Ahmednagar Jilha Maratha Vidya Prasarak Samaj's

New Arts, Commerce and Science College, Ahmednagar (Autonomous)

(Affiliated to Savitribai Phule Pune University, Pune)

Choice Based Credit System (CBCS)

Framework for Syllabus

Master of Science (M. Sc.)

Implemented from

Academic year 2021 -22

M. Sc.(Statistics) –Part -I

DSCC – Discipline Specific Core Course,

DSEC- Discipline Specific Elective Course

GE- General Elective T- Theory

Course Code-

MSCST	Year (1 or 2)	Semester (1 to 4)	Course Number (1 to 6

Semester – I

Course Type	Course Code	Course Title	Credits
DSCC (T)	MSCST-111	Linear Algebra Using SCILAB	04
DSCC (T)	MSCST-112	Mathematical Analysis	04
DSCC (T)	MSCST-113	Bivariate Probability Distributions	04
DSEC (T)	MSCST-114	Multivariate Analysis Using R (A) / Sampling Theory and Methods (B)	04
GE (T)	MSCST-115	Modeling and Simulation	02
Practical	MSCST-116	Practical Using Computational Tools –I	04

Semester - II

Course Type	Course Code	Course Title	Credits
DSCC (T)	MSCST-121	Probability Theory	04
DSCC (T)	MSCST-122	Statistical Inference	04
DSCC (T)	MSCST-123	Regression Analysis Using R	04
DSEC (T)	MSCST-124	Time Series Using ITSM (A) / Categorical Data Analysis (B)	04
GE (T)	MSCST-125	Computer Intensive Operation Research	02
Practical	MSCST-126	Practical Using Computational Tools –II	04

Note: The End of Semester (EoS) Examination of following courses will be programming or software based

GE (T)	MSCST-115	Modeling and Simulation	02
Practical	MSCST-116	Practical Using Computational Tools –I	04
GE (T)	MSCST-125	Computer Intensive Operation Research	02
Practical	MSCST-126	Practical Using Computational Tools –II	04

M. Sc.(Statistics) –Part -II

Semester -III

Course Type	Course Code	Course Title	Credits
DSCC (T)	MSCST-231	Bayesian Inference using R	04
DSCC (T)	MSCST-232	Stochastic Process	04
DSCC (T)	MSCST-233	Decision Theory	04
DSEC (T)	MSCST-234	Machine Learning	04
GE (T)	MSCST-235	Image Processing	02
Practical	MSCST-236	Practical Using Computational Tools –III	04

Semester - IV

Course Type	Course Code	Course Title	Credits
DSCC (T)	MSCST-241	Design and Analysis of Experiments Using R	04
DSCC (T)	MSCST-242	Survival Analysis	04
DSCC (T)	MSCST-243	Applications of Statistics in Clinical Trials	04
DSEC (T)	MSCST-244	Non Parametric Inference Using R	04
GE (T)	MSCST-245	Practical Using Computational Tools –IV	04
Project	MSCST-246	Project	02

Ahmednagar Jilha Maratha Vidya Prasarak Samaj's

New Arts, Commerce and Science College, Ahmednagar

(Autonomous)

(Affiliated to Savitribai Phule Pune University, Pune)

Choice Based Credit System (CBCS)

Master of Science (M. Sc.) Statistics

Syllabus of

M. Sc. Part II (Semester-III & IV) Statistics

Implemented from

Academic year 2022 -23

Ahmednagar Jilha Maratha Vidya Prasarak Samaj's

New Arts, Commerce and Science College, Ahmednagar (Autonomous)

Board of Studies in Statistics

Sr. No.	Name	Designation
1	Prof. M.S. Kasture	Chairman
2	Dr. A. A. Kulkarni	Member
3	Dr. A.J. Shivagaje	Academic Council Nominee
4	Dr. A.K. Khamborkar	Academic Council Nominee
5	Prof. S. Kawale	Vice-Chancellor Nominee
6	Dr. S.B.Pathare	Alumni
7	Dr. V. P. Narkhede	Subject Expert, Special Nomination by Hon. Principal
7	Mr. Anirudha Deshmukh	Industry Expert
8	Prof. S.A Tarate	Member (co-opt)
9	Dr. S.D Jagtap	Member (co-opt)
10	Dr.N.T Shelke	Member (co-opt)
11	Dr. B.P. Thakur	Member (co-opt)

1. Prologue/ Introduction of the Programme:

It is known that in economic activities are of three types, agriculture, industrial and service. In the same way the subject Statistics is a SERVICE SCIENCE having potential to address the problems in these three fields. In research application of Statistics is mandatory. In the present days, apart from traditional field of career, Data Science, Data Analytics, Data Mining, Data Visualization are the upcoming field of career for Statistics students. In these field student must have mathematical ability, statistical thinking, computer (Software and programming) knowledge and communication (Verbal and written). These points are taken into consideration to design the syllabus and examination pattern of Statistics. In addition to academics, the department takes care to arrange a series of lectures on interview skills, preparation of CV, improve communication skill and overall personality development. The students are given the task of event management so that they can practice the principles of management such as leadership, creativity, communication, time management, group activity, team work, etc. In general, through curricular, co-curricular and extra-curricular activities student in three years is developed as thought provoker, problem solver, technologically sound, with command on communication, strong self-confidence.

