

Total No. of Questions : 5]

SEAT No. :

PE-4806

[Total No. of Pages : 2

[6521]-1

M.Sc. (Part - I)

STATISTICS

STS-501-MJ : Fundamental of Analysis and Calculus
(2023 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 numbers is countable set.
- b) Define uniformly continuous function.
- c) Find limit inferior and limit superior of a sequence $a_n = \sin\left(\frac{n\pi}{2}\right)$.
- d) Define continuity of a function from $\mathbb{R}$ to $\mathbb{R}$.
- e) Define convex function with an illustration.

Q2) Attempt any three of the following :

[5 each]

- a) Show that every bounded infinite subset of $\mathbb{R}$ has atleast one limit point in $\mathbb{R}$.
- b) State and prove ratio test for convergence of series.
- c) State and prove mean value theorem.
- d) Define power series and radius of convergence. Obtain the radius of

convergence for the following series: $\sum_{n=1}^{\infty} \frac{(-1)^n n^2 x^n}{5^n \sqrt{n^5}}$.

P.T.O.

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

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

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

- b) 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?
- c) Show that arbitrary intersection of closed sets is closed.
- d) If f and g are defined on $[a, b]$ and are differentiable at a point $x \in [a, b]$, then show that $(f \cdot g)$ is also differentiable and $(f \cdot g)'(x) = g(x)f'(x) + f(x)g'(x)$.

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

- a) Define compact set. Show that $[0, 1]$ is compact set.
- b) Show that a series of nonnegative terms converges if and only if its partial sums form a bounded sequence.
- c) State and prove Taylor's series expansion.
- d) Show that $\int_0^1 \frac{\sin(1/x)}{\sqrt{x}} dx$ is convergent and also absolutely convergent.

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

- a) i) If $S_n = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!}$, for $n = 1, 2, 3, \dots$. Prove that $2 < \lim_{n \rightarrow \infty} S_n < 3$. [7]
- ii) Define open set and interior point. Show that every neighborhood set open set. [8]
- b) i) Define the following Terms: [8]
- A) Absolute continuity B) Functions of bounded variation
- C) Concave function D) Directional derivative
- ii) State and prove Cauchy criterion for convergence of a sequence in $\mathbb{R}$. [7]



Total No. of Questions : 5]

SEAT No. :

PE-4807

[Total No. of Pages : 2

[6521]-2
M.Sc. (Part - I)
STATISTICS
STS-510-MJ : Optimization Techniques
(2023 Pattern) (Semester - I) (4 Credits)

Time : 3 Hours]

[Max. Marks : 70

Instructions to the candidates:

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

Q1) Attempt all questions :

[5 × 2 = 10]

- a) Write the concept of looping and dangling in network diagram.
- b) State the applications of zero-one integer programming problem.
- c) Explain the reason of restrictions in assignment problem.
- d) What are the advantages of duality (any two).
- e) Find the cut for equation : $\frac{5}{3} = x_1 + \frac{4}{3}x_2 - \frac{1}{4}s_1$ (x_1 source row)

Q2) Attempt any THREE of the following :

[3 × 5 = 15]

- a) Write the rules for constructing dual from the primal.
- b) Describe the method of solving the simplex method of linear programming problem in case of unrestricted variables with suitable example.
- c) What is the algorithm of branch and bound method.
- d) Explain how a trans-shipment problem can be solved as a transportation problem.

P.T.O.

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

- a) Explain the method of solution to the problem in case of unbounded solution and infeasible solution in assignment problem.
- b) Compute the solution of following dynamic programming problem:
Minimize $Z = y_1^2 + y_2^2 + y_3^2$
Subject to the constraints
 $y_1 + y_2 + y_3 \geq 15$
 $y_1, y_2, y_3 \geq 0$
- c) Write the steps of Wolf's modified simplex method.
- d) Narrate the polynomial time algorithm.

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

- a) What is the input model conditioning and pivot rule complexity in linear programming?
- b) State the minimal spanning tree algorithm of network.
- c) Explain the concept of dynamic programming and the relation between dynamic and linear programming approach.
- d) Write the application of quadratic programming in case of demand for electricity.

Q5) Attempt any ONE of the following :

- a) i) Obtain the solution to the following linear programming problem; **[9]**
Max $Z = x_1 + 2x_2 + 3x_3 - x_4$
Subject to constraint
 $x_1 + 2x_2 + 3x_3 = 15$
 $2x_1 + x_2 + 5x_3 = 20$
 $x_1 + 2x_2 + x_3 + x_4 = 10$
 $x_1, x_2, x_3, x_4 \geq 0$
ii) Describe the Karmarkar's polynomial time algorithm with one illustration. **[6]**
- b) i) Explain resource analysis in network scheduling. **[5]**
ii) Find the difference between linear and non-linear programming by using graphical method. **[5]**
iii) Write a procedure to solve the maximization transportation problem. **[5]**



Total No. of Questions : 5]

SEAT No. :

PE-4808

[Total No. of Pages : 2

[6521]-3
M.Sc.
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 of the following.

[5 × 2 = 10]

- a) Define the term Quality.
- b) Find the hazard for the Weibull distribution with scale parameter α and shape parameter β .
- c) Define Renewal Reward Process. Hence give one real life example of it.
- d) Justify the use of 3σ control limits for the control charts when characteristic under study is not normally distributed.
- e) Derive the hazard rate for the Pareto distribution.

Q2) Attempt any Three questions out of Four questions.

[3 × 5 = 15]

- a) Explain the construction of Conforming Run Length (CRL) chart for process fraction defective.
- b) Prove the following implications
 $DMRL \Rightarrow NBUE$ and $NBUE \Rightarrow HNBUE$
- c) Write a short note on Double sampling plan.
- d) Suppose that quality characteristic has normal distribution with specification limits as 95 ± 5 . A random sample of size 30 gives $\bar{X} = 97, \hat{\sigma} = 1.6$. Compute C_p, C_{pk} . Also interpret it.

P.T.O.

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

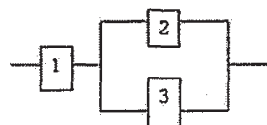
- State the Sigmund's approximation formula for ARL and determine the parameters of CUSUM chart.
- Determine the control limits of EWMA control chart for process variation.
- Prove that any subset of associated random variables are associated.
- Define term associated random variables and prove that increasing function of associated random variables are associated.

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

- For the 2-out of -3 system find minimal path sets and hence represent the R.B.D. of the system using minimal path sets.
- Explain age replacement policy and random age replacement policy.
- A system consist of 4 components functions when both the components 1 and 4 functions and at least one of the remaining component functions. Find the reliability of the system.
- Show by graphical presentation working of synthetic control chart.

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

- Define Module of Coherent System. Hence find modules of the coherent system $(C, \Phi, \text{ where } C = \{1,2,3,4,5\} \text{ and } \Phi(\underline{X}) = X_1 [1 - (1 - X_2)(1 - X_3)][1 - (1 - X_4)(1 - X_5)]$ [7]
 - State and prove IFRA characteristic property. [8]
- Consider the coherent system with the following Reliability block diagram (R.B.D.) [9]



- Find I) Structure function for the coherent system
 II) Minimal path and cut vector for the coherent system
 III) Represent coherent system using minimal path set.
- Give t two definitions of Increasing failure rate class of life distribution and prove that both the definitions are equivalent. [6]



Total No. of Questions : 5]

SEAT No. :

PE-4809

[Total No. of Pages : 4

[6521]-4
M.Sc. (Part - I)
STATISTICS
STS-512-MJ : Actuarial Statistics
(2023 Pattern) (Semester - I) (4 Credits)

Time : 3 Hours]

[Max. Marks : 70

Instructions to the candidates:

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

Q1) Attempt each of the following :

[5 × 2 = 10]

a) Show that, $\frac{d}{dx} {}_tP_X = {}_tP_X(\mu_x - \mu_{x+t}).$

b) Define Gompertz's force of mortality.

c) Examine whether following can serve as survival function.

$$s(x) = \exp\{x - 0.7(2^x - 1)\}$$

d) Suppose the life length r.v. X has a distribution with survival function:

$$s(x) = 1 - \frac{x^2}{100} \quad 0 \leq x \leq 10.$$

Find median future life time of (5).

e) In how many years will a sum of money double itself at compound interest with effective rate $i = 0.05$?

P.T.O.

Q2) Attempt any three of the following :

[3 × 5 = 15]

- a) Suppose the life length $r.v.$ X is modelled by a uniform distribution over $(0,100)$. Find the p.d.f. of $T(25)$.
- b) Suppose the life length $r.v.$ X is modelled by
 - i) Gompertz's force of mortality and
 - ii) Makeham's force of mortality.In each case, find the p.d.f. $g(t)$ of $T(x)$.
- c) With effective rate of interest 6% per annum find:
 - i) Nominal rate of interest convertible quarterly.
 - ii) Effective rate of discount.
 - iii) Accumulated value of 25000 at the end of 4th year.
 - iv) Present value of Rs. 10000 due after the end of 5th year.
 - v) Force of interest.
- d) Write present value $r.v.$ and Actuarial Present Value of all insurance products when benefit payable at the moment of death.

Q3) Attempt any three of the following :

[3 × 5 = 15]

a) Let $F(x) = 1 - \left(1 - \frac{x}{120}\right)^{1/6}; 0 \leq x \leq 120$

Calculate,

- i) μ_x
- ii) q_x for $x = 20$
- iii) Probability that a newborn life survives beyond age 30.
- iv) Probability that a life aged 30 dies before age 50.
- v) Probability that a life aged 40 survives beyond age 65.

- b) For a certain animal population, $l_x = \frac{l_0}{(1+x)^2} \quad (x \geq 0)$

Calculate,

- i) the complete expectation of life at birth.
 - ii) the force of mortality at age 1 year.
 - iii) the chance that a newly-born animal will die between ages 1 and 2 years.
- c) If $s(x) = \exp\{-0.00133[(1.07152)^x - 1]\}$; $0 \leq x \leq 100$ find the following probabilities:
- i) (40) will die in the next decade.
 - ii) (40) will die between his 50th and 60th birthday.
 - iii) (40) will die either between (45, 50) or between (60, 70).
- d) Explain the following forms of annuity related to life insurance:
- i) n-year temporary life annuity
 - ii) Whole Life Annuity.

