Total No. of Questions : 4]

SEAT No. :

PD3696

[Total No. of Pages : 2

[6487]-101

M.Sc. - I

STATISTICS

STS-501-MJ : Fundamentals of Analysis and Calculus

(2024 Credit Pattern) (Semester-I) (2 Credits)

Time : 2 Hours] *[Max. Marks : 35*

Instructions to the candidates:

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

Q1) Attempt all questions. **[5×1=5]**

a) Calculate limit supremum of $a_n = (-1)^n \left(1+\frac{1}{n}\right)$.

b) What is the difference between absolute convergence and conditional convergence?

c) State the Cauchy criterion for convergence.

d) State Roll's theorem.

e) Define the concept of a limit point.

Q2) Attempt any two of the following: **[2×5=10]**

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

b) State and prove mean value theorem.

c) 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 in f A = –*sup*(–A).

P.T.O.

Q3) Attempt any two of the following: **[2×5=10]**

a) Suppose $a_1 \geq a_2 \geq \ldots \geq 0$. Then show that the series $\Sigma\, a_n$ converges if and only if the series $\sum_{k=0}^{\infty} 2^k\, a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + \ldots$ converges.

Hence prove that $\Sigma \frac{1}{n^p}$ converges if $p > 1$ and diverges if $p \leq 1$.

b) State and prove Taylor's series expansion.

c) Show that arbitrary intersection of closed sets is closed.

Q4) Attempt any one of the following: **[1×10=10]**

a) i) If $s_n = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{n!}$, for $n = 1, 2, 3, \ldots.$ Prove that $2 < \lim_{n \to \infty} s_n < 3$. **[7]**

ii) Show that [0, 1] is compact set. **[3]**

b) i) Show that every neighborhood set open set. **[3]**

ii) Prove that $\lim_{n \to \infty} \frac{1}{\sqrt{n!}} = 0$. **[5]**

iii) Define convex and concave function. **[2]**

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PD3697

[Total No. of Pages : 3

[6487]-102

First Year M.Sc.

STATISTICS

STS-502-MJ : Linear Algebra

(2023 & 2024 Credit 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 Row Space and Column Space of a Matrix.

b) If A and B are invertible matrices of the same order, then show that AB is invertible.

c) Represent the quadratic form $4x^2 - 6xy + 3y^2$ in matrix form.

d) Check whether the vectors (2,1,3) and (–1,6,5) are linearly independent.

e) State any two properties of inverse of a matrix.

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

a) If $v_1 = v_2 = \mathbb{R}^2$ and $f : v_1 \rightarrow v_2,\ f(x, y) = (ax + by, cx + dy)$ check whether f is linear transformation.

b) Define similar matrices. Show that similar matrices have same eigenvalues.

c) State Cayley Hamilton theorem. Hence verify Cayley Hamilton theorem for a matrix

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 4 \end{bmatrix}$$

d) Prove or disprove:

i) Any subset of linearly independent vectors is linearly independent.

ii) Any superset of linearly dependent vectors is linearly dependent.

P.T.O.

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

a) Check whether a matrix A is diagonalizable, where $A = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{bmatrix}$

b) If $\vec{u}$ and $\vec{v}$ are two vectors in an inner product space, then show that

i) $<\vec{u},\vec{v}>=0$ if and only if $\|\vec{u}+\vec{v}\|=\|\vec{u}-\vec{v}\|$.

ii) $<\vec{u}+\vec{v},\vec{u}-\vec{v} \geq 0$ if and only if $\|\vec{u}\|=\|\vec{v}\|$.

c) Define basis of a vector space. Check whether the set B={(3,1,–4), (2,5,6), (1,4, 8)} forms a basis for $\mathbb{R}^3$.

d) Define idempotent matrix. Show that for an idempotent matrix A,

$$rank\ (A) = trace\ (A)$$

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

a) Let $V = \mathbb{R}^3$, the vector space of ordered triplets of real numbers. If $W = \{(x, y, z) \in \mathbb{R}^3 2x + 3y + 4z = 0\}$, then show that W is a subspace of V.

b) If $\vec{v}_1, \vec{v}_2, \ldots.\vec{v}_r$ are eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots\ldots\lambda_r$ then show that $\{\vec{v}_1, \vec{v}_2, \ldots.\vec{v}_r\}$ is linearly independent.

c) Consider the system of equation AX = 0, where A is $m \times n$ matrix with *rank*(A) = r. Show that there are $n - r$ linearly independent solutions to this system.

d) Define determinant of matrix. Consider a matrix $A = (a_{ij})$, where

$$a_{ij} = \begin{matrix} a & if & i=j=1,\ldots\ldots\ldots n \\ b & if & i\neq j=1,\ldots\ldots\ldots n \end{matrix}$$

then show that $\det(A) = [a + (n - 1)b](a - b)^{n-1}$

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

a) i) Solve the following system of linear equation by using LU decomposition.

$$-5x_1 + 4x_2 + x_4 = -17$$

$$-30x_1 + 27x_2 + 2x_3 + 7x_4 = -102$$

$$5x_1 + 2x_2 + 2x_4 = -7$$

$$10x_1 + x_2 - 2x_3 + x_4 = -6$$

ii) Define trace of a matrix of order n. Let A and B be two square matrices. Show that trace (A+B) = trace (A) + trace (B). **[10+5]**

b) i) Find spectral value decomposition of a matrix $A = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}$.

ii) State and prove the necessary and sufficient condition for a quadratic form to be non negative definite. **[10+5]**

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PD3698

[Total No. of Pages : 3

[6487]-103
First Year M.Sc.
STATISTICS
STS-503-MJ : Probability Distributions
(2024 Credit 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. **[5×2=10]**

a) Suppose the random variable X takes non-negative integer values, i.e. X is a count random variable. Prove that $E(X)=\sum_{n=0}^{\infty}P(X>n)$.

b) Let X and Y be two independent random variables with probability generating function $P_X(t)$ and $P_Y(t)$ respectively. Prove that, $P_{X+Y}(t) = P_X(t)P_Y(t)$.

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

d) State characteristic properties of univariate distribution function.

e) Define monotone field with an example.

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

a) Give an example of random variable which has exactly finite moment upto order 5 but for which the higher order moments do not exist. Justify your answer.

b) Let X~Poisson(λ) distribution. Show that P(P(s)) is probability generating function of random variable Y, where P(s) is probability generating function of random variable X. Obtain P(Y = 0) and E(Y).

c) Let $F(\bullet)$ be the distribution function (DF) of RV X where

$$F(\bullet)=\begin{cases}0 & \text{if } x<0\\ \dfrac{x}{4} & \text{if } 0\le x<1.\\ 1 & \text{if } x\ge 1\end{cases}$$ Obtain decomposition of DF F(•)

d) Let $\Omega=\{a,b,c,d\}, A=\{\phi,\Omega,\{a,b\},\{c,d\}\}$,

$X(a)=X(b)=-1, X(c)=1, X(d)=2.$ Examine whether X is A measurable.

P.T.O.

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

a) Let Q = $X'AX$ is a quadratic form, where $A_{m\times n}$ is a symmetric matrix then Q~χ^2 distribution with r degrees of freedom iff A is an idempotent matrix with rank r.

b) Let X and Y be two independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and g_1 and g_2 are Borel functions, then $g_1(X)$ and $g_2(Y)$ are also independent random variables.

c) Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed U(0,1) random variables. Obtain the probability density function of sample range $R = X_{(n)} - X_{(1)}$.

d) Let X and Y have joint pmf: $P(X = x, Y = y) = \frac{1}{3x}; x = 1,2,3$ and y = 1, 2, ... , x. Find the value of E(Y|X = 3].

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

a) Let (X, Y) follows bivariate beta distribution with parameters λ_1, λ_2 and λ_3. Obtain conditional distribution of X|Y.

b) Define Marshall-Olkin's exponential distribution. Also, identify the marginal distribution of variables involved in it.

c) Let (X,Y) be the random variable having joint pdf as $f(x,y) = xe^{-x(y+1)}; x \ge 0$ and y ≥ 0. Obtain the regression curve of y on x.

d) 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.

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

a) i) State and Prove Fisher-Cochran Theorem. **[6]**

ii) Let X and Y be random variables with common probability density function $f(x)=\begin{cases}\alpha e^{-x}, & x>0,\alpha>0\\ 0, & \text{Otherwise}\end{cases}$. Find the probability distribution of X + Y using convolution method. **[4]**

iii) Derive the probability density function of non-central F-distribution. Also, express the density function with non-centrality parameter $\lambda = 0$. **[5]**

b) i) 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]**

ii) An insurance company models the total claim amount *S* over a certain period using a compound distribution. Let N be the number of claims, modeled by a Poisson distribution with mean $\lambda = 5$. The individual claim amounts X_i are independent and follow an Exponential distribution with mean $\mu = 2000$. Derive the probability distribution of the total claim amount *S*. Also find its mean and variance. **[7]**

✤✤✤✤

Total No. of Questions : 4]

SEAT No. :

PD3699

[Total No. of Pages : 2

[6487]-104
M.Sc. - I
STATISTICS
STS-510-MJ : Optimization Techniques
(2024 Credit Pattern) (Semester - I)

Time : 2 Hours] *[Max. Marks : 35*

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 : **[5×1=5]**

a) Write any one property of Gomory's algorithm.

b) Define spanning tree in network routine problem.

c) What is the meaning of the term augmenting path.

d) Explain the supply node, transshipment nodes for solid waste management.

e) Write the necessary condition for minimum cost flow problem to have any feasible solution.