M.Sc. in Statistics program is of two years' duration, with semester pattern. The important feature of the syllabus is that, all practical's form first year second year will be conducted on computer using R, Python SciLab, SPSS, ITSM programming and Tableau.

The another feature is at the in all courses, end of every chapter, self-learning activities are listed. These self-learning activities will play important role in creating interest in the subject and also boost their confidence. Further group activities will give the chance to explore their creativity and ideas. In addition, the verbal and written communication will be improved. These self-learning activities are expected to motivate students to participate in various student related academic events organized by home college or by other colleges too.

The course on Tableau will give an opportunity to learn thousands of various data presentation types and to present the complex data by easy way. The practical examinations of all courses will be on computer.

The syllabus is framed with appropriate weightage of theory, applied and skill enhancement courses. After receiving M.Sc. degree, student is expected to have minimum

knowledge of various courses and student will have ability to analyze the data with relevant interpretation of results.

2. **Programme outcomes (PoS)**

Students enrolled in the program complete a curriculum that exposes and trains students in a full range of essential skills and abilities. They will have the opportunity to master the following objectives.

I. Student will achieve the skill of understanding the big data handling

II. Student will have skill to write a story using data visualization.

III. Student will understand the interdisciplinary approach to correlate the statistical concepts with concepts in other subjects.

IV. Students will get good and unique combination of statistical concepts with computational tools which they use in feature

V. Student will be handle the real life example and industrial problems in project

VI. Students will demonstrate conceptual domain knowledge of the Statistics in an integrated manner.

VII. Student will get Sufficient knowledge to compete in any examination with great zeal like (NET / SET, UPSC, MPSC) etc.

I. Programme Structure and Course Titles

Sr. No.	Class	Semester	Course Code	Course Title	Credits
1.	M.Sc.	I	MSCST-111	Linear Algebra Using SciLab	04
2.	M.Sc.	I	MSCST-112	Mathematical Analysis	04
3.	M.Sc.	I	MSCST-113	Bivariate Probability Distributions	04
4.	M.Sc.	I	MSCST-114	Multivariate Analysis Using R (A) / Sampling Theory and Methods (B)	04
5.	M.Sc.	I	MSCST-115	Modeling and simulation	02
6.	M.Sc.	I	MSCST-116	Practical's Using Computational Tools-I	04
7.	M.Sc.	II	MSCST-121	Probability Theory	04
8.	M.Sc.	II	MSCST-122	Statistical Inference	04
9.	M.Sc.	II	MSCST-123	Regression Analysis Using R	04
10	M.Sc.	II	MSCST-124	Time Series Using ITSM (A) / Categorical Data Analysis (B)	04
11	M.Sc.	II	MSCST-125	Operations Using Software	02
12	M.Sc.	II	MSCST-126	Practical's Using Computational Tools-II	04
13	M.Sc.	III	MSCST-231	Bayesian Inference using R	04
14	M.Sc.	III	MSCST-232	Stochastic Process	04
15	M.Sc.	III	MSCST-233	Decision Theory	04
16	M.Sc.	III	MSCST-234	Machine Learning	04
17	M.Sc.	III	MSCST-235	Image Processing	02
18	M.Sc.	III	MSCST-236	Practical's Using Computational Tools-III	04
19	M.Sc.	IV	MSCST-241	Design of Experiments Using R	04
20	M.Sc.	IV	MSCST-242	Survival Analysis	04
21	M.Sc.	IV	MSCST-243	Applications of Statistics in Clinical Trials	04
22	M.Sc.	IV	MSCST-244	Non Parametric Inference using R	04
23	M.Sc.	IV	MSCST-245	Practical's Using Computational Tools-IV	04
24	M.Sc.	IV	MSCST-246	Project	02

Detailed syllabus of MSc. Part II Semester III & Semester IV Courses

.

Semester – III	Paper – I
Course Code: MSCST 231 T	Title of the Course: Bayesian Inference Using R
Credits: 04	Total Lectures: 60 Hrs.