Q4) Attempt any three of the following :

[3 × 5 = 15]

- a) Let $F(x) = 1 - \left(1 - \frac{x}{120}\right)^{1/6}$; $0 \leq x \leq 120$;
- i) Find μ_x
 - ii) Calculate q_x for $x = 20$ and for $x = 110$
 - iii) Find e_x^0
 - iv) Calculate e_x^0 for $x = 30$ and $x = 80$
- b) Suppose the p.d.f: $g(t)$ of $T(30)$ is given by,

$$g(t) = \begin{cases} 0.12e^{-0.12t} & ; 0 \leq t < 5 \\ 0.18e^{-0.18t+0.3} & ; t \geq 5 \end{cases}$$

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

- c) Prove that, $\ddot{a}_{x:\overline{n}|} = \sum_{k=0}^{n-1} v^k p_x$
- d) How much money should be deposited in an account which has 5% annual effective rate of interest on April 01, 2024 if we want to make a withdrawal of 5000 on April 01, 2025, of 10000 on April 01, 2026, of 15000 on April 01, 2027.

Q5) Attempt any One of the following :

- a) i) It is given that $q_{28} = 0.135$, $q_{29} = 0.146$, $q_{30} = 0.159$, $q_{31} = 0.173$, $q_{32} = 0.188$ and $i = 0.05$ then [8]
- A) Find the actuarial present value of benefit of 1000 in a 5 year term insurance, 5 year endowment insurance issued to (28).
- B) Find the actuarial present value of benefit of 1000 in a 2 year deferred 2 year term insurance, 2 year deferred 2 year endowment insurance issued to (28).
- ii) Show that $a_{x:\overline{n}|} < \bar{a}_{x:\overline{n}|} < \ddot{a}_{x:\overline{n}|}$ [7]
- b) i) It is given that the force of interest is $\delta = 6\%$ and the actuarial present value at age 27 of a unit benefit to be paid at the moment of death in a 5 year endowment insurance with force of interest δ is 0.7395 and with force of interest 2δ is 0.6006. Find the actuarial present value at age 27 of a continuous 5 year temporary life annuity payable at the rate of 10,000 per annum. Also find standard deviation of the annuity. [8]
- ii) An aviary of birds which has a constant intake 1500 new born birds per year experiences the following mortality rates. [7]

Age(x)	0	1	2	3	4	5
q_x	0.3	0.1	0.2	0.4	0.7	1

Construct the life table with columns P_x , l_x , d_x , L_x , T_x and e_x .



Total No. of Questions : 5]

SEAT No. :

PE-4810

[Total No. of Pages : 2

[6521]-6

M.Sc.

STATISTICS

STS-551-MJ : Modern Statistical Inference (4 Credits)
(2023 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 Ancillary statistics with an illustration.
- b) Prove or disprove: If T is an unbiased estimator for θ then T^2 is unbiased estimator for θ^2 .
- c) Explain the term shortest expected length confidence intervals.
- d) State the factorization criterion for a sufficient statistic.
- e) Define the term uniformly most accurate family of confidence sets.

Q2) Attempt any THREE of the following :

[3 × 5 = 15]

- a) State and prove Basu's theorem.
- b) Let $X_1, X_2, \dots, X_n$ be a random sample from *Bernoulli* (p) find the MVBUE of p .
- c) If $X_1, X_2, \dots, X_n$ is random sample from $f(x, \theta)$, $\theta \in \Omega$ which is member of one parameter exponential family and joint pmf is given by

$$L(x, \theta) = u(\theta) T(x) + nV(\theta) + W(x)$$

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

- d) Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, 1)$, show that $T = \sum_{i=1}^n X_i$ is complete statistic for θ .

P.T.O.

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

- a) Define a sufficient statistic and obtain sufficient statistics for θ when $X_1, X_2, \dots, X_n$ are iid Poisson random variables with parameter θ .
- b) Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$ for $\mu \geq \mu_0$ find maximum likelihood estimator of μ .
- c) Define pitman family. If $a(\theta) \downarrow$ and $b(\theta) \uparrow$ then prove that $\max\{a^{-1}(X_{(1)}), b^{-1}(X_{(n)})\}$ is minimal sufficient statistic for θ .
- d) State and prove Lehmann-Scheffe theorem.

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

- a) Define the term fisher Information. Compute it in case of $N(\mu, 1)$ and $N(0, \sigma^2)$ distribution.
- b) Let $X_1, X_2, \dots, X_n$ is a random sample of size n from discrete distribution with $P(X_1 = 1) = \frac{2(1-\theta)}{2-\theta}, P(X_2 = 2) = \frac{\theta}{2-\theta}$ where $0 < \theta < 1$ is an unknown. Find the estimator of θ by using method of moments.
- c) State and Prove Rao-Blackwell theorem.
- d) Define monotone likelihood ratio property of a distribution. Does *Bernoulli* (p) and *Poisson* (λ) possess an MLR property?

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

- a) i) State and prove Neyman-Pearson Lemma. **[7]**
ii) Let $X_1, X_2, \dots, X_n$ be a random sample from $U(-2\theta, 5\theta)$ find maximum likelihood estimator of θ . **[4]**
iii) Verify the distribution having following pdf is from one parameter exponential family or not.

$$f(x, \theta) = \frac{1}{\pi} \left[1 + (x - \theta)^2 \right]^{-1}; -\infty < x < \infty, -\infty < \theta < \infty. \quad [4]$$

- b) i) Let $X_1, X_2, \dots, X_n$ be random sample from $f(x, \theta) = \frac{\theta}{x^{\theta+1}}, x > 1, \theta > 0$
Obtain shortest Expected Length CI for θ . **[7]**
ii) Let X be a random sample from $P(\lambda)$. The parameter λ has prior distribution over Θ , given by the pdf $\pi(\lambda) = e^{-\lambda}$ if $\lambda > 0$ and zero otherwise. Under squared error loss function find the Bayes estimator for λ . **[8]**



Total No. of Questions : 5]

SEAT No. :

PE-4811

[Total No. of Pages : 3

[6521]-7

M.Sc.

STATISTICS

**STS-560-MJ : Advances in Generalized Linear Models
(2023 Pattern) (Semester - II) (4 Credits)**

Time : 3 Hours]

[Max. Marks : 70

Instructions to the condidates:

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

Q1) Attempt all of the following questions :

[5 × 2 = 10]

- a) What is the key feature of the complementary log-log model?
- b) Give the intuitive motivation for the 'log' link function in modeling Poisson Count Data.
- c) Define Likelihood ratio and Relative risk ratio.
- d) What are the key components of a GLMM and how do they differ from those of a Generalized Linear Model (GLM)?
- e) Define quasi-likelihood and provide an example of its application in statistical modeling.

Q2) Attempt any 3 questions out of 4 questions :

[3 × 5 = 15]

- a) Explain the key difference between a classical Generalized Linear Model (GLM) and a Bayesian Generalized Linear Model (GLM).
- b) Explain Poisson regression model in detail and state any two applications.
- c) Let Y_{ii} has a $N(\mu_i, \sigma^2)$ distribution, $i = 1, \dots, n$. Formulate the normal linear model as a special case of a GLM, specifying the random component, linear predictor, and link function.
- d) Show that the exponential dispersion family representation for the gamma distribution.

P.T.O.

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

- a) Describe the purpose of the link function g , and why the identity link is the standard one with normal responses but is not often used with binary or count responses.
- b) Define Pearson residual and Write down the form of the deviance residual for (i) a binomial GLM, (ii) a Poisson GLM.
- c) Give the general frameworks of the (i) Likelihood Ratio test (ii) Wald Test
- d) Explain Logistic regression model for case control studies.

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

- a) For a binomial random variable with parameter π consider the null model. Explain how to invert the Wald, likelihood-ratio, and score tests of $H_0: \pi = \pi_0$ against $H_1: \pi \neq \pi_0$ to obtain 95% confidence intervals for π
- b) Explain Multinomial model in Detail.
- c) Discuss the following goodness of model fit measures i) Pearson Chisquare ii) Akaike Information Criteria.
- d) Let $\{Y_i\}$ are independent Poisson observations from a single group. Find the likelihood equation for estimating $\mu = E(Y_i)$. Show that $\hat{\mu} = \hat{y}$ regardless of the link function.

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

- a) i) The following are output table from a logistic regression model:[8]

	Coef	Std error	z	P > z	[0.025	0.975]
Intercept	-1.2542	0.253	-4.958	0.000	-1.751	-0.757
Hours Studied	0.6107	0.123	4.963	0.000	0.369	0.853
Prep Exams	0.3151	0.071	4.436	0.000	0.176	0.454

Interpret the coefficient estimates for the intercept, Hours_Studied, and Prep_Exams variables.

- A) Assess the statistical significance of each coefficient using the p-values.
 - B) Determine whether the Hours_Studied and Prep_Exams variables are statistically significant predictors of the outcome.
 - C) Discuss the implications of the confidence intervals for the coefficient estimates.
- 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) For a binomial distribution with beta prior, show that the marginal distribution of $s = n y$ is the beta-binomial. State its mean and variance. [8]
- ii) Compare and contrast the probit model with the logit model in the context of binary classification tasks and discuss their similarities, differences, advantages, and limitations, examples. [7]



Total No. of Questions : 5]

SEAT No. :

PE-4812

[Total No. of Pages : 3

[6521]-8

M.Sc.

STATISTICS

STS-561-MJ : Statistical Methods in Epidemiology

(2023 Pattern) (Semester - II) (4 Credits)

Time : 3 Hours]

[Max. Marks : 70

Instructions to the candidates:

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

Q1) Attempt each of the following questions :

[2 each]

- a) Define: Infection period and incubation period.
- b) Differentiate between Cohort studies and Case-Control studies.
- c) Define the reproduction number (R_0) and state which factors affect (R_0).
- d) Discuss any one compartment model in short.
- e) Association between meat consumption and enteritis necroticans in Papua New Guinea is given as follows:

Disease (enteritis necroticans)	Exposure (recent meat ingestion)			
		Yes	No	Total
	Yes	50	11	61
	No	16	41	57
Total		66	52	118

Calculate odds ratio of cases and control

P.T.O.

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

- a) Explain the role of Markov Chain Monte Carlo (MCMC) methods in Bayesian inference for latent variables
- b) What are the different types of tests used in epidemiology, and how do they contribute to disease detection and control.
- c) Explain the mathematical relationship among prevalence, incidence, and the average duration of disease.
- d) Define sensitivity in the context of diagnostic testing in epidemiology. How is it calculated, and what does it represent?

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

- a) Illustrate all the modes of direct transmission.
- b) Describe case-control study design with the help of diagram and give an example of it.
- c) Define ROC curve and explain how it is constructed.
- d) Explain the Reed-Frost chain binomial epidemic model.

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

- a) Explain spatial clustering in epidemiology.
- b) Describe randomized control studies and draw the well labelled diagram for it.
- c) Discuss the importance of quantifying spatial patterns and clusters in epidemiological studies.
- d) The cohort study design of Radiation Exposure and Cancer Occurrence is given below.