Q2) Attempt any two of the following : **[2×5=10]**

a) Explain the way of solving linear programming problem using the dynamic programming approach.

b) Write the steps of Gomory's Mixed-Integer Programming Algorithm.

c) Describe how to express transhipment problem as a networking problem.

Q3) Attempt any two of the following : **[2×5=10]**

a) Explain the procedure of Wolfe's modified simplex method.

b) Use dynamic programming problem to solve the following problem:

Min $Z = y_1^2 + y_2^2 + y_3^2$

Subject to constraints

$y_1 + y_2 + y_3 \geq 15$

$y_1, y_2, y_3 \geq 0$

c) Write the key types of minimum spanning tree problem.

P.T.O.

Q4) Attempt any one of the following : **[1×10=10]**

a) i) Describe the minimum cost flow problem.

ii) A company rent a car. This company is developing a replacement policy for its car fleet over a 4-year planning horizon. At the start of each year, a car is either replaced or kept in operation for an extra year. A car must be in service from 1 to 3. Formulate the model using shortest route method and find the replacement cost by using following information:

Equipment acquired at start of the year	Replacement cost for given in operations		
	1	2	3
1	4000	5400	9800
2	4300	6200	8700
3	4800	7100	-
4	4900	-	-

b) i) Explain the way of solving linear programming problem using the dynamics programming approach.

ii) The non-linear programming problem is

Max $Z = x_1 + x_2$

Suject to the constraints

$x_1 x_2 - 2x_2 \geq 3$

$3x_1 + 2x_2 \leq 24$

$x_1, x_2 \geq 0$

Draw the graph for the above problem and find the solution.

☯☯☯☯

Total No. of Questions : 4]

SEAT No. :

PD3700

[Total No. of Pages : 2

[6487]-105
M.Sc. - I
STATISTICS
STS 511-MJ : Statistical Quality Control
(2024 Credit Pattern) (Semester - I) (2 Credits)

Time : 2 Hours] ***[Max. Marks : 35***

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. **[5×1=5]**

a) Define the precision-to-tolerance ratio.

b) Write the control limits for EWMA chart for monitoring process mean.

c) A process is in control with $\overline{\overline{X}}$ = 100 and $\overline{s}$ = 16. Compute the estimates of mean and standard deviation. (C_4 = 0.94)

d) State the relationship between C_p and C_{pm}.

e) Define the term Average Run Length.

Q2) Attempt any Two of the following: **[2×5=10]**

a) Explain the need of process capability analysis.

b) Write a short note on Duncan's model for economic designing of control charts.

c) Obtain the 100(1– α)% confidence interval for capability index C_p and give the testing procedure of it.

P.T.O.

Q3) Attempt any Two of the following: **[2×5=10]**

a) A process is in control with $\overline{\overline{X}}=120$ and $\overline{R}=4$, n = 5. The process specifications are 115 ± 10. Compute C_p, C_{pk} and C_{pm}. ($d_2 = 2.326$)

b) Explain the working of tabular CUSUM chart for monitoring the process mean, when standard deviation is known.

c) Define the term conforming run length (CRL). Explain the construction and working of CRL charts.

Q4) Attempt any One of the following. **[1×10=10]**

a) Write a short note on; **[10]**

i) Chain sampling plan.

ii) Double sampling plan.

b) i) Explain the construction and working of Hotelling T^2 control chart for process mean vector when dispersion matrix is known and unknown.

ii) Short note on group run chart. **[6+4]**

❋ ❋ ❋

Total No. of Questions : 4]

SEAT No. :

PD3701

[Total No. of Pages : 2

[6487]-106

M.Sc. - I

STATISTICS

STS-512-MJ: Actuarial Statistics

(2024 Credit Pattern) (Semester - I) (2 Credits)

Time : 2 Hours] *[Max. Marks : 35*

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. **[5×1=5]**

a) Obtain probability distribution function of future life time random variable.

b) For the uniform distribution over (0, 80), find the force of mortality.

c) State the different assumptions of fractional ages.

d) Suppose force of interest per annum is 0.06. Find $i^{(2)}$.

e) Find the present value and accumulated value of a 15-year annuity immediate of Rs. 1500/- per annum if the effective rate of interest is 4%.

Q2) Attempt any 2 questions out of 3 questions. **[2×5=10]**

a) Show that $A_x = vq_x + vp_x A_{x+1}$.

b) It is given that $A_{x+20} = 0.40$, $A_x = 0.25$ $A_{x:\overline{20|}} = 0.55$ and $i = 0.03$ Calculate $1000\ \bar{A}_{x:\overline{20|}}$ under the assumption of uniformity in each unit age interval.

c) Under UDD assumption show that $\bar{A}_x = \frac{i}{\delta} A_x$.

Q3) Attempt any 2 questions out of 3 questions. **[2×5=10]**

a) Define the following terms.

   i) Loss at Issue Random Variable.

   ii) Equivalence Principle of premium.

   iii) Percentile premium.

b) Define annuity. Also, derive the expression of actuarial present value for n-year temporary life annuity and whole life annuity.

c) A select and ultimate life table gives $l_x = 100 - x$ and $l_{[x]+t} = 4\,(0.3 - 0.01t)\,l_{x+t}$. Compute the probability that a life now aged 22, selected 1 year ago will survive to age 26.

P.T.O.

Q4) Attempt any 1 question out of 2 questions. **[1×10=10]**

a) i) Write a note on retrospective reserve and prospective reserves. **[4]**

ii) Prove that the retrospective and prospective reserves are equal at time t for an immediate annuity (payable annually in arrears) of amount B with initial expenses I and actual renewal expenses R. **[6]**

b) i) Assume that premiums are calculated on the basis of the equivalence principle. If ${}_kL$ is the prospective loss random variable for a fully discrete whole life insurance of 1000 issued to (x). It is given that $A_x = 0.125$, $A_{x+k} = 0.4$, ${}^2A_{x+k} = 0.2$, $d = 0.05$. Then calculate $E\,({}_kL)$, $Var\,({}_kL)$ and the aggregate reserve at time k for 100 policies of this type. **[7]**

ii) Calculate ${}_1V_{40}$ given that $P_{40} = 0.01536$, $p_{40} = 0.99647$ and $i = 0.05$. **[3]**

ⓘⓘⓘⓘ

Total No. of Questions : 5]

SEAT No. :

PD3702

[Total No. of Pages : 2

[6487]-107

First Year M.Sc.

STATISTICS

STS-541-RM-MJ : Research Methodology

(2023 & 2024 Credit 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) What are mixed and multiplicative generators?

b) Prove that the random variable X generated by the rejection method has density function f.

c) Negate the statement. If x is prime then $\sqrt{x}$ is not a rational number.

d) What are mixed and multiplicative generators?

e) Explain convolution method briefly to generate random sample.

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

a) Write a short note on design of the research project.

b) Describe the stages involved in the research process. How do they relate to each other?

c) What is the purpose of a literature review in research? How do you conduct a literature review?

d) What are the different types of research designs? Explain the strengths and limitations of each.

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

a) Let A and B be subsets of some universal set. If $A = B \cup \{x\}$ where $x \notin B$, then any subset of A is either a subset of B or a set of the form $C \cup \{x\}$, where C is a subset of B.

b) Give a method for simulating a Gamma random variable.

c) If n is small, how we can simulate a random sample from $(X) = \sum_{i=1}^{n} P_i F_i(x), P_i \geq 0, \sum_{i=1}^{n} P_i = 1$?

d) Let A and B be subsets of some universal set. Then show that $A-(A-B) = A \cap B$.

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

a) Consider the relation $R = \{(x, x) : x \in \mathbb{Z}\} \text{ on } \mathbb{Z}$? Check whether is R reflexive, symmetric or transitive?

b) Write a R-code to simmulate a random sample from the exponential distribution with parameter $1/\lambda$ using inverse transform method.

c) What is Power BI? Explain how it is useful for data visualization.

d) What is ggplot? Explain how it is useful for data visualization.

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

a) i) Give a method for simulating a random sample from

$$F(X) = \begin{cases} \dfrac{1-e^{-2x}+2x}{3}, & 0 < x < 1 \\ \dfrac{3-e^{-2x}}{3} & 1 < x < \infty \end{cases}$$ **[8]**

ii) For statements P and Q.

a) Show that, the conditional statement $P \rightarrow Q$ is logically equivalent to $\sim P \vee Q$.

b) The statement $\sim(P \rightarrow Q)$ is logically equivalent to $P \wedge \sim Q$.

[7]

b) i) Define Cartesian product with an example. Also, sketch Cartesian product $\{1, 2, 3\} \times \{-1, 0, 1\}$ on the xy-plane. **[7]**

ii) Let A, B and C be sets. Then show that : **[8]**

$$A \times (B \cup C) = (A \times B) \cup (A \times C)$$

Total No. of Questions : 4]

SEAT No. :

PD3703

[Total No. of Pages : 2

[6487]-201

First Year M.Sc.