Course Outcomes (COs):

a) On completion of this course students will get theoretical background about the Bayesian approach to data analysis (including choice of prior distributions and calculation of posterior distributions) with an emphasis on practical applications to inference problems in social and behavioural sciences.
b) Students will get insight about estimation methods in Bayesian setup such as point estimation, interval estimation.
c) Students will get insight about testing of hypothesis in Bayesian setup
d) course discusses some Bayesian computation techniques such as EM algorithm, Monte Carlo Sampling, Gibbs Sampling, Metropolis Algorithm and Metropolis- Hasting Algorithm

Detailed Syllabus:

Unit-I		**School of Bayesian Thought's**	15
	1.1	Introduction, difference between Frequentist and subjective approach, learning about the proportion of Heavy sleepers using a Discrete Prior, Informative and non-informative priors Beta Prior, Histogram Prior, drawbacks of Histogram approach for Prediction, Two Equivalent Ways of Using Bayes Theorem, Bayes Theorem for Binomial with Discrete Prior, Important Consequences of Bayes Theorem, Bayes' Theorem for Poisson with Discrete Prior, Relative likelihood approach, problems with relative likelihood approach, matching given functional form approach, CDF approach. applications of Bayesian inference in various fields, advantages of being Bayesian, Paradoxes in classical statistics, Summary of R Functions.	
	1.2	Classes of priors, goals while choosing class of priors, neighborhood class of priors, density ratio class of priors, conjugate class of priors, conjugate prior for the regular one exponential family distributions, conjugate prior for the multi-parameter exponential family distributions, summary of R function.	
	1.3	Conjugate prior for the non-exponential family members, Jeffery Prior, invariant property of Jeffry Prior, Jeffery priors for beta family, Poisson family, geometric family etc., Hierarchical Models.	
	1.4	R coding for different computations and data visualization in 1.1 to 1.3	
Unit-II		**Bayesian Inference in Statistical Analysis**	15
	2.1	Point estimator, Baye's estimator for all the three types of loss function. Improper priors, example of improper prior, Concept Maximum a Posterior (MAP) estimator, MAP estimator as mode of posterior distribution, difference between MAP estimator and maximum likelihood estimator, summary of R functions.	

	2.2	Credible set, Credible interval, Highest posterior density credible interval, examples on credible set and credible interval, Computation of credible interval and HPD interval using R software, difference between credible interval and HPD credible interval, testing null using credible interval, summary of R functions.	
	2.3	**Bayesian Inference for Binomial Proportion**: Using a Beta Prior, Choosing Your Prior, Summarizing the Posterior Distribution, Estimating the Proportion, Bayesian Credible Interval **Comparing Bayesian and Frequentist Inferences for Proportion:** Frequentist Interpretation of Probability and Parameters, Point Estimation, Comparing Estimators for Proportion, Interval Estimation, Hypothesis Testing, Testing a One-sided Hypothesis, Testing a Two-sided Hypothesis, Summary of R Functions.	
	2.4	R coding for different computations and data visualization in 2.1 to 2.3	
Unit-III		**Bayesian Inference about Mean**	15
	3.1	**Bayesian Inference for Normal Mean:** Bayes ' Theorem for Normal Mean with a Discrete Prior, Bayes Theorem for Normal Mean with a Continuous Prior, Choosing Your Normal Prior, Bayesian Credible Interval for Normal Mean, Predictive Density for Next Observation, summary of R functions.	
	3.2	**Comparing Bayesian and Frequentist Inferences for Mean**: Comparing Frequentist and Bayesian Point Estimators, Comparing Confidence and Credible Intervals for Mean, Testing a One-sided Hypothesis about a Normal Mean, Testing a Two-sided Hypothesis about a Normal Mean, summary of R functions.	
	3.3	**Bayesian Inference for Difference Between Means:** Independent Random Samples from Two Normal Distributions, Case I: Equal Variances, Case 2: Unequal Variances, Bayesian Inference for Difference Between Two Proportions Using Normal Approximation, Normal Random Samples from Paired Experiments, summary of R functions.	
	3.4	R coding for different computations and data visualization in 3.1 to 3.3	
Unit-IV		**Robustness, sensitivity, Model selection and Bayesian Computation**	15
	4.1	Measures of Robustness and sensitivity of priors, need of sensitivity to the choice of prior, belief function and plausibility function, measures of robustness using belief function, interactive robust Bayesian Analysis, Global measures of sensitivity, Local measure of sensitivity, Loss robustness, summary of R functions.	
	4.2	Model Selection Criteria: Bayesian model selection BIC, Bayes factors, limit of posterior distributions, consistency and asymptotic normality of posterior distribution, summary of R functions.	
	4.3	Bayesian Computation: Need of Bayesian computation techniques, motivating example for the development of Expectation Maximization algorithm, EM algorithm, A Beta-Binomial Distribution for over distribution, Approximations Based on Posterior Modes, Weal law of large numbers, Monte Carlo Sampling as alternative to numerical integration,	