	Case	Non-Case	At Risk
Exposed	52	2820	2872
Non-Exposed	6	5043	5049

Calculate the Relative Risk

Q5) Attempt any one of the following questions :

[15 each]

- a) i) Briefly explain the use of Bayesian inference in estimating unobserved epidemiological parameters. **[7]**
- ii) Describe how spatial patterns of infection can be quantified and visualized using real-world data. **[8]**
- b) i) Use the SIR model to demonstrate how intervention strategies like social distancing or vaccination alter the epidemic curve. **[8]**
- ii) Table shows data from a cohort study of oral contraceptive (OC) use and Myocardial Infarction (MI) among women aged 16-49 years. **[7]**

		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.



Total No. of Questions : 5]

SEAT No. :

PE-4813

[Total No. of Pages : 3

[6521]-9

M.Sc.

STATISTICS

**STS-562-MJ : Discrete Data Analysis Equivalent Course
(2023 Pattern) (Semester - II)**

Time : 3 Hours]

[Max. Marks : 70

Instructions to the condidates:

- 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) State the relationship between Odd's Ratio and Relative Risk.
- b) Discuss about the relationship between Multinomial and Poisson distribution.
- c) Explain the concept of conditional Odd's ratio.
- d) Define 'Deviance' for a generalized linear model.
- e) What is probit model?

Q2) Attempt any 3 questions out of 4 questions :

[3 × 5 = 15]

- a) Explain Score test and construct the confidence interval using it.
- b) Construct the confidence interval for Binomial parameter using wolds test.
- c) Write a short note on Proportional Odd's cumulative logit model.
- d) Write a short note on classes of log linear model.

Q3) Attempt any 3 questions out of 4 questions :

[3 × 5 = 15]

- a) Give the test procedure for testing the parameters in the logistic regression model.
- b) Distinguish between population-averaged and subject-specific models for longitudinal data.
- c) Explain Poisson regression model in detail and state any two applications of it.
- d) Explain the terms:
 - i) Sensitivity
 - ii) Specificity

P.T.O.

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

- a) Explain the effect of sparseness of x^2 and G^2 .
- b) Write a short note on the Cochran-Mantel-Haenszel test for $2 \times 2 \times k$ contingency table.
- c) Write a short note on log-linear model.
- 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?

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

- 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) Discuss Cumulative logit model for multinomial response model with example. **[7]**

- b) i) Explain the logistic regression model with single explanatory variable. Also, obtain the maximum likelihood equation for it. **[7]**
- ii) 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.
- A) 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.
- B) Construct and interpret a 95% confidence interval for the probability of greater relief with the new analgesic. **[8]**



Total No. of Questions : 4]

SEAT No. :

[Total No. of Pages : 2

PE4814

[6521]-11

M.A./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 tables and scientific calculator is allowed.*
- 4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt all questions.

[5×1=5]

- a) Define countable set.
- b) What is the relationship between continuity and uniform continuity?
- c) What is the definition of a Cauchy sequence?
- d) Define Jacobian.
- e) Find the limit supremum of $a_n = \cos(n)$.

Q2) Attempt any two of the following:

[2×5=10]

- a) For the power series $\sum_{n=1}^{\infty} nx^n$, find the radius of convergence R. Test convergence at $x = 1$ and $x = -1$.
- b) Prove or disprove: If series $\sum a_n$ is convergent then series $\sum |a_n|$ is also convergent.
- c) Show that arbitrary union of open set is open.

Q3) Attempt any two of the following:

[2×5=10]

- a) State and prove mean value theorem.
- b) Determine the convergence of the given series.

i) $\sum \frac{2^n}{3^n + 5}$

ii) $\sum \frac{\sqrt{n}}{n+1}$

- c) Define compact set. Show that $[0, 1]$ is compact set.

P.T.O.

Q4) Attempt any one of the following:

[1×10=10]

- a) i) Discuss convergence of log series. Also find it's radius of convergence. **[4]**
- ii) Find the Taylor polynomial of degree 2 of $f(x) = \ln(1+x)$ about $x = 0$, and bound the remainder $R_2(0.5)$ using Lagrange form of remainder. **[6]**
- b) i) Discuss Cauchy criterion for convergence with illustration. **[7]**
- ii) Prove or disprove: Set of irrational numbers is uncountable. **[3]**



Total No. of Questions : 5]

SEAT No. :

PE-4815

[Total No. of Pages : 3

[6521]-12

M.Sc. (Part - I)

STATISTICS

STS-502-MJ : Linear Algebra

(2024 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) Define Inner product space.
- b) Find the characteristic polynomial of $A = \begin{bmatrix} 7 & -1 \\ 2 & 6 \end{bmatrix}$.
- c) If a matrix A is invertible then show that inverse of A is unique.
- d) Show that positive definite real symmetric matrix is non singular.
- e) Define Geometric and Algebraic multiplicity.

Q2) Attempt any THREE of the following :

[3 × 5 = 15]

- a) Let $V = \mathbb{R}^2$ with operations $(x_1, y_1) + (x_2, y_2) = (x_1, y_1 + y_2)$ and $k(x_1, y_1) = (kx_1, ky_1)$. Check whether V is vector space or not.
- b) Explain Gram-Schmidt orthogonalization process.
- c) Define G - inverse and obtain two distinct G-inverse for the following matrix.

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

- d) Show that a real symmetric matrix is positive definite if all its eigen values are positive.

P.T.O.

Q3) Attempt any THREE of the following :

[3 × 5 = 15]

a) Find eigenspace corresponding to each eigenvalue of $A = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 3 & -1 \\ 2 & 1 & 3 \end{bmatrix}$.

b) Find the rank and nullity of the matrix $A = \begin{bmatrix} 1 & 4 & 5 & 6 & 9 \\ 3 & -2 & 1 & 4 & -1 \\ -1 & 0 & -1 & -2 & -1 \\ 2 & 3 & 5 & 7 & 8 \end{bmatrix}$.

- c) Define Symmetric matrix and Skew Symmetric matrix with an illustration and Show that every square matrix A can be expressed as sum of Symmetric matrix and Skew Symmetric matrix.
- d) State and prove the necessary and sufficient condition for a quadratic form to be non negative definite.

Q4) Attempt any THREE of the following :

[3 × 5 = 15]

- a) Let S be a set with two or more vectors in a vector space V . Then show that S is linearly dependent if and only if at least one of the vectors in S is expressible as a linear combination of the rest of the vectors in S .
- b) State and prove Caley-Hamilton theorem.
- c) Define linear dependent and linear independent set of vectors. Hence show that $S = ((1, 1, 1), (1, 1, 0), (1, 0, 0))$ is linearly independent in $\mathbb{R}^3$.
- d) Solve following system of equation.

$$x_1 + 2x_2 - 3x_3 - 4x_4 = 0$$

$$x_1 + 3x_2 - x_3 - 2x_4 = 0$$

$$2x_1 + 5x_2 - 2x_3 - 4x_4 = 0$$

Q5) Attempt any ONE of the following :

[1 × 15 = 15]

- a) i) Define spectral decomposition of a real symmetric matrix. Obtain the same for $A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$.
- ii) If A is an $n \times n$ matrix and λ is a real number, then λ is an eigenvalue of A if and only if $\det(\lambda I - A) = 0$.
- iii) If k is a positive integer, λ is an eigenvalue of a matrix A and X is a corresponding eigenvector then show that λ^k is an eigenvalue of A^k .

[7+4+4]

- b) i) Let $S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_r\}$ be a set of vectors in a vector space V . Then show that $L(S)$ is a subspace of V . Hence show that $L(S)$ is the smallest subspace of V containing S .
- ii) Define quadratic form. Hence find the maximum and minimum value of the quadratic form $Q = (X_1, X_2, X_3) = 3X_1^2 + 2X_2^2 + 3X_3^2 + 2X_1X_3$ subject to the constraint $X_1^2 + X_2^2 + X_3^2 = 1$.

[8+7]



Total No. of Questions : 5]

SEAT No. :

PE4816

[Total No. of Pages : 3

[6521]-13

First Year M.A./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) Define :

- i) Measurable space
- ii) Characteristic function

b) State Hogg and Craig Theorem.

c) State the inversion theorem of characteristic function.

d) Show that ϕ and Ω are always the members of the field $\mathcal{F}$.

e) The pdf of a random variable X is given by $f(x) = \begin{cases} \frac{1}{4}, & |x| \leq 1 \\ \frac{1}{4x^2}, & \text{Otherwise} \end{cases}$.

Find $P\left(-\frac{1}{2} \leq X \leq 2\right)$.

P.T.O.

Q2) Attempt any 3 question out of 4 questions.

[3×5=15]

- a) If $P(s)$ is the probability generating function of a non-negative integer valued random variable X , then find $\int_0^1 P(s^2) ds$.
- b) Let X is a RV having cdf as,
$$F(x) = \begin{cases} 0, & x < 0 \\ 1/3, & x = 0 \\ \frac{1+x}{3}, & 0 < x < 1 \\ 1, & x \geq 1 \end{cases}$$
 Find $E(X)$
- c) Let X be a random variable having $df \ f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, -\infty < x < \infty$
Obtain the distribution of X^2 .
- d) Define bivariate exponential distribution. State the memory less property satisfied by this distribution.

Q3) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Consider a random variable X with $df \ f(x) = \begin{cases} \theta e^{-\theta(x-a)} & x > a \\ 0, & \text{otherwise} \end{cases}$. Find $\xi_{\frac{1}{2}}(X)$. Also find $E \left| X - \xi_{\frac{1}{2}}(X) \right|$.
- b) Derive probability density function of Dirichlet distribution as a bivariate beta distribution.
- c) Let $X_1, X_2, \dots, X_n$ be a random sample from a population with continuous density function. Show that, $Y_1 = \min(X_1, X_2, \dots, X_n)$ is exponential with parameter λ if each X_i is exponential with parameter λ .
- d) Let X be a random variable of the continuous type with probability density function $f(x)$. Let $y = g(x)$ be differentiable for every x and either $g'(x) > 0$ or $g'(x) < 0$ for every x . Show that the probability density function of continuous random variable $y = g(x)$ is,

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

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

Q4) Attempt any 3 questions out of 4 questions.

