	14
ab	41	43	36	39

Analyze the data from this experiment and draw conclusions.

b) i) Analyze the following data : **[8]**

Block-I	Block-II	Block-III	Block-IV
A = 4	C = 3	A = 5	B = 7
B = 6	D = 4	D = 4	C = 3

ii) Explain the concept signal to noise (S/N) ratio. **[7]**

⌘⌘⌘

Total No. of Questions : 5]

SEAT No. :

P-2666

[Total No. of Pages : 2

[6074]-314
M.Sc. (Part - II)
STATISTICS
ST - 34 : Machine Learning
(2019 Pattern) (4 Credits) (Semester - III)

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

Instructions to the candidates:

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

Q1) **Attempt all questions. (Each carries 2 marks).**

a) What is kernel trick?

b) Define supervised machine learning with example.

c) What is tree pruning?

d) What is cross validation?

e) What are various signal functions of Neurons?

Q2) **Attempt any 3 questions out of 4 questions (Each carries 5 marks).**

a) What are the characteristics of clustering algorithms?

b) What is cross validation?

c) What are basic steps in Apriori algorithm?

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 (Each carries 5 marks).**

a) What is K-nearest Neighbor (KNN) algorithm and explain how does it works?

b) What are hard-margin and soft margin support vector machines?

c) Explain Biological Neural Network and Artificial Neural Network.

d) What are the advantages and disadvantages of backpropagation?

Q4) **Attempt any 3 questions out of 4 questions (Each carries 5 marks).**

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

b) What is a confusion matrix explain all the term use in it.

c) Explain the CART algorithm for decision trees.

d) Explain the basic components of learning process.

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

a) i) Explain the different linkage methods used in the Hierarchical clustering Algorithm. **[7]**

ii) The species classification according to attributes color, legs, height and smelly is given in the table.

Color	Legs	Height	Smelly	Species
White	3	Short	Yes	M
Green	2	Tall	No	M
Green	3	Short	Yes	M
White	3	Short	Yes	M
Green	2	Short	No	H
White	2	Tall	No	H
White	2	Tall	No	H
White	2	Short	Yes	H

Using Naïve Bayes classifier estimate the probability values for new instance (Color = Green, Legs = 2, Height = Tall, Smelly = No) **[8]**

b) i) Explain the k-means clustering algorithm. Why Euclidean distance prefer than Manhattan distance in it? **[7]**

ii) What do you mean by random Forest Algorithm and explain how it works? **[8]**

▽▽▽▽

Total No. of Questions : 5]

SEAT No. :

P2667

[Total No. of Pages : 2

[6074]-315

M.Sc.-II

STATISTICS

ST 32 (B) : Statistical Quality Control

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

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

Instructions to the candidates:

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

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

a) Give the use and interpretation of C_p and C_{pk}.

b) How is process shift determined on control chart?

c) State any two advantages of sampling inspection.

d) For the $\overline{X}$ control chart, find the probability of type I error, assuming σ is constant. (Take $C_4 = 0.9213$ and $L = 3$).

e) Describe Nelson control chart for low defect counts.

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

a) Distinguish between attribute and variable control charts.

b) Write note on effect of non - normality on process capability ratio.

c) Discuss EWMA control chart for monitoring process mean.

d) Give various sensitizing rules for Shewhart control charts.

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

a) Write a note on multivariate control chart.

b) A car has gone out of control during a snowstorm and struck a tree. Construct a cause - and - effect diagram that identifies and outlines the causes of the accident.

c) Obtain the 100 $(1 - \alpha)$ % confidence interval for capability index C_p and C_{pk}. Also, give the testing procedure for testing C_p.

d) Write a note on Tabular CUSUM chart for monitoring the process mean.

P.T.O.

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

a) Write a short note on an acceptance sampling plan.

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

c) Describe the situations were Moving range chart can be used.

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

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

a) i) Sample of n = 6 items each are taken from a process at regular intervals. Quality characteristics is measured $\overline{X}$ and R values are calculated for each sample. After 50 sample we have $\sum_{i=1}^{50} X_i = 2000$ and $\sum_{i=1}^{50} R_i = 200$. **[8]**

Compute the control limits for the $\overline{X}$ and R control chart.

1) If the specification limits are 41$\pm$5. What is your conclusion regarding the ability of process to produce items conforming to specifications?