		Rejection Sampling, resampling. Use of Monte Carlo sampling to solve complicated integrals in Bayesian Inference, need of Gibbs sampler, Gibbs Sampler Algorithm, Metropolis algorithm and Metropolis- Hasting Algorithm, difference between Metropolis and Metropolis- Hasting Algorithm, Summary of R functions.	
	4.4	R coding for different computations and data visualization in 4.1 to 4.3	

Suggested Readings Books:

1. Ghosh J.K., Delampady M. and T.Samantha(2006) An Introduction to Baysian Analysis :Theory and Methods, Springer.
2. James O Berger (1985), Statistical Decision theory and Bayesian Analysis, Springer Text in Statistics.
3. Bolstad W M(2007), Introduction to Baysian Statistics 2 nd edition Wiley.
4. Congdon P (2006), Baysian Statistical Modeling, Wiley.
5. Christensen R. Johnson W. Branscum A. and Hanson T. E. (2011). Baysian Ideas and Data Analysis: An introduction for Scientists and Statisticians, Chapman and Hall
6. Michael Baron, Bayesian Statistics Lecture Notes.

Semester – III	Paper – II
Course Code: MSCST 232 T	**Title of the Course: Stochastic Processes**
Credits: 04	**Total Lectures: 60 Hrs.**

Course Outcomes (COs):

a) Course is aimed with intension that students will get fundamental knowledge of concepts in stochastic processes.
b) Analytical thought process will be developed in such way that students can correlate real life problem to the concepts in the course.
c) Use of computational tools will accelerate the speed of solving complex problem.
d) R coding in the course will develop the logical thought process.

Unit-I		**Markov Chains (MC): Introduction**	15
	1.1	Definition, Transition probability matrices of MC, some MC models (An Inventory model, The Ehrenfest Urn model, MC in Genetics, A Discrete Queueing MC), First Step Analysis (Simple First Step Analyses, The General Absorbing MC)	
	1.2	Some Special MC (The Two-State MC, MC Defined by Independent random variables, One-Dimensional Random Walks, Success Runs), Functionals of Random Walks and Success Runs (The general Random Walk, Cash management, the success Runs MC)	
	1.3	Martingales: introduction, Definition & Examples, Properties of Martingales (Maximal inequality for non-negative martingales, Martingales convergence theorem, Optional Stopping theorem) continuous parameter martingales.	
	1.4	Study of different concepts in 1.1 to 1.3 using R coding.	
		Self learning: 1. Prepare a PPT on Transition probability matrices of MC 2. Write a summary on Three web pages of A Discrete Queueing MC 3. Write a summary on two videos of MC in Genetics 4. Prepare PPT on Properties of Martingales \ 5. Write a short note on functional of random walks and success runs.	
Unit-II		**Poisson Processes**	15
	2.1	The Poisson distribution and the Poisson Process (The Poisson distribution, the Poisson processes, Nonhomogeneous Processes, Cox processes), The Law of Rare Events (The law of rare events and the Poisson process, theorems related to rare events with proof)	
	2.2	Distributions Associated with the Poisson Process	

		The Uniform Distribution and Poisson Process(shot Noise, Sum Quota Sampling),Spatial Poisson Process, Compound and Marked Poisson Process (Compound Poisson Processes, Marked Poisson Processes)	
	2.3	Study of different concepts in 2.1 to 2.2 using R coding.	
		Self learning: 1. Write a summary on five web pages of Poisson Process. 2. Prepare a PPT on compound and Marked Poisson process. 3. Write a short note on Spatial Poisson Process. 4. Write a summary of Three videos on shot noise and sum quota sampling.	
Unit-III		**Continuous Time Markov Chain**	15
	3.1	Pure Birth Process: (Postulate for the Poisson process, Pure birth process, the Yule process), Pure Death Process: (The Linear Death Process, Cable Failure Under Static Fatigue) Birth and Death Processes(Postulates, Sojourn Times, Deferential Equations of Birth and Death process)	
	3.2	The Limiting Behaviour of Birth and Death processes, Birth and Death processes with absorbing states (Probability of Absorption into state 0, Mean time until absorption) Finite state continuous time Markov chain, A Poisson Process with a Markov Intensity	
	3.3	Study of different concepts in 3.1 to 3.2 using R coding.	
		Self learning: 1. Prepare PPT on Pure Birth Process 2. Write a summary on four web pages on Linear Death Process 3. Prepare a PPT on Limiting Behaviour of Birth and Death Process. 4. Write a summary on three videos on finite state continuous time Markov chain. 5. Write a summary on five web pages of Yule Process.	
Unit-IV		**Brownian Motion, Renewal and Branching Process**	15
	4.1	**Brownian Motion**: Wiener process as a limit of random walk; process derived from Brownian motion, Brownian motion and Gaussian processes (A little history, the Brownian motion stochastic process, the central limit theorem and the invariance principle, Gaussian processes) , The maximum variable and the reflection principle (The reflection principle, the time to first reach a level, the zeros of Brownian motion), Variations and Extensions (Reflected Brownian motion, Absorbed Brownian motion, The Brownian Bridge, Brownian meander).	