[3×5=15]

- a) Define the cumulative hazard function $\Lambda(t)$ and explain how it is related to the hazard rate function $\lambda(t)$. Derive the survival function $S(t)$ in terms of the cumulative hazard function $\Lambda(t)$.
- b) Define non-central Student's t-random variable and obtain its density function.
- c) Prove that, two quadratic forms $Q_1 = X'AX$ and $Q_2 = X'BX$ are said to be independent if $AB = 0$.
- d) Let X and Y be independent and uniformly distributed over $[0,1]$. Derive the PDF of the sum $Z = X + Y$. Compute the convolution of their distributions and sketch the resulting PDF for Z .

Q5) Attempt any 1 questions out of 2 questions.

[1×15=15]

- a) i) Consider the following 3-dimensional mean vector and covariance

matrix. $\mu = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $\Sigma = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. Calculate the conditional

distribution of $X_1|X_2 = 2, X_3 = 3$. Hence, find the conditional mean and variance of X_1 . Also, find $P(X_1 > 2|X_2 = 2, X_3 = 3)$. **[8]**

- ii) Define truncated normal distribution. Derive its mean and variance. **[7]**
- b) i) State and prove Fisher Cochran's theorem. **[7]**
- ii) Show that normal distribution is a particular case of χ^2 distribution for 1 degrees of freedom. Also prove that for large degrees of freedom, χ^2 distribution tends to normal distribution. **[8]**

* * *

Total No. of Questions : 4]

SEAT No. :

PE4817

[Total No. of Pages : 2

[6521]-14

M.A./M.Sc. - I

STATISTICS

**STS- 510 -MJ: Optimization Techniques
(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 a scientific calculator is allowed.*
- 4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt all queations.

[5×1=5]

- a) Derive the general form of Non Linear Programming problem.
- b) What is meant by Stochastic Dynamic programming?
- c) Define the Transshipment problem.
- d) Explain the term Network Flow problems?
- e) Define spanning tree in network routine problem.

Q2) Attempt any two of the following.

[2×5=10]

- a) Discuss how to express the transshipment Problem as a networking problem?
- b) Explain the types of Integer Programming Problems with suitable examples.
- c) Describe the minimum cost flow problem.

Q3) Attempt any two of the following.

[2×5=10]

- a) Write the steps of Gomory's Mixed-Integer Programming Algorithm.
- b) Explain Beale's Method for solving Quadratic Programming Problems.
- c) Describe Bellman's Principle of Optimality and its significance.

P.T.O.

Q4) Attempt any one of the following.

[1×10=10]

- a) i) Find the Minimum Spanning Tree (MST) for the following network:

Edges	Cost
A-B	4
A-C	5
C-D	3
B-C	6
A-D	7
B-D	8

- ii) Define Quadratic Programming and write its standard form.

- b) i) Solve the following Integer Programming Problem using the Branch and Bound Method

$$\text{Maximize } z = 7x_1 + 9x_2$$

$$\text{Subject to: } -x_1 + 3x_2 \leq 6$$

$$7x_1 + x_2 \leq 35$$

$$x_2 \leq 7$$

$$x_1, x_2 \geq 0 \text{ and integer}$$

- ii) Write short notes on:

- 1) Chicken and Dog Puzzle
- 2) Three Glass Puzzle



Total No. of Questions : 4]

SEAT No. :

PE4818

[Total No. of Pages : 2

[6521]-15

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

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 each of the following.

[5×1=5]

- a) Define the term Average Run Length.
- b) A process is in control with $\bar{\bar{X}}=100$ and $\bar{S} = 16$. Compute the estimates of mean and standard deviation. (Take $C_4 = 0.94$).
- c) What are the disadvantages of implementing V-mask procedure?
- d) State the relationship between C_p and C_{pm} .
- e) Define the term acceptance sampling.

Q2) Attempt any Two of the following.

[2×5=10]

- a) Explain the working of tabular CUSUM chart for monitoring the process mean, when standard deviation is known.
- b) Write a short note on DMAIC procedure.
- c) Obtain the relation between number of defectives and capability index C_p .

Q3) Attempt any Two of the following.

[2×5=10]

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

P.T.O.

- b) A process is in control with $\bar{\bar{X}}=150$ and $\bar{s}=2.5$, $n=5$. The process specification are at 145 ± 10 . The quality characteristic has normal distribution. Compute C_p and C_{pk} . Interpret it.
- c) State the equivalence between sampling plans and testing of hypothesis problem.

Q4) Attempt any one of the following.

[1×10=10]

- a) i) Explain the need of sampling plans in Industry.
- ii) Find the parameters of V-mask CUSUM chart when tabular CUSUM chart have parameters $k = 0.5$ and $h = 4.77$ ($A=1$).

[7+3=10]

- b) The data in the following table represent individual observations on viscosity (x_i) taken hourly from a chemical process. The target value of molecular weight is 10 and the process standard deviation is $\sigma=1$. Set up an EWMA control chart for $\lambda = 0.10$ and $L=2.7$ and interpret from the plot.

period i	x_i	period i	x_i
1	9.45	7	8.04
2	7.99	8	11.46
3	9.29	9	9.2
4	11.66	10	10.34
5	12.16	11	9.03
6	10.18	12	11.47



Total No. of Questions : 4]

SEAT No. :

PE4819

[Total No. of Pages : 2

[6521]-16
F.Y.M.A./M.Sc.
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 indicates 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) Suppose force of interest per annum is 0.07 then find d and v .
- b) Explain the term annuity.
- c) Define loss at issue random variable.
- d) Show that, $\mu_x = \frac{f(x)}{1 - F(x)}$.
- e) Explain what the following represents:
 - i) $l_{[x]+r}$.
 - ii) ${}_n|m q_x$

Q2) Attempt any 2 questions out of 3 questions.

[2×5=10]

- a) Find the amount to which Rs. 25,000/- will be accumulated after 15 years if the rate of interest is
 - i) 6% as the force of interest
 - ii) 6% as the effective annual rate of interest
 - iii) 6% per annum convertible monthly
- b) Prove that $\bar{a}_x < \ddot{a}_x^{(m)} < \ddot{a}_x$, under the assumption of uniformity in an unity age interval.
- c) Suppose life length random variable X has a survival function,
 $F(x) = 1 - \frac{x}{120}$ if $0 \leq x \leq 100$. Find ${}_{3|7}q_7$.

P.T.O.

Q3) Attempt any 2 questions out of 3 questions.

[2×5=10]

- a) Define the future life time and curtate future life time random variable. Hence obtain its probability distribution and mass function.
- b) Compute the expressions for the variance of the prospective random variable in an n-year term insurance, n-year pure endowment insurance and n-year endowment insurance and whole life insurance in fully continuous case.
- c) For the reserve of a semi-continuous policy show that ${}_kV_x^{(m)} = {}_kV_x \left(1 + P_x^{(m)} \beta(m)\right)$. It is assume that the mortality pattern in each unit age interval is according to the uniform distribution.

Q4) Attempt any 1 question out of 2 questions.

[1×10=10]

- a) Consider the following mortality law,

Age	40	41	42	43	44	45	46	47	48	49
q_x	0.027	0.029	0.032	0.034	0.037	0.039	0.043	0.046	0.050	0.054

For this set up $i = 0.05$. Find the actuarial present value of 5-year temporary discrete life annuity and continuous life annuity and hence compare both present values.

- b) A person wants Rs. 2000/- at the end of each year for 12 years. The interest rates will vary over the years as follows: For the first 5 years: 5% per annum convertible semi-annually, for the next 4 years, 6% per annum convertible quarterly and for the remaining years: 7% per annum convertible daily. Calculate the present value of this annuity.

* * *

Total No. of Questions : 5]

SEAT No. :

PE-4820

[Total No. of Pages : 3

[6521]-17

M.Sc.

STATISTICS

STS-541-RM : Research Methodology
(2023 & 2024 Pattern) (Semester - I) (Credit 4)

Time : 3 Hours]

[Max. Marks : 70

Instructions to the condidates:

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

Q1) Attempt all questions :

[5 × 2 = 10]

- a) Define the term: i) Cartesian Product of Sets ii) Relation
- b) If $R = \{(x,y); x,y \in \mathbb{Z} \text{ } x^2 + 3y^2 \leq 8\}$ is a relation on set of integers $\mathbb{Z}$, then find the domain of R^{-1} .
- c) Define the term union of an indexed family of sets.
- d) Explain inverse transformation method.
- e) Give the definition of power set of a set with an example.

Q2) Attempt any 3 questions out of 4 questions :

[3 × 5 = 15]

- a) How to formulate research problem statistically?
- b) Discuss the difference between research objectives and the purpose of research in an academic study. How does each component contribute to the overall success of the research?
- c) Explain the differences between exploratory, descriptive, and explanatory research? Provide an example of each.
- d) What is the purpose of data visualization in research? How do you create effective data visualizations?

P.T.O.

Q3) Attempt any 3 questions out of 4 questions :

[3 × 5 = 15]

- a) If A , B , and C are subsets of some universal set U , then prove that $A \times (B \cap C) = (A \times B) \cap (A \times C)$.
- b) Let $n \in \mathbb{N}$ and let $a, b \in \mathbb{Z}$, then prove that $a \equiv b \pmod{n}$ if and only if a and b have the same remainder when divided by n .
- c) Let $R = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid |x| + |y| = 4\}$. Then R is a relation on $\mathbb{R}$. Is R an equivalence relation on $\mathbb{R}$? If not, is reflexive, transitive or symmetric?
- d) Assume the universal set is the set of real numbers. Let $A = \{-3, -2, 2, 3\}$, $B = \{x \in \mathbb{R} \mid x^2 = 4 \text{ or } x^2 = 9\}$, $C = \{x \in \mathbb{R} \mid x^2 + 2 = 0\}$ and $D = \{x \in \mathbb{R} \mid x > 0\}$.
 - i) Is the set A equal to the set B ?
 - ii) Is the set A subset of the set B ?
 - iii) Is the set C a subset of the set D ?
 - iv) Is the set A a subset of the set D ?

Q4) Attempt any 3 questions out of 4 questions :

[3 × 5 = 15]

- a) Describe the rejection method. Prove that the random variable X generated by the rejection method has density function f .
- b) Write an algorithm to draw a random sample from beta distribution with parameter (r, s) by using the relation with gamma distribution.
- c) Write an algorithm and R - code to simulate a random sample from $\chi^2(n)$ distribution.
- d) Let $X_1 \sim \text{Gamma}(2, 2)$ and $X_2 \sim \text{Gamma}(2, 4)$ be independent. Compare the histograms of the samples generated by the convolution $S = X_1 + X_2$ and the mixture $FX = 0.5Fx_1 + 0.5Fx_2$ by using R-code.