STATISTICS

STS-551-MJ : Modern Statistical Inference

(2024 Credit Pattern) (Semester-II) (2 Credits)

Time : 2 Hours] *[Max. Marks : 35*

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. **[5×1=5]**

a) Define Ancillary statistics with an illustration.

b) Prove or disprove: If there exist two unbiased estimators of θ, then there exist infinitely many unbiased estimators of θ.

c) Define term simple and composite hypothesis with an illustration.

d) Define Type-I and Type-II error.

e) In a study with 50 hypotheses, you rejected 12 hypotheses. Out of the 12 rejected hypotheses, 3 were false positives (Type I errors). Calculate the False Discovery Rate (FDR) for this study.

Q2) Attempt any Two of the following: **[2×5=10]**

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

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

c) Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with parameter θ. Show that $\sum_{i=1}^{n} X_i$ is sufficient for θ.

P.T.O.

Q3) Attempt any Two of the following: **[2×5=10]**

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

b) Let $X_1, X_2, \ldots, X_n$ be a random sample from

$$P(X = x) = \begin{cases} \theta(1-\theta)^x & ,\theta > 0, x = 0,1,.... \\ 0 & ,\text{otherwise.} \end{cases}$$

Show that $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is UMVUE for $\frac{1-\theta}{\theta}$. Examine whether it attains CRLB.

c) 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	0.10	0.10	0.20	0.18	0.30

Find the MP test of size 0.1.

Q4) Attempt any One of the following: **[1×10=10]**

a) i) State Neyman-Pearson Lemma. **[3]**

ii) Let $x_1 = 6, x_2 = 7, x_3 = 5, x_4 = 8, x_5 = 4, x_6 = 6$ be a random sample of size 6 from $N(\mu, \sigma^2 = 4)$. Find the likelihood ratio test at the $\alpha = 0.05$ significance level. **[7]**

b) i) Why is controlling the False Discovery Rate important in multiple hypothesis testing? **[3]**

ii) If X the number of trials needed to produce the first k successes in a sequence of Bernoulli trials with the probability of success in a single trial being θ. Find the MLE of θ and the asymptotic variance of MLE. **[7]**

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PD3704

[Total No. of Pages : 2

[6487]-202

First Year M.Sc.

STATISTICS

STS-552-MJ : Regression Analysis and Applications

(2023 & 2024 Credit Pattern) (Semester -II) (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 scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

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

a) Show that solution of normal equation actually minimizes the residual sum of squares.

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

c) Define PRESS residuals and give the procedure to calculate it.

d) Derive the relationship between the ANOVA F-test and R^2 in a multiple linear regression model.

e) How are the condition number and condition index used to detect multicollinearity in regression analysis?

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

a) Derive the likelihood ratio test for testing $H_0 : \underline{\beta} = \underline{\beta}_0$.

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) State and prove Gauss-Markoff theorem.

d) 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.

P.T.O.

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

a) Define spline and knots. Also, explain the piecewise polynomial model fitting using cubic spline.

b) For the simple linear regression model estimate the parameters β_0, β_1 and σ^2 by using the method of maximum likelihood estimation.

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

d) Explain the use of PRESS statistic and Cook's D statistics.

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

a) Obtain the parameters in logistic regression model using the method of maximum likelihood estimation.

b) Write a short note on Poisson regression.

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

d) For the generalized linear model, explain the following link function.

   i) Binomial function

   ii) Inverse function

   iii) Inverse binomial

Q5) Attempt any one of the following. **[15 Each]**

a) Explain the process of Least Squares Estimation (LSE) in non-linear regression model $y = \theta_1 e^{\theta_2 + \theta_3 x} + \varepsilon$. How is it different from linear regression LSE? **[15]**

b) i) Derive the canonical link function for the Gamma distribution in a GLM setting. **[8]**

   ii) Define an estimable linear parametric function and state the necessary and sufficient condition for estimability. **[7]**

Total No. of Questions : 5]

SEAT No. :

PD3705

[Total No. of Pages : 3

[6487]-203
First Year M.Sc.
STATISTICS
STS-553-MJ : Multivariate Analysis and Applications
(2023 & 2024 Pattern) (Credit System) (Semester - II)

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

Instructions to the candidates:

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

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

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

b) 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.

c) Determine the first principal component for the covariance matrix given below : $\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$

d) State the additive property of Wishart distribution.

e) Write note on Fisher discriminant function.

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

a) Let $\underline{X} \sim N_p\left(\underline{\mu}, \Sigma\right)$. If A be any qXp matrix then prove that $A\underline{X} \sim N_q\left(A\underline{\mu}, A\Sigma A'\right)$. If $\underline{X} \sim N_3\left(\underline{\mu}, \Sigma\right)$ where $\underline{\mu} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$

Obtian the distribution of $\begin{bmatrix} X_1 - X_2 \\ X_1 + 2X_2 - X_3 \end{bmatrix}$.

P.T.O.

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

c) Explain Two Way MANOVA.

d) Show that the principal components are uncorrelated and have variances equal to the eigen values of Σ.

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

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

b) Let $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ be a random sample of size n from $N_p(\underline{\mu}, \Sigma)$, where Σ is unknown. Obtain the likelihood ratio test for testing $H_0 : \underline{\mu} = \underline{\mu}_0$. Establish its relationship with Hotelling-T^2 statistic.

c) Define Hierarchical cluster analysis. Also gives it's characteristics.

d) Consider the following density function :

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

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

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

a) Let $\underline{X} \sim N_4(\underline{\mu}, \Sigma)$ with $\underline{\mu} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 2 & 1 & 1 & 1 \\ & 2 & 1 & 1 \\ & & 2 & 1 \\ & & & 3 \end{bmatrix}$, then find the conditional distribution of $\begin{bmatrix} X_3 \\ X_4 \end{bmatrix} / \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

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

c) Derive the region of classification to classify an observation into one of the two Normal populations when $\Sigma_1 \neq \Sigma_2$.

d) Write a note on multiple correlation coefficient.

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

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

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

2) Find the first pair of canonical variables.

[8]

ii) Derive the characteristic function of Wishart distribution. **[7]**

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

ii) Let X̲ of order PX1 has mean μ̲ and covariance matrix Σ. Obtain K-principal components of standardized vector Z̲ of X̲. **[7]**

✪ ✪ ✪

Total No. of Questions : 4]

SEAT No. :

PD3706

[Total No. of Pages : 2

[6487]-204

M.Sc. - I

STATISTICS

STS-560-MJ: Advances in Generalized Linear Models

(2024 Credit Pattern) (Semester - II)

Time : 2 Hours] ***[Max. Marks : 35***

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. **[1 Each]**

a) What is deviance in GLM?

b) Justify the use of the log link function in Poisson GLMs.

c) Define quasi-likelihood and provide an example of its application in statistical modeling.

d) Define baseline-category logits.

e) Explain the exponential dispersion family.

Q2) Attempt any two of the following. **[5 Each]**

a) Explain the key difference between a classical Generalized Linear Model (GLM) and a Bayesian GLM.

b) Explain the concept of Generalized Estimating Equations.

c) Explain the concept and structure of a zero-inflated model (ZIP or ZINB).

Q3) Attempt any two of the following. **[5 Each]**

a) Give the general frameworks of the

i) Likelihood Ratio test

ii) Wald Test

b) What are the differences between marginal models and GLMMs?

c) Write a short note on the Poisson Regression Model by stating its assumption, need and use. Also, provide the tests for determining statistical significance of regression coefficients.

P.T.O.

Q4) Attempt any one of the following. **[10]**

a) A study is conducted to determine whether smoking status predicts the presence of a disease which is given in the following data.

Subject	1	2	3	4	5	6
Smoking	Yes	No	Yes	No	Yes	No
Disease	Yes	No	Yes	No	No	No

Using Newton-Raphson method, compute MLE for β_0 and β_1 (solve upto one iteration). Also, compute the odds ratio and interpret it.

b) Explain Poisson regression model and state any two applications.

ⓘⓘⓘⓘ

Total No. of Questions : 4]

SEAT No. :

PD3707

[Total No. of Pages : 1

[6487]-205
M.Sc. - I
STATISTICS
STS-561-MJ : Statistical Methods in Epidemiology
(2024 Credit Pattern) (Semester - II) (2 Credits)

Time : 2 Hours] *[Max. Marks : 35*

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. **[5×1=5]**

a) Define fatality rate and infection rate.

b) What is meant by "chains of infection"?

c) Write two key features of descriptive epidemiology.

d) State the difference between basic and preventive reproduction number.

e) What are asymptomatic carriers? Give one example.

Q2) Attempt any TWO of the following: **[2×5=10]**

a) Describe measures of disease frequency with examples.

b) Explain the concept and use of spatial exposure assessment.

c) What are the limitations of the Reed-Frost model?

Q3) Attempt any TWO of the following: **[2×5=10]**

a) Describe how SEIR models improve upon SIR models.

b) What is the significance of latent period in epidemic modeling?

c) Explain the role of doubling time in monitoring an epidemic.