	4.2	**Renewal Processes:** Renewal process (Renewal process in discrete time, Relation between F(s) and P(s), renewal interval, generalised form: Delayed Recurrent event, renewal theory in discrete time), Renewal processes in continuous time (Renewal function and renewal density), Stopping time: Wald's equation (Stopping time, Wald's equation), Renewal theorems (Elementary renewal theorem, Applications, some definitions, renewal theorem (Blackwell's and Smith's), Central limit theorem for Renewals).	
	4.3	**Branching Processes:** Definition and examples branching processes, generating function, properties of generating function of Branching process, mean and variance, moments of X_n, Galton-Watson branching process(Branching processes with Immigration, processes in varying and random environments, multi type Galton-Watson process, controlled Galton-Watson process), probability of extinction(Asymptotic distribution of X_n, distribution of total number progeny, conditional limit laws critical processes and sub critical processes)	
	4.4	Study of different concepts in 4.1 to 4.3 using R coding.	
		Self learning: 1. Prepare PPT on Gaussian Process. 2. Prepare a PPT on Various Process in 4.1, 4.2, 4.3. 3. Write a short note on Galton-Watson Branching Process. 4. Write a summary on five web pages of Branching Process with immigration. 5. Write a summary of three videos on conditional limit laws critical processes.	

Suggested Readings Books:

1. **J. Medhi, Stochastic Processes, 4rd Edition 2017, New Age International.**
2. **Mark A. Pinsky Samuel Karlin, An introduction to Stochastic Modelling. ELSEVIER Publication 2011**
3. S.M. Ross, Stochastic Processes, 2nd Edition, Wiley, 1996.
4. S Karlin and H M Taylor, A First Course in Stochastic Processes, 2nd edition, Academic Press, 1975.
5. S.M. Ross. 11th edition by Academic Press in 2014. Introduction to Probability Models Some but not all chapters are covered.
6. S. Resnick. Birkhauser 1994. Adventures in Stochastic Processes
7. A. Loffler and L. Kruschwitz, ¨ Springer. 2019. The Brownian Motion: A rigorous but gentle introduction for economists.

Semester – III	Paper – III
Course Code: MSCST 233 T	Title of the Course: Decision Theory
Credits: 04	Total Lectures: 60 Hrs.

Course Outcomes (COs):

a. Course content will give the new perception of scientific approach in decision making.
b. Students will understand broad range of applications of decision theory.
c. Logical development of students will blend statistical, mathematical concepts to address real life issues in decision making.
d. Students will be next generation HR of data driven decision makers.

Detailed Syllabus:

Unit-I		**Foundation of Decision Theory – Part-I**	15 L
	1.1	Coherence – The "Dutch Book" theorem (Betting odds, Coherecne and the axioms of probability, Coherent conditional probabilities, the implications of Dutch Book theorems), Temporal coherence, Scoring rules and the axioms of probabilities.	
	1.2	St. Petersburg paradox, Expected utility theory and the theory of means (Utility and means, associative means, functional means), The expected utility principle, The von Neumann-Morgenstern representation theorem (Axioms, Representation of preferences via expected utility), Allais' criticism.	
	1.3	Utility in action – The "standard gamble", Utility of money (Certainty equivalents, Risk aversion, A measure of risk aversion), Utility functions for medical decisions (Length and quality of life, Standard gamble for health states, The time trade-off methods, Utilities for time in ill health, Difficulties in assessing utility)	
		Self Learning Activity: 1.Write summary of FIVE web pages on Butch Book theorem. 2.Write summary of THREE videos on St. Petersburg paradox. 3.Prepare PPT on von Neumann-Morgenstern representation theorem. 4.Prepare PPT on applications of utility functions in different field.	
Unit-II		**Foundation of Decision Theory – Part-II**	
	2.1	Ramsey's theory, Savage's theory (Notation and overview, the sure thing principle, Conditional and a posterior preference, Subjectivity probability, Utility and expected utility, Ellsberg paradox.	
	2.2	State Independence – Horse lotteries, State-dependent utilities, State-independent utilities, Anscombe-Aumann representation theorem	
		Self Learning Activity: 1. Write summary of FIVE web pages on Ramsey's theory. 2. Write summary of THREE videos on Savage's theory. 3. Prepare PPT on Ellsberg paradox. 4. Prepare PPT on Anscombe-Aumann representation theorem	
Unit-III		**Statistical Decision Theory Part-I**	