Q5) Attempt any 1 question out of 2 questions :

[1 × 15 = 15]

- a) i) Using truth tables prove that $P \vee (Q \wedge R) \equiv (P \vee Q) \wedge (P \vee R)$. [7]
- ii) Use the choose-an-element method to prove $A - (B \cap C) = (A - B) \cap (A - C)$. [8]
- b) i) Let $*$ be a binary operation on $Q - \{-1\}$ defined by $a * b = a + b + ab$ for all $a, b \in Q - \{-1\}$, then : [8]
- Show that $*$ is both commutative and associative on $Q - \{-1\}$.
 - Find the identity element in $Q - \{-1\}$.
 - Show that every element of $Q - \{-1\}$ is invertible.
 - Find inverse of an arbitrary element.
- ii) With the usual notations, for any predicate $P(x)$, prove that :
 $\neg (\forall x) [P(x)] \equiv (\exists x) [\neg P(x)]$ and $\neg (\exists x) [P(x)] \equiv (\forall x) [\neg P(x)]$. [7]



Total No. of Questions : 4]

SEAT No. :

PE4821

[Total No. of Pages : 2

[6521]-21
F.Y.M.A./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 the term completeness.
- b) Explain the concept of sufficiency.
- c) What factors affect the power of a statistical test?
- d) What is the definition of a “false discovery” in the context of FDR?
- e) State Cramer - Rao Lower Bound.

Q2) Attempt any two of the following.

[2×5=10]

- a) State and prove Rao-Blackwell Theorem.
- b) Let X_1, X_2 and X_3 is a random sample of size 3 from a population with mean μ and variance σ^2 . Let T_1, T_2 and T_3 are the estimators used to estimate mean value μ where $T_1 = X_1 + X_2 - X_3, T_2 = 2X_1 + 3X_3 - 4X_2$ and $T_3 = \frac{\lambda X_1 + X_2 + X_3}{3}$
 - i) Are T_1 and T_2 unbiased estimators?
 - ii) Find value of λ such that T_3 is unbiased estimator of μ .
- c) Let X_1, X_2 and X_3 be three independent observations on a random variable $X \sim B(1, \theta)$. Show that $T = X_1 + 2X_2 + 3X_3$ is not sufficient for θ .

P.T.O.

Q3) Attempt any two of the following.

[2×5=10]

- a) State and prove Basu's theorem.
- b) Define monotone likelihood ratio property of a distribution. Show that the binomial distribution satisfies MLR property.
- c) Suppose X is a single observation from $Geometric(p)$. Obtain MP test of size $\alpha = 0.05$ for testing $H_0 : \theta = \frac{1}{2}$ against $H_1 : \theta = \frac{1}{4}$. Also obtain power of the test.

Q4) Attempt any one of the following.

[1×10=10]

- a) i) State Neyman-Pearson Lemma. **[3]**
- ii) Given $f(x, \theta) = \frac{2(\theta - x)}{\theta^2}, 0 < x < \theta$, find UMP test (if they exist) for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ **[7]**
- b) i) Let $X_1, X_2, \dots, X_n$ be a random sample from $Geometric(p)$, find UMVUE for p . **[7]**
- ii) Discuss the significance of FDR in the context of multiple hypothesis testing. **[3]**



Total No. of Questions : 5]

SEAT No. :

PE-4822

[Total No. of Pages : 2

[6521]-22

M.Sc.

STATISTICS

**STS-552-MJ : Regression Analysis and Applications
(2024 Pattern) (Semester - II) (4 Credits)**

Time : 3 Hours]

[Max. Marks : 70

Instructions to the candidates:

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

Q1) Attempt all questions.

[5 × 2 = 10]

- a) Prove that, the hat matrix H and $(I - H)$ are idempotent.
- b) Explain any two criteria for the model adequacy.
- c) Obtain the least square estimators of the regression coefficients for centered model.
- d) If $e_2 = 2.0256$, $MS_{res} = 3.8625$ and $h_{22} = 0.0707$ then obtain the standardized residual (d_2) and state whether the related observation is outlier or not?
- e) In polynomial regression model, test the significance of the highest order term.

Q2) Attempt any 3 questions out of 4 questions.

[3 × 5 = 15]

- a) 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.
- b) Derive the likelihood ratio test for testing $H_0 : \underline{\beta} = \underline{\beta}_0$.
- c) Consider, the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$, with $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$ and ε are uncorrelated. Show that,
$$Cov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{x}\sigma^2 / S_{xx} \text{ and } Cov(\bar{y}, \hat{\beta}_1) = 0$$
- d) Discuss the fitting of polynomial regression of y on single regressor variable X using orthogonal polynomials. Hence, give the first three polynomials in X .

P.T.O.

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

- a) Discuss the Durbin-Watson test to detect the presence of autocorrelation in errors.
- b) Assuming regression model with two regressor variables, discuss the consequences of multicollinearity on $X'X$ matrix and on least square estimates of regressor parameters.
- c) Explain the measure of influence Cook's D-Statistic.
- d) Explain the development of the PRESS statistic.

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

- a) For the multiple linear regression model, explain the role of extra sum of squares principal in testing the relative importance of regressor variables. Interpret, the various components of regression sum of squares in case of model with three regressor variables.
- b) Derive the distribution of error sum of squares and regression sum of squares.
- c) Discuss Likelihood ratio test in logistic regression model.
- d) For the nonlinear regression model $y = \theta_1 e^{\theta_2 + \theta_3 x} + \varepsilon$, using iterative procedure obtain least squares normal equation.

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

- a)
 - i) Explain the terms: leverage points and influential points. Also, describe the importance of detecting influential observations. [6]
 - ii) State and Prove Gauss Markov Theorem. [6]
 - iii) Show that solution of normal equation actually minimizes the residual sum of squares. [3]
- b) Let $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, $i = 1 : k$ and $j = 1 : n_i$ be the one-way linear model.
 - i) Obtain the normal equation for, it and hence show that, its BLUE is unique almost surely. [7]
 - ii) Show that, $V(\hat{\underline{\beta}}) = S^{-1}\sigma^2$, $V(\underline{\lambda}'\hat{\underline{\beta}}) = \underline{\lambda}'S^{-1}\underline{\lambda}\sigma^2$ and
$$Cov(\underline{\lambda}'_{(1)}\hat{\underline{\beta}}, \underline{\lambda}'_{(2)}\hat{\underline{\beta}}) = \underline{\lambda}'_{(1)}S^{-1}\underline{\lambda}'_{(2)}\sigma^2.$$
 [5]
 - iii) Show that, covariance between any function belongs to error space and any BLUE is zero. [3]



Total No. of Questions :5]

SEAT No. :

PE4823

[Total No. of Pages : 4

[6521]-23

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

STATISTICS

STS-553-MJ: Multivariate Analysis and Applications
(2023 & 2024 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.

[5×2=10]

- a) Define Multivariate Normal Distribution (MVND). Write characteristics function of MVND.

- b) Let $\underline{X} \sim N_3(\underline{\mu}, \Sigma)$ with $\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ & 5 & 0 \\ & & 2 \end{bmatrix}$. Check whether following

random variables are independent.

- i) $\frac{X_1 + X_2}{2}$ and X_3
- ii) X_1 and $X_2 - \frac{5}{2}X_1 - X_3$
- c) Why Wishart distribution is called generalization of χ_n^2 distribution. Justify your answer.
- d) Write note on Mahalanobis D^2 statistic.
- e) If $\underline{X} \sim N_2(\underline{0}, \Sigma)$, where $\Sigma = \begin{bmatrix} 1 & 0.7 \\ & 1 \end{bmatrix}$ then find $P(X_1 < X_2)$.

P.T.O.

Q2) Attempt any three of the following.

[3×5=15]

- a) Let $\underline{X} \sim N_3(\underline{\mu}, \Sigma)$. A random sample of size $n = 5$ is taken and corresponding sample matrix is,

$$X = \begin{bmatrix} 9 & 2 & 6 & 5 & 8 \\ 12 & 8 & 6 & 4 & 10 \\ 3 & 4 & 0 & 2 & 1 \end{bmatrix}.$$

Compute

- i) Sample mean vector
 - ii) Sample dispersion matrix.
- b) If $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ and $\underline{X}$ is partition as $\begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$. Find the conditional distribution of $\underline{X}^{(2)}$ given $\underline{X}^{(1)} = \underline{x}^{(1)}$.
- c) Define the Hotelling- T^2 statistic. If $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ and $C'\underline{X}$ is a non-singular transformation of $\underline{X}$ then show that Hotelling - T^2 based on $\underline{X}$ and $C'\underline{X}$ are same.
- d) Write a note on Two Way MANOVA.

Q3) Attempt any three of the following.

[3×5=15]

- a) Let $\underline{X} \sim N_4(\underline{\mu}, \Sigma)$ where $\underline{\mu} = (1, 2, 1, 2)'$ and $\Sigma = 4I$. Derive the joint distribution of $Y_1 = 2X_1 + X_2 + 2X_3$ and $Y_2 = X_2 + 2X_3 + X_4$.
- b) Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ and $\underline{X}$ is partition as $\begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$, then obtain the marginal distributions of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$.
- c) Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$. obtain the maximum likelihood estimators of $\underline{\mu}$ and Σ .
- d) Define canonical correlations and canonical variables. Derive the first canonical correlation coefficient and the corresponding canonical variables. Also show that first canonical correlation explains maximum correlation.

Q4) Attempt any three of the following:

[3×5=15]

- a) Let $\underline{X} \sim N_3(\underline{\mu}, \Sigma)$ with $\underline{\mu} = [2 \ 3 \ -1]'$ and $\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ & 3 & 2 \\ & & 3 \end{bmatrix}$, then find the distribution of:

i) $\begin{bmatrix} 2X_1 - X_2 \\ X_1 + X_3 \end{bmatrix}$

ii) $\begin{bmatrix} X_1 - X_2 \\ X_1 + X_2 \\ X_1 + X_3 \end{bmatrix}$

- b) Let $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$ be a random sample of size n from $N_p(\underline{\mu}, \Sigma)$. Derive test statistic to test $H_0 : \underline{\mu} = \underline{\mu}_0$.

i) When Σ is unknown.

ii) When Σ is known.

- c) Obtain the population principal components for the following variance-covariance-matrix and also calculate the proportion of the total population variance explained by the first two principal component.

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

- d) Define orthogonal factor model (state assumption clearly). Explain the term communalities.