Q4) Attempt any ONE of the following: **[1×10=10]**

a) i) Discuss the application of Bayesian inference in modeling the spread of COVID-19. **[7]**

ii) What do you mean by 'confounding' in epidemiology? **[3]**

b) i) Explain the SIR and SEIR models with reference to the 2009 H1N1 outbreak. **[7]**

ii) What is spatial clustering in epidemiology? **[3]**

✪ ✪ ✪

Total No. of Questions : 4]

SEAT No. :

PD3708

[Total No. of Pages : 2

[6487]-206
First Year M.Sc.
STATISTICS
STS-562-MJ : Discrete Data Analysis Linear Models
(2024 Credit Pattern) (Semester - II) (2 Credits)

Time : 2 Hours] *[Max. Marks : 35*

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 : **[1 each]**

a) State any two applications of poisson regression model.

b) In a survey, 3 brands of a product are preferred with probabilities 0.4, 0.35, and 0.25. If 10 people are surveyed, find the probability that exactly 4 prefer Brand A, 3 prefer Brand B, and 3 prefer Brand C.

c) What are clustered binary data?.

d) Explain the comprehensive log linear model.

e) Write the likelihood function for a sample of size n from a Binomial distribution with parameters n and p.

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

a) In Poisson regression, what is the role of the log-link function? Why is it commonly used?

b) Explain the use of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in model selection.

c) Explain the concept of Generalized Estimating Equations.

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

a) Explain likelihood ratio test and construct the confidence interval using it.

b) Construct the confidence interval for Binomial parameter using Score test.

c) Explain how linear trend is alternative to independence.

Q4) Attempt any one of the following : **[10 each]**

a) Compare the marginal modeling with Generalized Estimating Equations (GEE) in the context of analyzing correlated or clustered data and discuss their respective strengths, limitations, and applications.

b) i) Explain Simpson's Paradox phenomenon.

ii) Explain the concept of generalized linear model for cluster responses.

☯☯☯☯

Total No. of Questions : 4]

SEAT No. :

PD3709

[Total No. of Pages : 2

[6487]-301

M.Sc. - II

STATISTICS

STS-601-MJ : Probability Theory

(2024 Credit Pattern) (Semester-III) (2 Credits)

Time : 2 Hours] *[Max. Marks : 35*

Instructions to the candidates:

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

Q1) Attempt all questions. **[1×5=5]**

a) Define convergence in distribution.

b) State Weak Law of large numbers.

c) Define independents of events.

d) Define probability Space of a random vector.

e) Define arbitrary random variable.

Q2) Attempt any two of the following: **[2×5=10]**

a) If $x_n \xrightarrow{P} x,\ y_n \xrightarrow{P} y,$ then show that $ax_n + by_n \xrightarrow{P} ax, + by$.

b) State and prove Jordan decomposition theorem.

c) If $\{X_n\}_{n\geq 1}$ is a sequence of random variables then prove that $X_n \to 0$ in probability if and only if $E\left(\frac{|X_n|}{1+|X_n|}\right) \to 0$ in probability as $n \to \infty$

P.T.O.

Q3) Attempt any two of the following: **[2×5=10]**

a) Suppose $\{y_n, n \geq 1\}$ are i.i.d. with U(0,1) and $X_{(n)} = \min\{y_1, y_2, ..., y_n\}$ then show that $nX_{(n)} \to Y$ in distribution.

b) Show that $E|XY| \leq E^{\frac{1}{r}}|X|^r \,.\, E^{\frac{1}{s}}|Y|^S$ where $r > 1$ and $\frac{1}{r} + \frac{1}{s} = 1$

c) State and prove Borel 0-1 law.

Q4) Attempt any one of the following: **[1×10=10]**

a) Suppose X and Y are simple random variables and E(X) and E(Y) are exist then show that

i) $X \geq 0$ almost surely $\Rightarrow E(X) \geq 0$ (Non-negativity)

ii) $E(cX) = cE(X),\ c \in R$

iii) $X = 0$ almost surely $\Rightarrow E(X) = 0$

iv) $X \geq Y$ almost surely $\Rightarrow E(X) \geq E(Y)$

b) i) If $Y \leq X_n$, Y is integrable then $E\,\underline{\lim}\,X_n \leq \underline{\lim}\,EX_n$

ii) If $X_n \leq Z$, Z is integrable then $\overline{\lim}\ EX_n \leq E\overline{\lim}\ X_n$

iii) $Y \leq X_n \leq Z$, Y and Z are integrable and $X_n \xrightarrow{a.s.} X$ then $\lim_{n\to\infty} EX_n = EX$

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PD3710

[Total No. of Pages : 2

[6487]-302

M.Sc. - II

STATISTICS

STS 602 MJ : Stochastic Processes

(2023 & 2024 Credit 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) Write any two applications of stochastic processes.

b) Define a Brownian motion process.

c) State the additive property of Poisson distribution.

d) Explain extinction probabilities.

e) When a state of a Markov chain is called recurrent and transient?

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

a) Explain the postulates of Yule-Furry process and find an expression for $P_n(t)$.

b) Show that every stochastic process with independent increments is a Markov process.

c) Derive the relation between Poisson Process and Binomial Distribution.

d) Suppose that probability of dry day after rainy day is $\frac{3}{4}$ and that of rainy day after dry day is $\frac{2}{3}$ Let X_n be state of process after n^{th} day write state space and one step TPM of process $\{X_n, n \geq 1\}$ also find probability that second day is dry given that initial day is dry.

P.T.O.

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

a) Write short notes on the following

i) Stationary Distribution

ii) Communicative sets and their equivalence property

b) Let $\{X_n, n = 0.1.2,...\}$ be a Branching process with $X_0 = 1$. Find the mean and variance of X_n, in terms of those of the offspring distribution.

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

d) In a Poisson process with a rate of 4 arrivals per minute, what is the probability that there will be exactly 3 arrivals in a 30-second interval?

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

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

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

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

d) Derive the generating function relations satisfied by a Branching process

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

a) i) A Markov chain on states {1, 2 , 3, 4} has transition probability matrix

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

Classify the states are as recurrence or transient. **[8]**

ii) Explain the concept of inter-arrival time in a Poisson process. How can you compute the probability density function of inter-arrival times? **[7]**

b) i) Find the expected duration of game in the gambler ruin problem. **[7]**

ii) Derive forward and backward Kolmogorov differential equations for a Birth and Death process. **[8]**

✪ ✪ ✪

Total No. of Questions : 5]

SEAT No. :

PD3711

[Total No. of Pages : 3

[6487]-303

M.Sc. - II

STATISTICS

STS-603-MJ : Design and Analysis of Experiments

(2023 & 2024 Credit 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 each of the following: **[5×2=10]**

a) Define the following:
   i) Connected design
   ii) Orthogonal design

b) Obtain parameters of the following PBIBD:

Blocks	Treatments	
1	1	2
2	3	4
3	4	1
4	2	3

c) Define Confounding. Explain types of confounding.

d) Explain the linear and quadratic effect of factor.

e) Explain desirability function.

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

a) Construct 2_{IV}^{8-3} design.

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

c) Explain following method for comparing pairs of treatment means:
   i) Duncan's Multiple Range Test
   ii) Newman Keuls Test

d) In BIBD prove that trace of C-matrix is $b(k-1)$.

P.T.O.

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

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

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

c) Explain 3^3 design with illustration.

d) Write a note on Signal-to-Noise ratio.

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

a) Give the statistical analysis of a 2^3 factorial experiment in which 4 replicates are used and BC is totally confounded.

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) Confound AB^2C^2 in 3^3 factorial experiment in 3 blocks.

d) What is Simplex Centroid Design. Explain Simplex Centroid Design with p=3 components.

Q5) Attempt any one of the following:

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

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

ii) Explain Central Composite Design (CCD). Differentiate between Circumscribed(CCC), Face Centered(CCF) and Inscribed(CCI) designs of CCD for number of factors k=2. **[7]**

b) i) 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: priexercise, 3-month and post exercise. The data shown below: **[8]**

Subjects	Blood Pressure		
	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.

ii) Show that RBD is connected, orthogonal and variance balanced design. **[7]**

✤✤✤✤

Total No. of Questions : 4]

SEAT No. :

PD3712

[Total No. of Pages : 2

[6487]-304
M.Sc. - II
STATISTICS
STS-610-MJ : Survival Analysis
(2023 & 2024 Credit Pattern) (Semester - III) (2 Credits)

Time : 2 Hours] *[Max. Marks : 35*

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) Choose the correct alternative to each of the following questions : **[5×1=5]**

a) Let a component has failure rate r(t) = 8, then cumulative hazard rate of the component is;

i) $8t$ ii) $\frac{t}{8}$

iii) $4t$ iv) $8t^2$

b) The "New Better than Used" property is described as;

i) $\overline{F}(x).\overline{F}(t) < \overline{F}(x+t), \forall x,t > 0$ ii) $\overline{F}(x).\overline{F}(t) \geq \overline{F}(x+t), \forall x,t > 0$

iii) $\overline{F}(x)+\overline{F}(t) > \overline{F}(x+t), \forall x,t > 0$ iv) $\overline{F}(x)+\overline{F}(t) < \overline{F}(x+t), \forall x,t > 0$

c) In an experiment with Type II censoring scheme, the random variable of interest is the ______.

i) Termination time of the experiment
ii) No. of units failed before the termination time
iii) No. of units involved in the experiment
iv) Life time of the units at the experiment.

d) A lifetime distribution belongs to DFRA class iff,

i) Hazard rate is increasing function of time 't'
ii) Hazard rate is decreasing function of time 't'
iii) Hazard rate average function is decreasing function of time 't'
iv) Hazard rate average function is increasing function of time 't'

e) A non-parametric test for two sample problem given below is:

i) t-test ii) Chi-square test

iii) Mann-Whitney U test iv) Run test

P.T.O.