	3.1	Decision functions – Basic concepts (The loss function, Minimax, Expected utility principle, illustrations), Data-based decisions (Risk, Optimality principles, Rationality and the Likelihood Principle, Nuisance parameters)	
	3.2	The travel insurance problem, Randomized decision rules, Classification and hypothesis tests (Hypothesis testing, Multiple hypothesis testing, Classification), Minimax-Bayes connections	
		Self Learning Activity: 1. Prepare PPT on Randomized decision rules. 2. Prepare PPT on travel insurance problem. 3. Write summary to illustrate Minimax-Bayes connections. 4. Prepare PPT on Decision functions.	
Unit-IV		**Statistical Decision Theory Part-II**	15 L
	4.1	Admissibility – Admissibility and completeness, Admissibility and minimax, Admissibility and Bayes (Proper Bayes rules, Generlised Bayes rules), Complete classes (Completeness and Bayes, Sufficiency and the Rao-Blackwell inequality, The Neyamnn-Person lemma, Using the same α level across studies with different sample sizes is inadmissible.	
	4.2	Scoring rules – Betting and forecasting, scoring rules (Definition, proper scoring rules, the quadratic scoring rules, scoring rules that are not proper, Local scoring rules, Calibration and refinement (The well-calibrated forecaster, Are Bayesians well calibrated?)	
		Self Learning Activity: 1. Write summary of FIVE web pages on Admissibility and completeness. 2. Write summary of THREE videos on Admissibility and Bayes. 3. Prepare PPT on Completeness and Bayes. 4. Prepare PPT on travel insurance problem. 5. Write summary to illustrate "Using the same α level across studies with different sample sizes is inadmissible". 6. Prepare PPT on Betting and forecasting. 7. Write summary of FIVE web pages on Scoring rules 8. Prepare PPT on Calibration and refinement	

Suggested Readings Books:

1. **Decision Theory – Principles and Approaches, Giovanni Parmigiani and Lurdes Inoue (2009), Publisher- John Wiley and Sons Ltd.**
   https://www.webdepot.umontreal.ca/Usagers/perronf/MonDepotPublic/stt2100/Decision_theory.pdf
2. Optimal Statistical Decisions, Morris H. DeGroot, (1970), McGraw-Hill Book Company
3. https://people.kth.se/~soh/decisiontheory.pdf
4. http://probability.ca/jeff/ftpdir/anjali0.pdf
5. http://brian.weatherson.org/DTBook-15.pdf

Semester – III	Paper – IV
Course Code: MSCST 234 T	Title of the Course: Machine Learning
Credits: 04	Total Lectures: 60 Hrs.

Course Outcomes (COs):

e. Contents in the course will give insight of latest concepts applicable in Data Science.
f. Students will get an opportunity to analyses big data in terms of features and instances.
g. Students can blend artistic thought in scientific temperament to develop the algorithm in order to make sense of data.
h. Students will perform better analytics of real time data in internship (if any).

Detailed Syllabus:

Unit-I		**Data Cleaning, Data Wrangling and Data Transformation**	**15 L**
	1.1	Data Quality – Accuracy, Completeness, Consistency, Timelines, Believability and Interpretability. Major Tasks in Data Processing – Data Integration, Data Reduction, Dimensionality Reduction, Numerosity Reduction,	
	1.2	Data Cleaning – Missing Values (Ignore the tuple, filling the missing values manually, use of global constant to fill in the missing value, use of measure of central value of attribute, use the most probable value to fill in the missing value.), Noisy Data (Binning, Regression and outlier analysis), Data Cleaning as a process (Field overloading, Data Scrubbing tools, Data Auditing tools, Data Migration tools, Meta Data)	
	1.3	Data Transformation (Data Transformation by Normalization, Discretization by binning, Discretization by Histogram analysis, Discretization by cluster; decision tree and correlation analysis, concept hierarchy generation for nominal data)	
	1.4	R / Python Code for to study the quality of data of certain data set R / Python Code for to replace missing value by different method & study the change in data quality. R / Python Code for to apply transformation by different approaches& study change in data quality.	
		Self Learning Activity: 1. Select a data from Online Data Platforms & Discuss it. 2. Select a data from Online Data Platform & Discuss the Possibility of data cleaning. 3.Write a summary of five Web Pages on Data Quality. 4.Prepare Power Point Presentation on Data Processing. 5. Write summary of selected videos on data transformation.	
Unit-II		**Supervised Learning**	**15 L**