Q5) Attempt any one of the following:

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

1) $\frac{|D|}{|\Sigma|}$ is distributed as the product of p independent χ^2 variates with $n(n-1)\dots n-(p-1)$ d.f. respectively. [4]

2) $\frac{\sigma^{pp}}{d^{pp}} \sim X_{n-(p-1)}^2$, where d^{pp} and σ^{pp} are the last elements of D^{-1} and Σ^{-1} respectively. [4]

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

b) i) Explain following methods with example:

1) Average Linkage Method [4]

2) Hierarchical aggregative method [4]

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



Total No. of Questions : 4]

SEAT No. :

PE4824

[Total No. of Pages : 2

[6521]-24

M.A./M.Sc. - I

STATISTICS

**STS 560 MJ: Advances in Generalized Linear Models
(2024 Credit Pattern) (Semester - II)**

Time : 2Hours]

[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) List the key components of a Generalized Linear Model.
- b) Explain the terms concordance and discordance with respect to logistic regression.
- c) State the mean of a Poisson regression model.
- d) Define the Beta-Binomial model.
- e) Explain the normal linear mixed model (LMM).

Q2) Attempt any two 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) Explain the Quasi-Likelihood Method in statistical modeling. Discuss its advantages over traditional likelihood-based methods and give one example.

P.T.O.

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

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

- a) Compare Poisson and Negative Binomial regression models.
- b) Explain Logistic regression model for case control studies.
- c) Differentiate between likelihood ratio, Wald and Score tests.

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

- a) 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).

- b) Describe the use of GLMMs in modeling repeated binary responses.



Total No. of Questions : 4]

SEAT No. :

[Total No. of Pages : 2

PE4825

[6521]-25

M.A./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 incubation period and latent period.
- b) What is the difference between prevalence and incidence?
- c) State the basic assumptions of the SIR model.
- d) What is transmission intensity?
- e) What do you mean by flattening the curve?

Q2) Attempt any two of the following :

[2×5=10]

- a) Explain the concept of basic reproduction number (R_0) with an example.
- b) Discuss various types of disease occurrence patterns in epidemiology.
- c) Describe the Reed-Frost chain binomial epidemic model.

Q3) Attempt any two of the following :

[2×5=10]

- a) Discuss the structure and applications of the SEIR model.
- b) Explain how intervention strategies can be incorporated into epidemic models.
- c) Describe the concept of random networks for modeling epidemics.

P.T.O.

Q4) Attempt any one of the following :

[1×10=10]

- a) i) Develop and explain a mathematical model using the SIR framework for predicting COVID-19 spread. Include assumptions, parameters, and interpretations. **[7]**
- ii) Name any three types of epidemiological study designs. **[3]**
- b) i) Discuss in detail the applications of spatial epidemiology in understanding the spread and risk factors of H1N1 or COVID-19 with examples. **[7]**
- ii) What is the purpose of randomization in an RCT? **[3]**



Total No. of Questions : 4]

SEAT No. :

PE4826

[Total No. of Pages : 2

[6521]-26

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

STATISTICS

**STS - 562 - MJ : Discrete Data Analysis Linear Models
(2024 Credit Pattern) (NEP 2020) (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) What is meant by an interaction effect in a Generalized Linear Model (GLM)?
- b) Write down the loglinear model for a $2 \times 2 \times 2$ contingency table.
- c) Give confidence interval for binomial proportion.
- d) Explain the concept of conditional Odd's ratio.
- e) A call center receives an average of 4 calls per minute. What is the probability that in a given minute, exactly 5 calls are received?

Q2) Attempt any two of the following :

[5 each]

- a) What are graphical models in the context of log-linear analysis? Discuss their use in modeling and interpreting multi-way tables.
- b) Explain the effect of sparseness of χ^2 and G^2 .
- c) Write a short note on hierarchical log linear model.

P.T.O.

Q3) Attempt any two of the following :

[5 each]

- a) Write a short note on Fisher's exact test for 2×2 test.
- b) Briefly explain three components of GLM.
- c) Given a 3-way contingency table with variables A, B and C how would you check for conditional independence?

Q4) Attempt any one of the following :

[10 each]

- a) A coin is flipped twice. Let Y = number of heads obtained, when the probability of head for a flip equals π . Assuming $\pi = 0.50$, specify the probabilities for the possible values for Y , and find the distribution's mean and standard deviation. Find the bi-nomial probabilities for Y when π equals

i) 0.60 and

ii) 0.40.

Suppose $y = 1$ and π is unknown. Calculate and sketch likelihood function. Comment on it.

- b) i) Explain Simpson's Paradox phenomenon. **[4]**

ii) Derive the MLE method for logistic regression model. **[6]**

x x x

Total No. of Questions : 4]

SEAT No. :

PE-4827

[Total No. of Pages : 2

[6521]-31
M.A./M.Sc. - II
STATISTICS
STS-601-MJ : Probability Theory
(2024 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) Define the following terms.
(i) Cardinality of the SET (ii) Event
- b) Define convex function. Check the convexity of the function X^2
- c) Define monotone sequence with suitable example.
- d) What is independence of event for more than two events?
- e) Define Mixed Cumulative distribution functions of a random vector.

Q2) Attempt any TWO of the following :

[2 × 5 = 10]

- a) State and prove Borel-0-1 law.
- b) Define expectation of simple random variable If x and y are simple random variables then show that
 - i) $E(ax + by) = aE(x) + bE(y)$
 - ii) If $x > 0$ then $E(x) > 0$
- c) Let $(X_1, X_2, \dots, X_n)$ be a sequence of i.i.d. r.v. such that $\text{Uniform}(0, \theta)$ where $\theta > 0$ show that if $X(n) = \max(X_1, X_2, \dots, X_n)$ then $X_n \rightarrow \theta$ in probability.

P.T.O.

Q3) Attempt any TWO of the following :

[2 × 5 = 10]

- a) If $x_n \rightarrow x$ in probability $y_n \xrightarrow{P} y$, then show that $\frac{x_n}{y_n} \xrightarrow{P} \frac{x}{y}$, provided $y_n, y > 0$
- b) Show that $X_n \xrightarrow{P} c \Leftrightarrow X_n \xrightarrow{L} c$ If $F_n(x) \rightarrow F(x)$ where c is constant and $F(x) = \begin{cases} 0 & \text{if } x < c \\ 1 & \text{if } x \geq c \end{cases}$
- c) What is cumulative distribution function state its properties for bivariate random vector?

Q4) Attempt any ONE of the following :

[1 × 10 = 10]

- a) i) Show that $X_n \xrightarrow{P} 0$ if $E|X|^r \rightarrow 0, r > 0$
- ii) Write a note on independence of classes and independence of random variables.
- b) i) If $\{X_n\}_{n \geq 1}$ is a sequence of random variables then prove that $X_n \rightarrow 0$ in probability if and only if $E\left(\frac{|X_n|}{1+|X_n|}\right) \rightarrow 0$ in probability as $n \rightarrow \infty$
- ii) For random variable X and functions $g_1(x)$ and $g_2(x)$ show that $\{E[g_1(x)g_2(x)]\}^2 \leq \{E[g_1(x)]\}^2 \{E[g_2(x)]\}^2$



Total No. of Questions : 5]

SEAT No. :

PE-4828

[Total No. of Pages : 3

[6521]-32

M.Sc.

STATISTICS

STS - 602-MJ : Stochastic Processes

(2024 Pattern) (Semester - III) (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) Answer the following :

[5 × 2 = 10]

- a) Define the following terms:
 - i) Closed class of Markov chain
 - ii) Absorbing state
- b) What is a Markov property?
- c) Explain Non homogenous Markov chain
- d) State any two properties of Brownian motion
- e) Differentiate between a birth process and a death process.

Q2) Attempt any Three of the following :

[3 × 5 = 15]

- a) $N(t)$ is Poisson Process, derive its mean and variance.
- b) Define ergodic Markov chain and prove for an irreducible Markov chain there exist unique stationary distribution
- c) Define Doubly stochastic Matrix. Show that a markov chain having doubly stochastic matrix as its one step TPM has uniform distribution.
- d) Let a markov chain X_0, X_1, X_2 , has the TPM with state space $\{0,1,2\}$ is

$$\begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/4 \end{bmatrix}$$

and $P(X = i) = 1/3, i = 1,2,3$.

Find i) $P(X_3 = 1 | X_0 = 1)$

ii) $P(X_3 = 1, X_1 = 1, X_0 = 2)$

P.T.O.

Q3) Attempt any three of the following :

[3 × 5 = 15]

- a) State and prove chapman kolmogorov equation.
- b) Define persistent and transient states of a Markov chain. Show that being transient is a class property.
- c) Define Birth and death process and obtain its stationary distribution.
- d) If $N(t)$ denotes the number of renewals in the interval $[0, t]$, show that for the event $\{N(t) \geq n\}$, the process satisfies the renewal equation.

Q4) Attempt any three of the following :

[3 × 5 = 15]

- a) Given the transition probability matrix

$$P = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

- i) Determine whether the Markov chain is aperiodic or not, and find the period of each state.
 - ii) Check whether the chain is irreducible and doubly stochastic.
 - iii) Find the one-step transition probability from state 1 to state 3 and from state 2 to state 1.
- b) Define a weakly stationary and strictly stationary stochastic process. Show that if the process is strictly stationary, then it is also weakly stationary.
 - c) Explain Random walk markov chain, Show that random walk is persistent if $P = \frac{1}{2}$, otherwise it is transient.
 - d) Show that stationary distribution is unique.

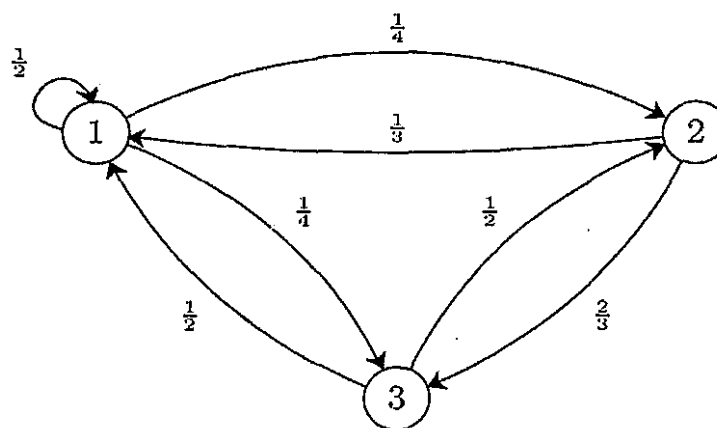
Q5) Attempt any one of the following :

[1 × 15 = 15]

- a) i) Consider the Markov chain with three states, $S = \{1, 2, 3\}$, that has the following transition matrix. **[7]**

$$P = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 0 & 2/3 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

- I) Draw the state transition diagram for this chain.
 II) What is the probability of transitioning from state 3 to state 2 in one step?
 III) Find the probability of remaining in state 1 after two steps
 IV) If we know $P(X_1 = 1) = P(X_1 = 2) = 1/4$, find $P(X_1 = 3, X_2 = 2, X_3 = 1)$
 V) Comment whether the chain is irreducible or not.
 VI) Determine whether state 2 is recurrent or transient.
- ii) State and prove addition property of Poisson process. **[8]**
- b) i) Define stationary distribution of Markov chain. Show that if π_1 and π_2 are two distinct stationary distributions then $\lambda\pi_1 + (1-\lambda)\pi_2$, $0 < \lambda < 1$ is also stationary distribution. State a condition under which stationary distribution is unique. **[8]**
- ii) Consider the Markov chain shown in Figure **[7]**



- I) Is this chain irreducible?
 II) Is this chain aperiodic?
 III) Find the stationary distribution for this chain.
 IV) Is the stationary distribution a limiting distribution for the chain?