2) Find the probability of scrap and rework when mean is 200.

ii) Write a note on synthetic control chart. **[4]**

iii) Write a short note on AOQ curve. **[3]**

b) i) Derive the control limits for $\overline{x}$ and s^2 chart when parameters are known and unknown. **[4]**

ii) Describe of MIL STD sampling plan. **[4]**

iii) 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. :

P2668

[Total No. of Pages : 2

[6074]- 411
M.Sc.-II
STATITICS
ST-41 : Asymptotic Inference
(2019 Pattern) (Semester - IV) (4 Credits)

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

Instructions to the candidates:

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

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

a) Define BAN estimator with illustration.

b) Give Crammers regularity conditions.

c) State any four properties of m.l.e.

d) Describe score test.

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

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

a) State and prove invariance property of CAN estimator.

b) Let y = $\alpha + \beta x + \in$ be a linear model with $E(\in) =$ & 0 *var* $(\in) = \sigma^2$ then show that $\hat{\beta}_n = \frac{S_{xy}}{S_{xX}} \xrightarrow{p} \beta$ *and* $\hat{a}_n = \overline{y}_n - \hat{\beta}_n \overline{x}_n \xrightarrow{p} \alpha$.

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

d) Let $X_1, X_2, ..., X_n$ is random sample from $U(\theta - 1, \theta + 1)$ Find maximum likelihood estimate of θ.

P.T.O.

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

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

b) Let $X_1, X_2,..., X_n$ 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) Write a note on Newton Raphson Method to obtain MLE with one example.

d) Let $X_1, X_2, - - - -, X_n$ be random sample from *Cauchy*(μ, λ). Find consistent estimator for (μ, λ)

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

a) Let $X_1, X_2,...,X_n$ 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) Consider density function.

$f(x,\theta) = \theta x^{\theta} e^{-x^{\theta}}; x > 0 \,\&\, \theta > 0.$ Obtain CAN estimator for θ based on percentile method. Also obtain its asymptotic variance.

d) Describe Delta method with one example.

Q5) Attempt any one equestions. **[1×15=15]**

a) i) Describe in detail Likelihood Ratio Test for Multinomials Pearson's Chisquare test for goodness of fit.

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

[8+7]

b) i) Describe general strategy of obtaining 100(1 – α)% asymptotic confidence interval (ACI) for a parameter θ based on sample size n from (*x*, θ). Construct 95% ACI for θ when sample of size n is drawn from Exponential (mean = θ).

ii) Prove the consistency property of MLE of parameter of distribution belongs one parameter exponential family.

[8+7]

◀▶ ◀▶ ◀▶

Total No. of Questions : 5]

SEAT No. : [ ]

P2669

[Total No. of Pages : 3

[6074]-412

M.Sc.-II

STATISTICS

ST - 44 (B) : Analysis of Clinical Trials

(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 meaning of blinding?

b) State the goals of clinical trials.

c) Write a note on combination trials.

d) Explain the difference between washout period and run in period.

e) Explain the role Biostatistician in the planning and execution of clinical trials.

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

a) Discuss testing procedure for carryover, direct drug and period effect in standard 2×2 crossover design using analysis of variance.

b) i) What are the primary and secondary pharmacokinetic parameters.

ii) What are the roots of administration of the drug?

c) What is Interim analysis and safety report in clinical trial?

d) Who are the members of advising committee? which important issues the advising committee is supported to address?

P.T.O.

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

a) What is patient compliance? What is the difference between missing values and dropouts?

b) Explain the method of permuted block randomization with example and its advantages over complete randomization.

c) Write a note on

i) Target population

ii) Patient selection

d) What is population model and invoke population model. State the difference between them.

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

a) Describe the following designs with the appropriate layouts:

i) Balance Incomplete block design.

ii) William's design and Balaam's design

b) Define the following terms:

i) Intersubject variability

ii) Treatment IND

c) Discuss the Higher-Order crossover design.

d) What are the difference between analysis of multicenter trial and meta-analysis?