Q2) Attempt any 2 questions out of 3 questions : **[2×5=10]**

a) Prove the following implications:

IFR $\Rightarrow$ IFRA and IFR $\Rightarrow$ DMRL

b) Describe the Cox's Proportional Hazards model of regression for complete data.

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

Q3) Attempt any 2 questions out of 3 questions : **[2×5=10]**

a) Suppose 20 items from an exponential distribution are put on life test and observed for 150 hours. During the period 15 items fail with the following life times. 3, 19, 23, 26, 27, 37, 38, 41, 45, 58, 84, 90, 99, 109, 138. Identify the type of censoring used in the experiment. Also find M.L.E. of the parameter and M.L.E. of average lifetimes of the items.

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

c) Obtain hazard rate for the Lehman family of life distribution.

Q4) Attempt any 1 question out of 2 questions : **[1×10=10]**

a) i) State and prove Cauchy functional equation.

ii) The following failure and censor times (in operating hrs.) were recorded on 12 turbine vanes : 142, 149, 320, 345+, 560, 805, 1130+, 1720, 2480+, 4210+, 5280, 6890. (+ indicates censored observation). Censoring was a result of failure mode other than wear out. Compute the Kaplan-Meier estimator of the survival function.

[5+5]

b) i) Derive Greenwood's formula for variance of actuarial estimator of the survival function.

ii) Following table shows the failure time of two machines, new and old.

	Failure times (day)
New machine	250, 476+, 355, 200, 355+
Old machine	191, 563, 242, 285, 16, 16 , 16, 257, 16

(+ indicates censored times)

Test whether the new machine is more reliable than the old one by using log rank test.

Total No. of Questions : 4]

SEAT No. :

PD3713

[Total No. of Pages : 2

[6487]-305

M.Sc. - II

STATISTICS

STS 611-MJ : Asymptotic Inference

(2023 & 2024 Credit Pattern) (Semester - III) (2 Credits)

Time : 2 Hours] ***[Max. Marks : 35***

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. **[5×1=5]**

a) Describe score test.

b) State invariance property of consistent estimator.

c) Define consistent estimator.

d) State any two properties of MLE.

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

Q2) Attempt any Two of the following: **[2×5=10]**

a) State and prove invariance property of CAN estimator.

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

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

P.T.O.

Q3) Attempt any Two of the following: **[2×5=10]**

a) Describe Delta method with on example.

b) Explain method of percentile to obtain CAN estimator.

c) Let $X_1, X_2, ...X_n$ is random sample from Exp (θ_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 $P(X < Y)$.

Q4) Attempt any One of the following. **[1×10=10]**

a) i) Describe in detail Likelihood Ratio Test for Multinomials Pearsons Chi-square test for goodness of fit. **[7]**

ii) Discuss the Wald test. **[3]**

b) i) Let $X_1, X_2,....., X_n$ is random sample from Poisson (θ). Check whether $T_n = \overline{X_n}$ is BAN estimator for θ. **[5]**

ii) 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. **[5]**

❋ ❋ ❋

Total No. of Questions : 4]

SEAT No. :

PD3714

[Total No. of Pages : 2

[6487]-306
M.Sc. - II
STATISTICS
STS-612-MJ: Machine Learning
(2023&2024 Credit Pattern) (Semester - III) (4 Credits)

Time : 2 Hours] ***[Max. Marks : 35***

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. **[5×1=5]**

a) What are the main types of Machine Learning?

b) Write common methods to handle missing data in Machine Learning?

c) What does the ROC curve represent in model evaluation?

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

e) Define Single Linkage in Agglomerative Hierarchical Clustering.

Q2) Attempt any Two of the following: **[2×5=10]**

a) What is Machine Learning (ML), and why is it important?

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

c) 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	2	0	5	9	8
C	6	5	0	4	5
D	10	9	4	0	3
E	9	8	5	3	0

P.T.O.

Q3) Attempt any Two of the following: **[2×5=10]**

a) What is the Apriori Algorithm, and how is it used for association rule mining?

b) How classifier performance is assessed using a confusion matrix, and what related measures can be derived from it?

c) For the following dataset of 8 points, with two features X_1 and X_2 . Classify the of new point ($X_1 = 4$, $X_2 = 4$) by using K- Nearest Neighbor algorithm with K = 3.

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

Q4) Attempt any One of the following: **[1×10=10]**

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

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

b) i) What are Artificial Neural Networks (ANNs), and how do they function in Machine Learning?

ii) A company wants to classify emails as either Spam or Not Spam based on certain features. The company uses two features: Contains "Offer" (whether the word "offer" appears in the email) and Contains "Click" (whether the word "click" appears in the email). For the following data from past emails. Classify a new email that contains the word "Offer" and does not contain the word "Click" using the Naive Bayes classifier.

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

ⓘ ⓘ ⓘ ⓘ

Total No. of Questions : 5]

SEAT No. :

PD3715

[Total No. of Pages : 2

[6487]-401
M.Sc. - II
STATISTICS
STS - 651 - MJ : Time Series Analysis
(2023 & 24 Credit Pattern) (Semester - IV)

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

Instructions to the candidates:

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

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

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

b) Write the SARIMA (0, 1, 1) X $(0, 1, 1)_{12}$ 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) Explain Dickey - Fuller test.

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

a) Write a note on Holt-Winters' smoothing technique

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

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

P.T.O.

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

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

b) Compute ψ_j & π_j coefficients for j = 1, 2,....,5 for the following processes :

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

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

c) Obtain ACVF of causal ARMA(1, 1) process.

b) Explain the Innovation Algorithm.

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

a) Explain Durbin - Levinson algorithm.

b) Write a note on ARCH and GARCH model.

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

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

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

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

Q5) Attempt any one of the following.

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

A) Exponential Smoothing 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	2	6	5	8

b) i) Estimate the parameters of stationery AR(p) process using Yule-Walker equations. **[8]**

ii) Determine the ψ_j and π_j coefficients of the ARMA (p, q) Process. **[7]**

✪ ✪ ✪

Total No. of Questions : 5]

SEAT No. :

PD3716

[Total No. of Pages : 2

[6487]-402

Second Year M.A./M.Sc.

STATISTICS

STS-652-MJ : Sampling Theory and Applications

(2023 & 2024 Credit Pattern) (Semester -IV)

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

Instructions to the candidates:

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

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

a) Define the first and second order inclusion probability.

b) Explain any two factors that are responsible for non-sampling error.

c) Define the sampling frame with an illustration.

d) Define the following with illustration.

i) Cluster sampling

ii) Stratified Sampling

e) Differentiate between the methods of SRS and varying probability scheme.

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

a) Explain Spatial Smoothing.

b) Show that ratio estimator of population mean is more precise than the regression estimator of the population mean.

c) Cost function for two phase sampling is C=C0+Cl m +C2 n. Then find minimum cost C for fixed mean squared error of the ratio estimator of population mean in case of two phase sampling.

d) Show that systematic sampling is more precise than simple random sampling without replacement, if the variance within systematic sample is larger than the population variance as a whole.

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

a) Give difference between Non response and Not at homes method.

b) Explain cluster sampling and discuss the effect of cluster size on relative efficiency.

c) Discuss Cumulative total method and Lahiri's method of selecting random sample by using PPSWR method.

d) Obtain Horvitz-Thompson (HT) estimator of population mean when Simple Random samplign without Replacement (SRSWOR) with varying probabilities of selection is used. Show that it is unbiased and derive expression for its variance.

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

a) Derive an unbiased ratio type estimator for population total. Also find variance of your estimator.

b) Explain the method of 'Post Stratification' giving an illustration, State the estimator of population mean in this method and verify whether it is unbiased estimator of population mean.

c) 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.

d) Explain the methods of drawing a sample by 'Circular systematic sampling' and 'Two dimensional systematic sampling', Compare them.

Q5) Attempt any one of the following question.

a) i) Define inclusion probabilities πi and πij for a sample of size n where i and j denote units in the population of N units. Obtain πi and πij for Simple Random Sampling without Replacement (SRSWOR) and Simple Random Sampling with Replacement (SRSWR). Derive the expression for variance of sample mean when SRSWOR is used. **[10]**

ii) Distinguish between sampling and non-sampling errors. **[5]**

b) i) The circular systematic sampling is usually preferred over linear systematic sampling. Discuss, why is it so? **[6]**

ii) Explain the following terms: Warner's model, Franklin's model, Jackknife technique. **[9]**

Total No. of Questions : 4]

SEAT No. :

PD3717

[Total No. of Pages : 2

[6487]-403

M.Sc. - II

STATISTICS

STS - 660-MJ : Advanced Statistical Learning Techniques and Applications

(2023 & 2024 Credit Pattern) (Semester - IV) (2 Credits)

Time : 2 Hours] ***[Max. Marks : 35***

Instructions to the candidates:

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

Q1) Attempt all of the following questions. **[5×1=5]**

a) What is the working principal of autoencoders and decoders?

b) Write down the difference between deep learning and machine learning.

c) State the salient features of Neural Network.

d) Give the basic elements of a Biological neuron.

e) What is baseline model in deep learning?