Total No. of Questions : 5]

SEAT No. :

PE-4829

[Total No. of Pages : 3

[6521]-33

M.A./M.Sc. (Part - II)

STATISTICS

STS-603-MJ : Design and Analysis of Experiments
(2024 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) In a single factor ANOVA problem involving five treatments, with a random sample of four observations from each one, it is found that $SST_r = 16.1408$ and $SSE = 37.3801$. Then obtain the value of test statistic.
- b) Check whether following block design is connected :

Block - I Block - II

1	1
2	4
3	5

- c) Explain the linear and quadratic effect of factor.
- d) Define the following:
 - i) Treatment contrast
 - ii) Elementary treatment contrast
- e) Define connected design. How to check whether design is connected or not?

P.T.O.

Q2) Attempt any three of the following :

[3 × 5 = 15]

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

School I : 9 7 6 5 8

School II : 7 4 5 4 8

School III : 6 5 6 7 6

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

- b) Give the statistical analysis of 3^2 factorial experiment.
c) Distinguish among the resolution III, IV and V designs.
d) Write a note on Central Composite Design (CCD).

Q3) Attempt any three of the following :

[3 × 5 = 15]

- a) Explain Friedman Test.
b) Define BIBD, with usual notations prove that :
i) $vr = bk$
ii) $\lambda(v - 1) = r(k - 1)$
c) Given below are the incidence matrices N_A and N_B of design A and B :

$$N_A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 2 & 2 & 2 \end{bmatrix}, N_B = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 2 & 4 & 2 \end{bmatrix}$$

Check the orthogonality of design A and B.

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

Q4) Attempt any three of the following :

[3 × 5 = 15]

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

Blocks	Treatments		
1	1	2	5
2	1	3	6
3	1	4	7
4	8	9	10
5	6	7	10
6	3	4	10
7	2	3	8
8	2	4	9
9	5	7	9
10	5	6	8

- c) What is confounding? Distinguish between total and partial confounding.
- d) Confound AB^2C^2 in 3^3 factorial experiment in 3 blocks.

Q5) Attempt any one of the following :

[8 + 7 = 15]

- a)
 - i) Show that BIBD is connected and variance balanced design.
 - ii) Explain Taguchi design.
- b)
 - i) Explain PBIBD(m). Give the statistical analysis of PBIBD(2).
 - ii) What is Simplex Lattice Design. Explain a [3, 3] Simplex Lattice Design?



Total No. of Questions : 4]

SEAT No. :

PE-4830

[Total No. of Pages : 2

[6521]-34
M.Sc. - II
STATISTICS
STS-610-MJ : Survival Analysis
(2023 & 2024 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 each of the following question :

[1 Each]

- a) Explain Right random censoring with an illustration.
- b) Define mean residual life.
- c) Under which assumption log-rank test is most appropriate?
- d) Give one example of a distribution with an increasing hazard.
- e) What is minimal and perfect pair.

Q2) Attempt any two of the following :

[5 each]

- a) Explain the different types of censoring with suitable examples.
- b) Explain the procedure to obtain the maximum likelihood estimators of parameters of Gamma distribution for complete data.
- c) Explain the concept of the bathtub failure rate curve.

Q3) Attempt any two of the following :

[5 each]

- a) Prove the implications $NBU \Rightarrow NBUE$ and $DMRL \Rightarrow NBUE$.
- b) State and prove Cauchy functional equation.

P.T.O.

- c) For the following life table obtain the actuarial estimator of the survival function.

Interval in Year	No. Alive at Beginning of I_i	No. of Deaths During I_i	Number of Censored
0-4	20	2	1
5-9	17	1	2
10-14	14	1	4
15-19	9	1	3
20-24	5	1	4

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

- a) i) Define the Cox proportional hazards model. Derive the partial likelihood and explain its role in estimation. **[6]**
- ii) Prove that F is IFR if and only if equilibrium distribution function is concave function of t. **[4]**
- b) i) Discuss the Gamma life distribution and sketch the hazard curve of it. **[6]**
- ii) State the purpose of a Goodness-of-Fit test for a Cox proportional hazards (Cox PH) model. **[4]**



Total No. of Questions : 4]

SEAT No. :

PE-4831

[Total No. of Pages : 2

[6521]-35

M.Sc. (Part - II)

STATISTICS

STS-611-MJ : Asymptotic Inference

(2023 & 2024 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) Write the test statistic of Pearson's χ^2 test.
- b) State any one criterion to compare two consistent estimators.
- c) What is a Likelihood Ratio Test?
- d) Define BAN estimator
- e) State one parameter Canonical Form.

Q2) Attempt any TWO of the following :

[2 × 5 = 10]

- a) Define consistent estimation. Establish the invariance of consistent estimator under continuous transformation with proof.
- b) Find CAN estimator of θ when random sample of size n is drawn from $U(\theta, \theta + 1)$.
- c) Let $X_1, X_2, \dots, X_n$ is random sample from $U(\theta - 1, \theta + 1)$. Find maximum likelihood estimate of θ .

P.T.O.

Q3) Attempt any TWO of the following :

[2 × 5 = 10]

- a) Write short notes on Asymptotic relative efficiency
- b) Consider density function $f(x, \theta) = \theta x^{\theta} e^{-x^{\theta}}$; $x > 0$ and $\theta > 0$. Obtain CAN estimator for θ based on percentile method. Also obtain its asymptotic variance.
- c) Explain in detail locally most powerful test with example.

Q4) Attempt any ONE of the following :

[1 × 10 = 10]

- a)
 - i) Develop the likelihood ratio test and derive the asymptotic distribution of $-2\log\lambda$ [6]
 - ii) State asymptotic properties of hypothesis tests. [4]
- b)
 - i) State Cramer-Huzurbazar theorem. [3]
 - ii) Explain method of percentile to obtain CAN estimator. Also find CAN estimator of θ when random sample of size n is drawn from $(U(\theta, \theta + 1))$. [7]



Total No. of Questions : 4]

SEAT No. :

PE-4832

[Total No. of Pages : 3

[6521]-36

M.Sc. - II (Statistics)

STS-612-MJ : Machine Learning

(2023 & 2024 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) Define Recall in the context of classifier performance.
- b) What is a cluster centroid in K-means?
- c) Define complete linkage in agglomerative hierarchical clustering.
- d) What are common methods to handle missing data in Machine Learning?
- e) Define support in association rule mining.

Q2) Attempt any Two of the following :

[2 × 5 = 10]

- a) What is Machine Learning (ML)? Discuss its need in modern data-driven systems and state the main objectives of ML.
- b) How classifier performance is assessed using a confusion matrix, and what related measures can be derived from it?
- c) Consider the following transaction dataset :

Transaction ID	Items Bought
T1	Bread, Butter, Milk
T2	Bread, Butter
T3	Bread, Milk
T4	Butter, Milk
T5	Bread, Milk

Using the Apriori algorithm, identify all frequent itemsets and generate the corresponding association rules. Assume the minimum support = 40% and the minimum confidence - 60%.

P.T.O.

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

- a) Explain Hierarchical Clustering in Machine Learning. Describe its working principle, differentiate between Agglomerative and Divisive approaches, and discuss the role of distance measures and linkage methods.
- b) What is Natural Language Processing? Explain its objectives and describe the key components involved in text pre-processing such as tokenization, stop-word removal, and stemming
- c) A company wants to determine whether a customer will purchase a product ("Buy") based on the following attributes : Age (Young, Middle-aged, senior), Income (Low, Medium, High), and Student status (Yes or No). Apply the ID3 Decision Tree Algorithm to determine the root node.

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

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

- a) Solve the following
 - i) Explain the working principle of Support Vector Machine (SVM). Describe the concepts of hyperplane, margins, support vectors, and the role of kernel functions.

- ii) The following table contains 8 data points, each with two features X_1 and X_2 along with class label. Use the K- Nearest Neighbor (KNN) algorithm to classify the new point ($X_1 = 4$, $X_2 = 3$) with $K = 5$.

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

b) Solve the following

- Explain the concept of an Artificial Neural Network (ANN). Discuss its architecture, working mechanism, and the role of activation functions.
- Describe Sigmoid, Tanh, ReLU and Softmax activation functions with their characteristics.



Total No. of Questions : 5]

SEAT No. :

PE-4833

[Total No. of Pages : 3

[6521]-41

M.Sc. (Part - II)

STATISTICS

STS-651-MJ : TIME SERIES ANALYSIS

(2024 Pattern) (Semester - IV) (4 Credits)

Time : 3 Hours]

[Max. Marks : 70

Instructions to the candidates:

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

Q1) Attempt all questions :

[5 × 2 = 10]

- a) Write any two asymmetric extension models for GARCH with their one real life use.
- b) What are the causality and invertibility conditions of stationary time series model?
- c) Explain the general approach for time series modelling.
- d) How large sample theory helps ARIMA?
- e) State the test for trend and seasonality.

Q2) Attempt any three of the following :

[3 × 5 = 15]

- a) Define SARIMA (p, d, q) X (P, D, Q)_s model.
- b) Compute ψ_j & π_j coefficients for $j = 1, 2, \dots, 5$ for the following processes:
 - i) $X_t - 0.4X_{t-1} = Z_t + 0.5Z_{t-1}$, where $\{Z_t\} \sim WN(0, \sigma^2)$
 - ii) $X_t + 0.8X_{t-1} = Z_t + 1.5Z_{t-1}$, where $\{Z_t\} \sim WN(0, \sigma^2)$
- c) Explain exponential-Smoothing method. Apply it to estimate the trend for the following data with damping factor 0.2:

t	1	2	3	4	5	6
X_t	7	5	8	11	9	14

- d) Define transfer function model. Write its characteristics and write its anyone special case for the input X_t .

P.T.O.