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

a) i) List out the kinds of uncertainty and biases involved in clinical trials? **[8]**

ii) Explain classical confidence interval for testing the average bioequivalence. **[7]**

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 the role of sampling distributions for the valid and unbiassed assessment of true efficacy and safety of the study medication. **[6]**

● ● ●

Total No. of Questions : 5]

SEAT No. :

P-2670

[Total No. of Pages : 3

[6074]-413
S.Y. M.Sc. (Semester - IV)
STATISTICS
ST-42(A) : Econometrics and Time Series
(2019 Pattern)

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) Explain the general approach for time series modelling.

b) Define the following :

i) Weakly stationary

ii) Strict stationary

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

d) Define AICC and BIC.

e) Explain Dickey-Fuller test.

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

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

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

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

i) Find ACVF.

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

P.T.O.

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

i) The turning point test

ii) The difference sign test

d) Obtain one step best linear predictor of $AR(1)$ process. Also obtain its mean square error.

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

a) Obtain ACVF of causal $ARMA(1, 1)$ process.

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

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

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

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

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

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

a) What is random walk? Prove that it is not stationary.

b) Prove that $E(X_{n+h} - P_n X_{n+h})^2 = v_0 - a'_n v_n(h)$

c) Obtain ACF of the following time series :

i) $X_t = Z_t - 1.2Z_{t-1} - 1.6Z_{t-2}$, where $\{Z_t\} \sim WN(0, 0.25)$

ii) $X_t = Z_t + 1.75Z_{t-1} + 0.625Z_{t-2}$ where $\{Z_t\} \sim WN(0, \sigma^2)$

d) Explain exponential-Smoothing method. Apply it to estimate the trend for the following data. Take $\alpha = 0.6$.

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

Q5) Attempt any one of the following :

a) i) Compute ψ_j & π_j coefficients for the following processes : **[8]**

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

ii) Estimate the parameters of stationary $AR(p)$ process using Yule-Walker equations. **[7]**

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

A. Graphical Method

B. Moving-Average Method

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

t	1	2	3	4
X_t	6	8	5	12

Total No. of Questions : 5]

SEAT No. :

P-2671

[Total No. of Pages : 2

[6074]-414
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 table & scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meanings.***

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

a) State the assumptions of LPP.

b) What is Basic feasible solution? Explain.

c) What is meant by transhipment problem?

d) Give any two applications of dual simplex method.

e) Define :

i) Slack variable

ii) Surplus variable

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

a) What is Assignment problem (A.P). Write down general from of A.P. State the assumptions clearly.

b) Define convex set. Prove that hyperplane is a convex set.

c) Solve the following L.P.P. by graphical method

Maximize $Z = 3x + 2y$

Subject to $-2x + 3y \leq 9$

$3x - 2y \leq -20$

$x, y \geq 0$

d) Discuss Big. M method solving L.P.P.

P.T.O.

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

a) State and prove optimality condition of solving LPP.

b) Write an algorithm of two phase method of solving LPP.

c) Prove that if the primal has an unbounded solution then its dual has no feasible solution.

d) Discuss 'MODI' method of solving unbalanced Transportation problem (T.P.).

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

a) Discuss various steps of simulation process.

b) Discuss the procedure of solving dynamic programming problem.

c) Explain Gomory's cutting plane method of solving pure integer programming problem (IPP).

d) Solve the following quadratic programming problem by Beale's method.

Maximize $Z = 10x_1 + 4x_2 - x_1^2 + 4x_1x_2 - 5x_2^2$

Subject to $x_1 + x_2 = 0$

$x_1, x_2 \geq 0$

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

a) i) Derive Kahn Tucker's necessary and sufficient condition for an optimal solution of Q.P.P.

ii) Explain Wolfe's method to solve Q.P.P.

[7 + 8 = 15]

b) i) Consider the following I.P.P.

Maximize $Z = 2x_1 + 3x_2$

Subject to $6x_1 + 5x_2 \leq 25$

$x_1 + 3x_2 \leq 10$

$x_1, x_2 \geq 0$ & integer

Solve I.P.P. using Branch & Bound method.

ii) State & prove the reduction theorem of Assignment problem (A.P.)

[8 + 7 = 15]

⌘⌘⌘

Total No. of Questions : 5]

SEAT No. :

P2672

[Total No. of Pages : 2

[6074]-415

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

Q1) Attempt each of the following question : **[2 Each]**

a) Obtain hazard rate for the Makeham family of life distribution.

b) Explain the concept of Stragged and aligned entries.

c) Define the following :

i) U-statistic.

ii) Covariate with example.

d) Show that empirical distribution function is an unbiased estimator of the distribution function. Also check the consistency of the estimator.

e) Suppose 15 items from an exponential distribution are put on life test and observed for 150 hours. During this period 10 items fail with the following life times, measured in hours : 3, 19, 23, 26, 27, 37, 38, 41, 45, 58.