	2.1	Concept of Supervised learning, training data and test data, Separation and Classification, Support Vector Machine (Maximum margin hyper plane, linear SVM for both separable and non-separable cases, non-linear SVM- different kernels , Characteristic of SVM), k-Nearest Neighbor Classifier (Algorithm & Characteristic of k-NN) Difference between Regression verses Classification and Inference verses Prediction	
	2.2	Decision Tree (Working of Decision Tree, building a Decision Tree, Attribute Test Condition, Selecting the best Split, Algorithm for Decision Tree Induction, Characteristics of Decision Tree Induction) Random Forest (Bias-Variance Decomposition, Randomized Induction Algorithms, Propoerties and Features, Out-of-bag estimates, Variable importance, Proximity measures, missing data and Consistency) Naïve Bayes (Use of Bayes theorem for classification, Naive Bayes Classifier, Bias Error rate, Bayesian Belief Network)	
	2.3	R / Python Code for Linear SVM R / Python Code Nonlinear SVM R / Python Code for K-NN R / Python Code for Decision Tree R / Python Code Random Forest R / Python Code for Naïve Bayes Classifier Use Sufficiently Big- Data to test the above Code.	
		Self Learning Activity: **1.** Write in detail various applications of different classifiers. 2. Write a summary of five web pages on Random Forest. 3.Write a summary of selected videos on Naïve Bayes classifier. 4. Write comparative account of different classifier. 5.Prepare PPT to illustrate one of the application of classifier. 6.Write R code to study the effect of size of training data on accuracy.	
Unit-III		**Unsupervised Learning**	**15 L**
	3.1	Concept of Unsupervised learning, Clustering – (characteristics, Characteristics of Clustering Algorithms, Prototype Based Clustering, Density Based Clustering, Graph Based Clustering, Scalable Clustering Algorithms, Grid Search based method), Evaluation of cluster quality (Assessing Clustering Tendency, Determining number of clusters, measuring cluster quality)	
	3.2	Market Basket Analysis (Association Analysis)- Concept of Association, Frequent Itemset Generation (Apriori Principle, Apriori Algorithms, Candidate Generation and Pruning, Support Counting, Computational Complexity) Rule Generation-(confidence Based Pruning)	

		Evaluation of Association rule (Support, Confidence, Gini Index, Lift, J-measures, Φ-Coefficient, Odds ratio, cosine, Interest, kappa) Compact representation of frequent Itemset (maximal & Close frequent Itemset) FP Growth Algorithm (FP Tree Representation, Frequent Itemset Generation in FP Growth Algorithm)	
	3.3	R / Python Code for Prototype Based Clustering R / Python Code for Density Based Clustering R / Python Code for Graph Based Clustering R / Python Code for Scalable Clustering R / Python Code for Grid Search based Clustering R / Python Code for number of clusters & cluster quality R / Python Code for Frequent Item Generation R / Python Code for Confidence Based Pruning R / Python Code for Evaluation of Association Rule R / Python Code for FP Growth Algorithm R / Python Code for FP Tree Representation	
		Self Learning Activity: 1. Write in detail various applications of different clustering methods. 2. Write a summary of five web pages on any one clustering method. 3.Write a summary of selected videos on any one clustering method. 4. Write comparative account of different clustering method. 5.Prepare PPT to illustrate one of the application of numbers of clusters & clusters quality. 6. Write in detail various applications of association rule. 2. Write a summary of five web pages on frequent item generation. 7.Write a summary of selected videos on evaluation of association rule. 8. On suitable data apply FP algorithm and evaluate association rule with appropriate interpretation by different measures. 9.Prepare PPT to illustrate one of the application of FP Tree Representation.	
Unit-IV		**Reinforcement Learning & Artificial Neural Network**	**15 L**
	4.1	Introduction The Learning Tasks, Q Learning (Q function, Algorithm, Convergence, Experimentation Strategies, updating sequence) Non Deterministic Rewards & Actions, Generalizing from examples, Relationship to dynamic programming, Temporal-Difference Learning: (TD Prediction, Advantages of TD Prediction Methods, Optimality of TD (0), Sarsa: On-Policy TD Control, Q-Learning: Off-Policy TD Control, Games, Afterstates, and Other Special Cases	