Q2) Attempt any Two of the following questions. **[2×5=10]**

a) Explain the Architecture of Convolution Neural Networks in detail?

b) How would you construct Alex Net layers and filters? Explain.

c) What is object detection and how is it different from image segmentation?

Q3) Attempt any Two of the following questions. **[2×5=10]**

a) Discuss the concept of backpropagation in neural networks. How is it used to calculate the error and update the weight and biases of the network during training?

b) What is image segmentation and how is it used in computer vision applications?

c) Explain the structure and learning process of single layer perceptron.

P.T.O.

Q4) Attempt any Two of the following questions. **[2×5=10]**

a) How are LSTM models used in video to text applications and what are some potential applications of this technology?

b) Explain the working principle of recurrent neural network.

c) Explain the process of gradient descent and its role in optimizing the weights and biases of neural network. How does it help in improving the accuracy of a deep learning model?

✪ ✪ ✪

Total No. of Questions : 4]

SEAT No. :

PD3718

[Total No. of Pages : 1

[6487]-404

M.Sc. - II

STATISTICS

STS-661-MJ: Design and Analysis of Clinical Trials

(2023 & 2024 Credit Pattern) (Semester - IV)

Time : 2 Hours] ***[Max. Marks : 35***

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. **[1 Each]**

a) Explain the term washout period.

b) What is mean by target population?

c) Describe cross-sectional design.

d) Define bias in the context of clinical studies.

e) Mention one use of sequential stopping in trials.

Q2) Attempt any two of the following. **[5 Each]**

a) Write a short note on Balaam's design.

b) In a clinical trial comparing two treatments, 40 patients are allocated to each group. The standard deviation is 10, and the true difference in means is 4. Find the power of the test at a 5% significance level.

c) How is sample size calculated for detecting treatment effects in Phase III trials?

Q3) Attempt any two of the following. **[5 Each]**

a) Differentiate between pharmacokinetics and pharmacodynamics.

b) What is a clinical trial protocol? Why is it important?

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

Q4) Attempt any one of the following. **[10 Each]**

a) Define bioavailability and pharmacokinetics. Describe the components of a two - compartment pharmacokinetic model and its application.

b) Describe the various types of bias in clinical trials and how randomization helps to minimize them. Also, explain any one method to detect outlying subjects.

ⓘⓘⓘⓘ

Total No. of Questions : 4]

SEAT No. :

PD3719

[Total No. of Pages : 2

[6487]-405

S.Y. M.Sc.

STATISTICS

STS - 662 - MJ : Bayesian Inference

(2023 & 2024 Credit Pattern) (Semester - IV) (2 Credits)

Time : 2 Hours] *[Max. Marks : 35*

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. **[1 Each]**

a) State the properties of large sample Bayes estimator.

b) Define Prior and Posterior distribution.

c) Give one application of Bayes theorem for performing hypothesis test.

d) State MCMC algorithm.

e) Represent the squared error loss function graphically.

Q2) Attempt any two of the following. **[5 Each]**

a) Write short note on informative prior and non-informative prior

b) Define zero-one (all-or-nothing) loss function and obtain the Bayes estimator under zero-one (all-or-nothing) loss function.

c) Suppose $X_1, X_2, \ldots, X_n$ 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.

P.T.O.

Q3) Attempt any two of the following. **[5 Each]**

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

b) Explain Bayes rule. A bag I contains 4 white and 6 black balls while other bag contains 4 white and 3 black balls. One ball is chosen at a random from one of the bags and it is found to be black. Find the probability that it was drawn from bag I. (Assuming bag is chosen randomly).

c) Discuss Gibbs sampling in briefly.

Q4) Attempt any one of the following. **[10 Each]**

a) Explain Bayesian information criterion (BIC) with illustartion.

b) 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.

✪ ✪ ✪

Total No. of Questions : 5]

SEAT No. :

PD3720

[Total No. of Pages : 3

[6487]-1001

First Year M.Sc.

STATISTICS

STS-501-MJ : Fundamentals of Analysis & Calculus

(2023 Credit 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 scientific calculator and statistical table is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Answer the following questions. **[2 Each]**

a) Show that set of rational number is countable set.

b) Define interior point of a set with an illustration.

c) Find limit inferior and limit superior of a sequence $t_n = \sin\left[\pi\left(n - \frac{1}{2}\right)\right]$.

d) Define continuity of a function from R^n to R.

e) State multiplication theorem for Jacobians.

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

a) Suppose s is an ordered set with least upper bound (lub) property. Suppose $B \subset S$ and B is nonempty set with bounded below. Let L be the set of all lower bounds of B then α = lub(L) exist in S and α = glb(B).

b) Prove that the following results:

i) If p > 0, then $\lim\limits_{n\to\infty} \frac{1}{n^p} = 0$.

ii) If p > 0, then $\lim\limits_{n\to\infty} \sqrt[n]{p} = 1$.

P.T.O.

c) State and prove mean value theorem for function of several variable.

d) Obtain the radius of convergence for the following series:

i) $$\sum_{n=1}^{\infty} \frac{(-1)^n n^2 x^n}{5^n \sqrt{n^5}}$$

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

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

a) If r is rational $(r \neq 0)$ and x is irrational, then prove that $r + x$ and rx are irrational. Hence prove that $\sqrt{12}$ is irrational number.

b) For any two real sequence $\{a_n\}$ and $\{b_n\}$, then prove that

$$\lim_{n\to\infty} \inf (a_n + b_n) \geq \lim_{n\to\infty} \inf a_n + \lim_{n\to\infty} \inf b_n$$

c) Arbitrary intersection of closed sets is closed.

d) State and prove Taylor's formula for function of several variables.

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

a) If $x \in \mathbb{R}$, $y \in \mathbb{R}$ and $x > 0$, then show that there exists a positive integer n such that $nx > y$.

b) State and prove root test for convergence.

c) If f and g are defined on $[a, b]$ and are differentiable at a point $x \in [a, b]$, then show that $(f+g)$ is also differentiable and $(f+g)'(x) = f'(x) + g'(x)$.

d) Suppose f is a differentiable function of two variables. Show that the maximum value of the directional derivative $D_u f(x)$ is $|\nabla f(x)|$. When does the maximum occur?

Q5) Attempt any one of the following: **[15 Each]**

a) i) State and prove Cauchy criterion for convergence of a sequence in $\mathbb{R}$. **[7]**

ii) Solve the following examples: **[8]**

1) If $x = (x_1, x_2), y = (y_1, y_2)$ are any two points in R^2 and define $d(x, y) = |x_1 - y_1| + |x_2 - y_2|$ then show that d is metric in R^2.

2) Find *lub* and *glb* of the set $\{x \in R \mid 10 < x^2 - 5x + 6 < 20\}$

3) Discuss convergence of the sequence $\left\{\frac{(-1^n)}{n}\right\}$

4) Discuss convergence of the series $\Sigma \frac{1+n}{2+3n^2}$.

b) i) Define the following terms: **[8]**

1) Differentials of composite function

2) Uniform Continuity

3) Convex function

4) Directional Derivative

ii) If Σa_n converges, and if $\{b_n\}$ is monotonic and bounded, prove that $\Sigma a_n b_n$ converges. **[7]**

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PD3722

[Total No. of Pages : 2

[6487]-1005
M.Sc. - I
STATISTICS
STS 511-MJ : Reliability & Statistical Quality Control
(2023 Credit 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 the following. **[5×2=10]**

a) Explain different types of causes with an illustration.

b) Discuss the concept of TQM.

c) Discuss acceptance rejection and acceptance rectification scheme.

d) Obtain the hazard rate for the Makeham family of life distributions.

e) State Renewal Reward theorem and give one real life application.

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

a) Prove the following implications

NBU $\Rightarrow$ NBUE and IFR $\Rightarrow$ DMRL

b) Find reliability function for the 2- out of -3 system.

c) Discuss the relationship between testing of hypothesis and control chart. Also define the terms: false alarm rate and probability of catching a shift.

d) Prove that the hazard rate of series system of components having independent lifetimes is the summation of component hazard rate.

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

a) Explain the procedure of tabular CUSUM chart.

b) Obtain the relation between C_p and C_{pk}.

c) Discuss EWMA control chart for process variability.

d) Prove the following implications

IFR $\Rightarrow$ IFRA and IFRA $\Rightarrow$ NBU

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

a) Obtain AOQ for the double sampling plan.

b) Write a short note on s-shaped property of reliability function.

c) State and prove IFRA closure property of life distributions.

d) Prove that increasing function of associated random variables are associated.

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

a) i) Explain Multivariate control charts for mean vector when dispersion matrix is known. Also, state its limitations. **[9]**

ii) Prove that hazard rate is constant if and only if underlying distributions is exponential distribution. **[6]**

b) i) Define Module of Coherent System. Hence find modules of the coherent system (C, Φ), Where $C = \{1, 2, 3, 4, 5\}$ and $\Phi(\underline{X}) = X1[1 - (1-X2)(1-X3)][1 - (1-X4)(1-X5)]$. **[8]**

ii) Obtain the hazard rate for the proportional hazard rate family and linear hazard rate family. **[7]**

❋ ❋ ❋

Total No. of Questions : 5]

SEAT No. :

PD3723

[Total No. of Pages : 3

[6487]-1006

M.Sc. - I

STATISTICS

STS-512-MJ-STE-104: Actuarial Statistics

(2023 Credit 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 Future life time random variable. Hence obtain its probability distribution function.

b) For the uniform distribution over (0, 80), find the force of mortality and comment on it.

c) Discuss the different assumptions of fractional ages.

d) Suppose force of interest per annum is 0.06. Find d, v and $i^{(2)}$.

e) Find the present value and accumulated value of a 15-year annuity immediate of Rs. 1500/- per annum if the effective rate of interest is 4%.