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

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

a) State and Prove Cauchy functional equation.

b) Prove the implications $\text{IFR} \Rightarrow \text{IFRA}$ and $\text{IFR} \Rightarrow \text{DMRL}$.

c) Discuss the concept of Bathtub failure rate with an illustration.

d) If random variable T has Weibull distribution with scale parameter λ and shape parameter γ then find the distribution of T^{γ}.

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

a) Explain the procedure to find maximum likelihood estimators of the parameters of Gamma distribution in case of complete data.

b) Discuss the two methods of testing the exponentiality of data graphically.

c) The following failure and censor time (in operating hours) were recorded on 12 turbine veins : 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.

d) Define the term is censoring. Hence discuss type I and type II censoring with an illustration.

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

a) Derive Green-woods formula for variance of actuarial estimator of the survival function.

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

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

d) Define spacing and Hence find distribution of spacing.

Q5) Attempt any one of the following question :

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

ii) Apply Gehan's test for the following hypothetical data and compute the test statistic. **[6]**

$R_X(A)$	3	5	7	9+	18
$R_X(B)$	3	5	7	9+	18

b) i) Explain Deshpande's test for testing exponentiality against IFRA class of life distribution. **[10]**

ii) If $T_1, T_2, ..., T_n$ are independent; $T_i \to Exp(\lambda_i)$ for $i = 1, 2, ...,n$ and $T = \min\{T_1, T_2, ..., T_n\}$ then show that $T \to Exp\left(\sum_{i=1}^{n} \lambda_i\right)$. **[5]**

✦✦✦✦

Total No. of Questions : 5]

SEAT No. :

P2673

[Total No. of Pages : 3

[6074]-416

S.Y.M.Sc.

STATISTICS

ST-43(B) : Categorical Data Analysis

(2019 Pattern) (4 Credits) (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) State any two real life application of categorical regression model.

b) Define a 'nominal variable' with an example.

c) State the Wald's statistic for testing a specified value of Poisson parameter.

d) Define the test statistic for testing the significance of individual regression coefficient in Binary Logistic regression.

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

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

a) Give the general frameworks of the

i) Likelihood Ratio test.

ii) Score Test.

b) Explain logit model in detail.

c) Write any four real life application of multinomial logistic regression.

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

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

P.T.O.

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

a) Attempt the following :

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

ii) State the assumptions of Poisson Regression Model.

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

b) Explain Log-Linear Analysis for analyzing dependency in contingency table.

c) Which scale of measurement is most appropriate for the following variables - nominal, or ordinal interval.

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

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

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

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

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

d) Discuss about Cochran Mantel Haenszel test statistics.

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

a) Explain Multiple Logistic regression in detail.

b) Write the Fisher exact test for 2*2 contigency table with example.

c) X = mother's alcohol consumption and Y = whether a baby has sex organ malformation. With scores (0, 0.5, 1.5, 4.0, 7.0) for alcohol consumption, ML fitting of the linear probability model has the output :

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

i) State the prediction equation and interpret the intercept and slope.

ii) Use the model fit to estimate the

A) probabilities of malformation for alcohol levels 0 and 7.0,

B) relative risk comparing those levels.

d) Explain GLM and its components.

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

a) A study used logistic regression to determine characteristics associated with Y = whether a cancer patient achieved remission (1 = yes). The most important explanatory variable was a labeling index (LI) that measures proliferative activity of cells after a 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
....						

i) State the prediction equation. Interpret. **[3]**

ii) Conduct a Wald test for the LI effect. Interpret. **[3]**

iii) Construct a Wald confidence interval for the odds ratio corresponding to a 1-unit increase in LI. Interpret. **[3]**

iv) Conduct a likelihood-ratio test for the LI effect. Interpret. **[3]**

v) Construct the likelihood-ratio confidence interval for the odds ratio. Interpret. **[3]**

b) Discuss the models for matched pair with example.

i) Comparing dependent proportions. **[5]**

ii) Logistic regression for matched pair. **[5]**

iii) Comparing margins of square contingency table. **[5]**

✦✦✦✦