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

- a) Derive the partial autocorrelation function (PACF) of a MA (1) process and its two important properties.
- b) Consider a filter with coefficients $[a_{-3}, a_{-2}, a_{-1}, a_0, a_1, a_2, a_3] = \left[\frac{1}{16}, \frac{-1}{8}, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}, \frac{-1}{8}, \frac{1}{16} \right]$ Show that this filter:
 - i) Passes a second-degree polynomial.
 - ii) Eliminates seasonal components with a period of 4.
- c) Explain the following tests for testing the estimated noise sequence:
 - i) The turning point test.
 - ii) The difference sign test.
- d) Determine which of the following processes are causal and which of them are invertible:
 - i) $X_t - 1.2 X_{t-1} + 0.5 X_{t-2} = Z_t + 0.3 Z_{t-1} + 0.2 Z_{t-2}$
 - ii) $X_t - 1.5 X_{t-1} + 0.7 X_{t-2} = Z_t + 0.4 Z_{t-1}$

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

- a) Describe the Augmented Dickey - Fuller (ADF) test in AR (p) model.
- b) Let Y_1, Y_2, Y_3 be the observations from MA(1) process. Find the coefficient 'b' such that estimate $\hat{Y}_3 = b Y_2$ minimizes $E(Y_3 - \hat{Y}_3)^2$.
- c) Let $\{X_t\}$ be a time series
 $X_t = Z_t + 0.5 Z_{t-2}$ where $\{Z_t\} = \text{WN}(0, 1)$.
 - i) Find ACVF
 - ii) Computer $V\left(\frac{1}{3} \sum_{i=1}^3 X_i\right)$
- d) Explain the concept of ARCH (1) model with its properties.

Q5) Attempt any one of the following :

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

A) Graphical Method

B) Moving-Average Method

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

t	1	2	3	4	5
X_t	7	9	4	11	6

- b) i) Describe any two model selection criteria that are used in ARIMA modeling. **[8]**

- ii) Define ARIMA model and discuss any one method for parameter estimation. **[7]**



Total No. of Questions : 5]

SEAT No. :

PE-4834

[Total No. of Pages : 3

[6521]-42

M.Sc.

STATISTICS

STS - 652-MJ : Sampling Theory and Applications
(2023 & 2024 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) Why sample survey methods provided over census method.
- b) Distinguish between SRSWR and SRSWOR.
- c) Explain Stratified Random Sampling. Also give a real life situation where Stratified random sampling is useful.
- d) Explain spatial smoothing.
- e) What is the role of time series in repeated sampling.

Q2) Attempt any three of the following :

[3 × 5 = 15]

- a) Under SRSWOR obtained an expression for variance of sample proportion.
- b) What is double sampling? How it is used in ratio and regression methods of estimation of population mean.
- c) Define Hansen-Hurwitz (HH) estimator $(\hat{Y}_{pps})$. Show that it is an u.b.e. of Y (Population total). Also show that $V(\hat{Y}_{pps}) = \frac{1}{n} \sum_{i=1}^N \left(\frac{Y_i}{p_i} - Y \right)^2 P_i$.
- d) Explain the method of robustness of design and predictors.

P.T.O.

Q3) Attempt any three of the following :

[3 × 5 = 15]

- a) Define Neyman's allocation. Under Neyman's allocation prove that $n_i \propto N_i S_i$. Also obtain an expression for sample size from i^{th} stratum.
- b) Under Systematic sampling obtain an expression for variance of sample mean in terms of intra class correlation coefficient. Compare systematic sampling and SRSWOR in terms of intra class correlation coefficient.
- c) Prove that the Horvitz-Thompson estimator $\hat{Y}_{HT} = \sum_{i=1}^n \frac{y_i}{\pi_i}$ is an u.b.e. of Y (Population total) with variance,

$$V(\hat{Y}_{HT}) = \sum_{i=1}^N \left(\frac{1 - \pi_i}{\pi_i} \right) Y_i^2 + \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) Y_i Y_j.$$

- d) Explain the method of sampling on successive occasions.

Q4) Attempt any three of the following :

[3 × 5 = 15]

- a) Explain Cluster Sampling. Obtain efficiency of single cluster sampling over SRSWR and SRSWOR.
- b) State the unbiased estimator of population mean in two stage sampling with equal number of first stage and second stage units. Derive the expression for its sampling variance.
- c) Define regression estimator $(\bar{Y}_{lr})$. In regression estimator let b is known constant then show that $\bar{Y}_{lr}$ is unbiased estimator of $\bar{Y}$. Obtain $V(\bar{Y}_{lr})$.
- d) Define Jack-Knife ratio estimator. Show that it is unbiased estimator for population mean.

Q5) Attempt any one of the following :

- a) i) Explain following methods to take PPSWR sample with example, [8]

I) Cumulative Total Method

II) Lahiri's Method

- ii) Prove that, under SRS the approximate value of Mean Square Error

of $\hat{R}$ is,
$$MSE(\hat{R}) = \frac{1}{X^2} [V(\bar{y}) + R^2 V(\bar{x}) - 2R\rho\sigma_{\bar{x}}\sigma_{\bar{y}}]$$

Where, ρ is the correlation coefficient between $\bar{x}$ and $\bar{y}$. [7]

- b) i) Explain Midzuno Sampling Design (MSD). Obtain single and double inclusion probabilities under MSD. [8]

- ii) Under Stratified random Sampling with cost function of the form C

$$= C_0 + \sum_{i=1}^k C_i \sqrt{n_i} .$$
 Prove that $n_i \propto \left(\frac{W_i^2 S_i^2}{C_i} \right)^{\frac{2}{3}}$ for minimising

variance subject to a fixed total cost. Also find out expression for n_i . [7]



Total No. of Questions : 4]

SEAT No. :

PE-4835

[Total No. of Pages : 2

[6521]-43

S.Y. MSc. (STATISTICS)

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

(2023 & 2024 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 are the limitations of the perceptron?
- b) Define different layers of Convolutional neural network.
- c) Which is the best algorithm of face detection?
- d) Define Batch Normalization.
- e) Write down any two drawbacks of RNN.

Q2) Attempt any Two of the following questions

[2 × 5 = 10]

- a) Explain LSTM (Long Short Term Memory). Give some applications of LSTM.
- b) What is unsupervised learning and how can deep neural network be used for unsupervised learning tasks?
- c) Write short notes on:
 - i) Shallow Network
 - ii) VGG Net

P.T.O.

Q3) Attempt any Two of the following questions

[2 × 5 = 10]

- a) Justify your answer, that how autoencoders are suitable compared to Principal Component Analysis (PCA) for dimensionality reduction.
- b) What is image segmentation and how is it used in computer vision applications?
- c) Explain Backpropagation with its algorithm.

Q4) Attempt any Two of the following questions

[2 × 5 = 10]

- a) How is automatic image captioning achieved using deep learning models and what are some potential applications of this technology?
- b) Discuss how batch gradient descent and Stochastic gradient descent are different.
- c) Explain the structure and learning process of a multilayer perceptron.



Total No. of Questions : 4]

SEAT No. :

PE-4836

[Total No. of Pages : 2

[6521]-44

M.Sc. (Part - II)

STATISTICS

STS-661-MJ : DESIGN AND ANALYSIS OF CLINICAL TRIALS

(2023 & 2024 Pattern) (Semester - IV) (2 Credits) (NEP-2020)

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 clinical end point.
- b) What is patient compliance?
- c) What is mean by blinding?
- d) State the goals of clinical trials.
- e) List out the primary pharmacokinetic parameters.

Q2) Attempt any two of the following :

[5 Each]

- a) What is a clinical trial protocol? Why is it important?
- b) List and briefly describe the four phases (Phase I-IV) of clinical trials.
- c) Explain the two-compartment model in pharmacokinetics with an example.

Q3) Attempt any two of the following :

[5 Each]

- a) What are the key differences between cross-sectional and longitudinal designs?
- b) State the null and alternative hypotheses for a 2×2 crossover design in bioequivalence testing.
- c) What are Case Report Forms (CRFs)? Discuss their role in data collection.

P.T.O.

Q4) Attempt any one of the following :

[10 Each]

- a) A 2×2 crossover bioequivalence study was conducted on 10 subjects comparing a Test (T) drug with a Reference (R) drug. The following are the log-transformed AUC (Area Under Curve) values for each subject in both periods.

Subject	1	2	3	4	5	6	7	8	9	10
Period 1	3.8	4.1	3.7	4.2	3.9	4.0	3.8	4.1	3.6	4.0
Period 2	3.9	4.0	3.6	4.0	3.8	3.9	3.7	4.2	3.7	4.1
Sequence	TR	TR	TR	TR	TR	RT	RT	RT	RT	RT

Estimate the mean difference in log AUC between Test and Reference drugs. Also, construct a 90% confidence interval for the difference. State whether the drugs are bioequivalent, assuming the equivalence range on log scale is (0.223, +0.223).

- b) Compare and contrast parallel design and crossover design in clinical trials.



Total No. of Questions : 4]

SEAT No. :

PE-4837

[Total No. of Pages : 2

[6521]-45
M.Sc. (Statistics)
STS-662-MJ: BAYESIAN INFERENCE
(2023 & 2024 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) Define non-informative prior.
- b) Distinguish between subjective probability and frequentist probability
- c) Define the loss function with an illustration.
- d) What is a credible interval?
- e) Define prior and posterior density.

Q2) Attempt any two of the following.

[5 Each]

- a) Define the squared error loss function and obtain the Bayes estimator under squared error loss function.
- b) If X follows the Poisson distribution with parameter λ and the prior distribution of λ is Gamma with parameters α and β . Find the posterior distribution of λ based on $Y = \sum_{i=1}^n X_i$, which is sufficient statistics for λ , where $X_1, X_2, \dots, X_n$ is a random sample of size n from the Poisson distribution.
- c) Explain Robust Bayesian analysis in briefly.

P.T.O.

Q3) Attempt any two of the following.

[5 Each]

- a) i) Let $X \sim N(\theta, 4)$ and assume that we are using squared error loss. Let $\hat{\theta} = x$. Find the expected loss (risk) for this estimator.
- ii) Let $X \sim B(n, \theta)$ and prior distribution of θ is $\beta(a, b)$. Find the posterior distribution of θ . Is given prior conjugate or not?
- 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 .
- c) Explain Bayes rule. A man is known to speak the truth two out of three times. He throws a die and reports that the number obtained is four. Find the probability that the number obtained is actually a four.

Q4) Attempt any one of the following.

[10 Each]

- a) Explain the term Bayes factor and Bayesian information criterion (BIC) with an illustration.
- b) What is the Gibbs sampling technique? Describe briefly.