OR

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

a) Suppose the life length random variable is modeled by a distribution with force of mortality as specified below.

$$\mu_x = \begin{cases} 0.04 & if\ 0 \le s < 15 \\ 0.08 & if\ 15 \le s < 25 \\ 0.12 & if\ 25 \le s < 35 \\ 0.18 & if\ s \ge 35 \end{cases}$$

Find $\bar{A}^{1}_{30:\overline{5|}}$ for $\delta = 0.06$.

b) Show that $A_x = vq_x + vp_x A_{x+1}$.

c) It is given that $A_{x+2} = 0.40, A_x = 0.25,\ A_{x:\overline{20|}} = 0.55$ and $i = 0.03$. Calculate

$1000\,\bar{A}_{x:\overline{20|}}$ under the assumption of uniformity in each unit age interval.

d) Under UDD assumption show that $\bar{A}_x = \dfrac{i}{\delta} A_x$.

P.T.O.

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

a) For n-year temporary life annuity due show that $\ddot{a}_{x:\overline{n}|} = \sum_{k=0}^{n-1} V^k \, {}_kP_x$.

b) Define annuity. Also, derive the expression of actuarial present value for n-year temporary life annuity and whole life annuity.

c) A select and ultimate life table gives $1_x = 100 - x$ and $1_{[x]+t} = 4\,(0.3 - 0.01t)\,1_{x+t}$. Compute the probability that a life now aged 22, selected 1 year ago will survive to age 26.

d) It is given that $\ddot{a}^{(4)}_{\overline{\infty}|} = 17.287$, $A_x = 0.1025$ and deaths are uniformly distributed over each year of age. Calculate $\ddot{a}^{(4)}_x$.

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

a) Show that $\overline{p}(\overline{A}_x) = \dfrac{1}{E(T(x))}$ for effective rate of interest is zero.

b) The following information is given for (x).

   i) The premium for a 20-year endowment insurance of 1 is 0.0349.

   ii) The premium for a 20-year pure endowment of 1 is 0.0230.

   iii) The premium for a 20-year deferred whole life annuity due of 1 year is 0.2087 and is paid for 20 years.

   iv) All premiums are fully discrete annual benefit premiums.

   v) i=0.05

   Calculate the premium for a 20-payment whole life insurance of 1.

c) The actuarial present value at age 27 of a unit benefit to be paid at the moment of death in 5-year endowment insurance with force of interest δ is 0.7395 and with force of interest 2δ is 0.6066. The actuarial present value at age 27 of unit benefit to be paid at the moment of death in a whole life insurance with force of interest δ is 0.5321. Find the annual premium, payable continuously for at most 5 years, for 5-year deferred continuous whole life annuity payable to (27) at the rate of 7000 per annum.

d) Define the following terms.

   i) Loss at Issue Random Variable.

   ii) Equivalence Principle of premium.

   iii) Percentile premium.

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

a) i) Write a note on retrospective reserve and prospective reserves. **[4]**

ii) Prove that the retrospective and prospective reserves are equal at time t for an immediate annuity (payable annually in arrears) of amount B with initial expenses I and actual renewal expenses R. **[6]**

iii) It is given that $P_x = 0.00646$, $P_{x:\overline{n}|} = 0.04166$, $P^{1}_{x:\overline{n}|} = 0.000211$ and $P_{x+n} = 0.01426$. Calculate i. **[5]**

b) i) Assume that premiums are calculated on the basis of the equivalence principle. If ${}_kL$ is the prospective loss random variable for a fully discrete whole life insurance of 1000 issued to (x). It is given that $A_x = 0.125$, $A_{x+k} = 0.4$, ${}^2A_{x+k} = 0.2$, d = 0.05. Then calculate $E({}_kL)$, $Var({}_kL)$ and the aggregate reserve at time k for 100 policies of this type. **[7]**

ii) Calculate ${}_1V_{40}$ given that $P_{40} = 0.01536$, $p_{40} = 0.99647$ and $i = 0.05$. **[4]**

iii) Under the assumption of a uniform distribution of deaths in each year of age, show that ${}_kV(\bar{A}_x) = i/\delta\ {}_kV_x$. **[4]**

ⓘⓘⓘⓘ

Total No. of Questions : 5]

SEAT No. :

PD3724

[Total No. of Pages : 3

[6487]-2001
First Year M.Sc.
STATISTICS
STS-551-MJ : Modern Statistical Inference
(2023 Credit 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 confidence interval.

c) Explain uniformly most powerful unbiased test.

d) Define sufficient statistic with an illustration.

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) Consider the Pareto distribution with pdf $f(x,\theta)=\frac{\theta}{x^{\theta+1}}, x\geq 1, \theta>0$. Based on random sample size n find the moment estimator of θ.

c) Define Joint sufficient statistic. Let $X_1, X_2, \dots, X_n$ 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 $X_1, X_2, \dots, X_n$ 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 $1(\mu, \sigma^2)$ if $X \sim N(\mu, \sigma^2)$.

b) State and prove Rao-Blackwell theorem.

c) Define pitman family. Let $X_1, X_2, \ldots, X_n$ 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 $X_1, X_2, \ldots, X_n$ be a random sample from exponential distribution with mean=θ. 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

$$f(x, \theta) = 2\theta x e^{-\theta x^2}, x > 0, \theta > 0, \text{obtain unbiased estimator of } \frac{1}{\sqrt{\theta}}.$$

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

d) State Basu's theorem. Let X_1 and X_2 be random sample from $N(\mu, \sigma^2)$, σ^2 is known then show that $M = X_1 + X_2$ and $T = X_1 - X_2$ are independent.

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

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

ii) The lifetime of electric bulbs produced by a new technology is assumed to follow the exponential distribution with mean θ. A random sample of n such bulbs were tested and the lifetime recorded for the first h hours only and then stopped. After h hours, m of the bulbs were still burning. Find the MLE of θ. **[7]**

b) i) State Neyman-Pearson Lemma. A sample of size 1 is taken from pdf $f(x) = \dfrac{2}{\theta^2}(\theta - x)$, where $0 \leq x < \theta$. Find an MP test of its size for testing H_0: $\theta = \theta_0$ Vs H_1: $\theta = \theta_1$ ($\theta_1 < \theta_0$). **[8]**

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. **[7]**

✤✤✤✤

Total No. of Questions : 5]

SEAT No. :

PD3725

[Total No. of Pages : 2

[6487]-2004
M.Sc. - I
STATISTICS
STS-560-MJ: Advances in Generalized Linear Models
(2023 Credit 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 of the following. **[2 Each]**

a) What are the key components of a Generalized Linear Model (GLM) and how do they differ from those of a GLM?

b) Define concordance and discordance in binary logistic regression.

c) Explain the concept of zero inflation in data analysis.

d) Define the Beta-Binomial model.

e) Define a normal linear mixed model (LMM).

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

a) Let y_i has a $N(\mu_i, \sigma^2)$ distribution, i = 1,2,...,n. Formulate the normal linear model as a special case of a GLM, specifying the random component, linear predictor, and link function.

b) Explain Multinomial model in detail.

c) Explain the Quasi-Likelihood Method in statistical modeling. Discuss its advantages over traditional likelihood-based methods and give one example.

d) Show that the exponential dispersion family representation for the gamma distribution.

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

a) Define Pearson residual and write the form of the deviance residual for

i) binomial GLM,

ii) Poisson GLM.

b) Explain Logistic regression model for case control studies.

c) For a GLM with canonical link function, explain how the likelihood equations imply that the residual vector $e = (y - \hat{\mu})$ is orthogonal with C(X).

d) Define quasi-likelihood. How does it differ from likelihood?

P.T.O.

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

a) Estimate the dispersion parameter ϕ and comment on the overdispersion for a Poisson GLM. The mean response $\hat{\mu}_i$ and deviance residuals are as follows:

Observations	$\hat{\mu}_i$	y_i	Residual
1	2.5	3	0.5
2	4.0	6	1.0
3	5.0	9	1.5

b) Define deviance and explain its use in model comparison.

c) Compare Poisson and Negative Binomial regression models.

d) Differentiate between likelihood ratio, Wald, and Score tests.

Q5) Attempt any one of the following. **[15 Each]**

a) i) The following table gives the number of insurance claims Y is modeled as a function of age X.

Age	25	30	40	50	60
Claims	1	3	4	6	9

Use normal equations to estimate β_0 and β_1. Also, predict the number of claims for age 35. Check if overdispersion is present (compute residual deviance and degrees of freedom). **[8]**

ii) Compare the marginal modeling with Generalized Estimating Equations (GEE) in the context of analyzing correlated or clustered data and discuss their respective strengths, limitations, and applications. **[7]**

b) i) Discuss the estimation procedure in mixed models using residual maximum likelihood. **[8]**

ii) Describe the use of GLMMs in modeling repeated binary responses. **[7]**

ⓘⓘⓘⓘ

Total No. of Questions : 5]

SEAT No. :

PD3726

[Total No. of Pages : 3

[6487]-2005
First Year M.Sc.
STATISTICS
STS - 561 - MJ : STATISTICAL METHODS IN EPIDEMIOLOGY
(2023 Credit 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 questions. **[2 each]**

a) Differentiate between Prevalence and incidence.

b) Define ROC curve and what are the components of ROC curve?

c) Explain the terms: Infection rate and fatality rate.

d) Describe sensitivity related to diagnostic testing.

e) Compare between non-communicable and communicable disease.

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

a) Explain the SIR model.

b) A study conducted on the association between postmenopausal hormone use and the risk of coronary heart disease (CHD) among postmenopausal women. A data is collected from a cohort of 1000 postmenopausal women over a 10-year period. Among these women, 200 use postmenopausal hormones (exposed group) and 800 do not use postmenopausal hormones (unexposed group). During the study period, observe that 20 postmenopausal hormone users and 40 non-users develop coronary heart disease. Calculate the Relative Risk (RR) and give the interpretation.

c) Consider a region with an initial number of confirmed COVID- 19 cases of 200 and after 5 days the no. of cases increases to 800. Calculate the doubling timing and interpret the result.

d) Write a brief note on flattening of the curve.

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

a) Explain the Cohort study with the help of diagram.

b) Write any four applications of SEIR model.

c) Discuss the significance of preventative strategies in epidemiology.

d) Let n = 3539 participants who attended the seventh examination of the offspring in the Framingham Heart Study. Descriptive statistics on variables measured in the sample are shown below:

Summary Statistics on Participants Attending the Seventh Examination of the Framingham Offspring Study (n = 3539)

	n	Mean ($\overline{x}$)	Standard deviation(s)
Systolic blood pressure	3534	127.3	19.0
Diastolic blood pressure	3532	74.0	9.9
Total Serum cholesterol	3310	200.3	36.8
Weight (lbs)	3506	174.4	38.7
Hight(in)	3326	65.957	3.749
Body Mass Index (BMI)	3326	28.15	5.32

Generate 95% confidence interval for systolic blood pressure using data collected in the Framingham offspring study.

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

a) Define the Risk and state associate measures of risk.

b) Explain the role of adaptive Markov Chain Monte Carlo (MCMC) methods in Bayesian inference for latent variables.

c) Suppose we followed a population of 150 persons for 1 year and 25 had a disease of interest at the start of follow up and another 15 new cases develop during the year.

   i) What is the period of prevalence for the year?

   ii) What is the point prevalence at the starts of the period?

   iii) What is the cumulative incidence for the 1-year period?

d) If a disease has a reproduction number of 2.5 and preventive measures reduced the probability of transmission by 70% while preventing 80% of contacts, Calculate preventative number.

Q5) Attempt any one of the following questions. **[15 each]**

a) i) Compare and contrast cohort studies, case-control studies, and randomized controlled trials. **[7]**

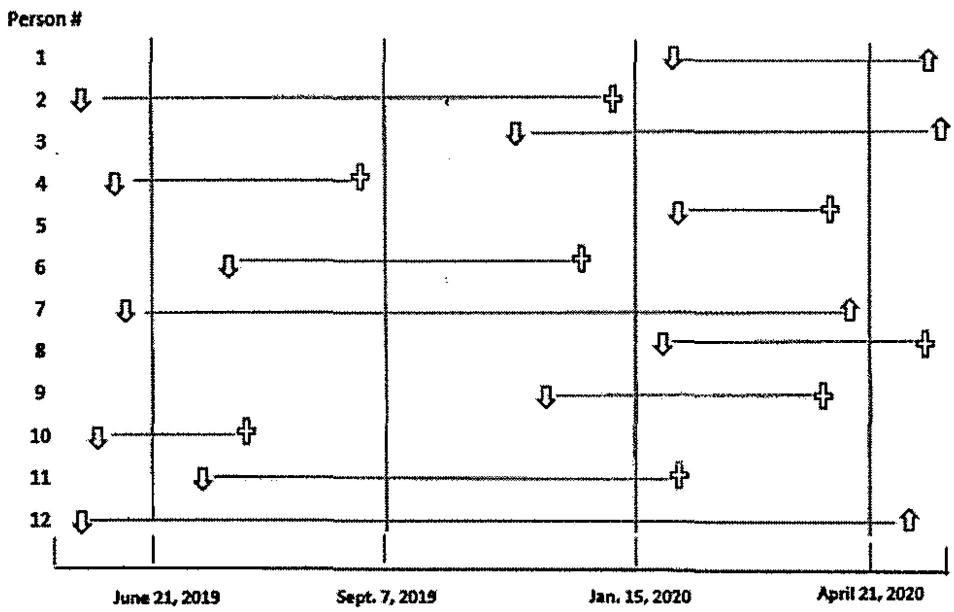
ii) Explain the Reed-Frost chain binomial model. **[8]**

b) i) Table shows data from a cohort study of oral contraceptive (OC) use and Myocardial Infarction (MI) among women aged 16-49 years.

		MI		
		Yes	No	Total
OC Use	Yes	27	455	482
	No	77	1831	1908
	Total	104	2286	2390

Calculate Absolute Risk for the above table and give your interpretation. **[7]**

ii) Figure (attached below) represents 12 new cases of illness over about 10 months in a population of 40 persons. Each horizontal line represents one person. The down arrow indicates the date of onset of illness. The solid red line represents the duration of illness. The up arrow and the cross represent the date of recovery and date of death, respectively. **[8]**



Using Figure, calculate the following:

1) Point prevalence on January 15, 2020
2) Period prevalence from June 21, 2019 to April 21, 2020
3) Incidence rate from June 21, 2019 to April 21, 2020 using the population alive on September 7, 2019 as the denominator. Express the rate per 100 (round up to whole number)

✪ ✪ ✪

Total No. of Questions : 5]

SEAT No. :

PD3727

[Total No. of Pages : 3

[6487]-2006

First Year M.Sc.

STATISTICS

STS-562-MJ : Discrete Data Analysis Equivalent Course

(2023 Credit 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 : **[2 each]**

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

b) Give any two measures of association for contingency table.

c) Explain the concept of marginal Odd's ratio.

d) Explain multinomial model.

e) Define 'Deviance' for a generalized linear model.

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

a) In the context of generalized linear models, explain the Wald statistic and its use in hypothesis testing. Derive the confidence interval for a parameter using the Wald method.

b) Explain the following terms with example :

i) Difference of proportions

ii) Relative Risk

iii) Odd's Ratio

c) Obtain the parameter of Poisson regression model using MLE method.

d) Construct the confidence interval for Binomial parameter using Wald's test.

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

a) Explain the effect of sparseness of χ^2 and G^2.

b) What are generalized linear models for clustered data? Discuss how correlation within clusters is handled in such models.

c) Write a short note on Fisher's exact test for 2×2 tables.

d) According to recent UN figures, the annual gun homicide rate is 62.4 per one million residents in the United States and 1.3 per one million residents in the UK. Compare the proportion of residents killed annually by guns using difference of proportions and relative risk. When both proportions are very close to 0, as here, which measure is more useful for describing the strength of association? Why?

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

a) Differentiate between polytomous logit models for ordinal and nominal response.

b) Explain the role of link functions in GLMs. Why is the logit link function commonly used in logistic regression?

c) A survey asks 200 people about their favorite ice cream flavor: Vanilla, Chocolate, or Strawberry. The observed counts for Vanilla are 80, for Chocolate are 70 and for Strawberry are 50. Construct the likelihood function and find the MLE of the category probabilities.

d) Compare and contrast the Pearson's chi-square test and the likelihood ratio test.

Q5) Attempt any one of the following : **[15 each]**

a) i) The following table comes from one of the first studies of the link between lung cancer and smoking, by Richard Doll and A. Bradford Hill. In 20 hospitals in London, UK, patients admitted with lung cancer in the previous year were queried about their smoking behavior. For each patient admitted, researchers studied the smoking behavior of a noncancer control patient at the same hospital of the same sex and within the same 5-year grouping on age. A smoker was defined as a person who had smoked at least one cigarette a day for at least a year.

Have Smoked	Lung Cancer	
	Cases	Controls
Yes	688	650
No	21	59
Total	709	709

Identify the response variable and the explanatory variable, also identify the type of study this was. Can you use these data to compare smokers with nonsmokers in terms of the proportion who suffered lung cancer? Why or why not? Summarize the association, and explain how to interpret it. **[8]**

ii) Explain the logistic regression model with single explanatory variable. Also, obtain the maximum likelihood equation for it. **[7]**

b) i) A sample of women suffering from excessive menstrual bleeding have been taking an analgesic designed to diminish the effects. A new analgesic is claimed to provide greater relief. After trying the new analgesic, 40 women reported greater relief with the standard analgesic, and 60 reported greater relief with the new one.

Test the hypothesis that the probability of greater relief with the standard analgesic is the same as the probability of greater relief with the new analgesic. Report and interpret the P-value for the two-sided alternative. Also, construct and interpret a 95% confidence interval for the probability of greater relief with the new analgesic. **[9]**

ii) Explain the concept of homogeneity of proportions and how it is tested in an $r \times c$ table. **[3]**

iii) Define the terms: **[3]**

1) Comprehensive log linear model
2) Hierarchical log linear model and
3) Graphical log linear model