	4.2	Artificial Neural Network Introduction, Biological Motivation, NN representation, Perceptron – Representational power of Perceptron, Perceptron Training rule, Gradient Descent and Delta Rule. Multilayer Networks and Back Propagation algorithm, Differentiable threshold units, Back Propagation Algorithm, Derivation of Back Propagation Rule.	
	4.3	R / Python Code for Q function & Q Algorithm R / Python Code for Convergence of Q Algorithm R / Python Code for TD Prediction R / Python Code for NN Representation R / Python Code for Backpropagation Algorithm	
		Self Learning Activity: 1.Write in detail various applications of Reinforcement Learning. 2. Write in detail various applications of Artificial Neural Network. 3.Write a summary of five web pages on Reinforcement Learning or Artificial Neural Network. 3.Write a summary of selected videos on Reinforcement Learning. 4.Write a summary of selected videos on Artificial Neural Network. 5.Prepare PPT to illustrate one of the application of Reinforcement Learning. 6.Prepare PPT to illustrate one of the application of Artificial Neural Network.	

Suggested Readings Books:

1. Data Mining Concept and Techniques (Jaiwei Han /Micheline Kamber / Jian Pei).
2. Introduction to Data Mining (Pang-nine tan Michel Steinbach vipin kumar).
3. Machine Learning in Action (Peter Harrington).
4. Machine Learning (Tom M. Mitchell).
5. Monetizing Machine Learning (Manuel Amunategui / Mehdi Roopaei).
6. Scaling up Machine Learning (Parallel & Distributed Approaches) Ron Bekkerman / Mikhil Bilenko / John Langford
7. Machine Learning (The Arts and Science of Algorithms that make sense of data. (Peter Flach)

Semester – III	Paper – V
Course Code: MSCST-235 T	**Title of the Course: Image Processing**
Credits: 02	**Total Lectures: 30 Hrs.**

Course Outcomes (COs):

a. On completion of this course student will have different perception about Digital Image Processing than a layman.
b. Students will understand the applications of basic concepts in image processing.
c. Students will understand the science behind GUI interface.
d. Students may undertake projects in Data science.

Detailed Syllabus:

Unit-I		Fundamental Concepts in image processing	15 L
	1.1	Concept of digital image processing, origin of digital image processing, applications of digital image processing, gamma ray imaging, X-ray imaging, imaging in ultra violet band, imaging in visible and infra-red band, imaging in microwave band, imaging in the radio band.	
	1.2	Fundamentals steps in digital image processing- Image acquisition, image enhancement (Contrast manipulation, histogram modification, noise cleaning, edge crispening), image restoration, colour image processing, wavelets, compression, morphological processing (Binary image connectivity, binary image heat or Miss transformations, binary image shrinking, thinning, thickening, skeletonizing, grace scale image morphological operations).	
	1.3	segmentation (Amplitude segmentation, clustering segmentation, region segmentation, boundary segmentation, texture segmentation and segment labelling), representation and description, recognition.	
Unit-II		Mathematics for image processing	15 L
	2.1	2-D Fourier transform, properties of Fourier transform (special frequencies, uniqueness, seperability, frequency response and Eigen functions of shift invariant systems, convolution theorem, inner product preservation). Some more properties of 2-D Fourier transform (Rotation, Linearity, Conjugation, Seperability, Scaling, Shifting, Modulation, Convolution, Multiplication, Spatial correlation and Inner product).	
	2.2	Fourier Series (Fourier transform of sequences)- 1-D and 2-D sequence, properties of Fourier transform of 2-D sequence (Linearity, Conjugation, Seperability, Shifting, Modulation, Convolution, Multiplication, Spatial correlation and Inner product).	

	2.3	Z Transform, region of convergence, transfer function, causality and stability, properties of 2-D Z transform (Rotation, Linearity, Conjugation, Seperabability, Shifting, Modulation, Convolution, Multiplication).	
	2.4	Hands on Experience: 1. Study of Rotation, Linearity, Conjugation, Seperability, Scaling, Shifting, Modulation, Convolution, Multiplication, Spatial correlation and Inner Product Using 2-D Fourier Transform. 2. Study of Linearity, Conjugation, Seperabability, Shifting, Modulation, Convolution, Multiplication, Spatial correlation and Inner product using Fourier Series. 3. Study of Rotation, Linearity, Conjugation, Seperabability, Shifting, Modulation, Convolution, Multiplication using Z Trnasform.	

Suggested Readings Books:

1. **Fundamentals of Digital Image Processing: Anil K. Jain.**
2. Digital Image Processing: William K. Pratt.
3. Digital Image Processing: Rafael C. Gonzalez and Richard E. Woods.
4. Digital Image Processing: Walter G. Kropatsch and Horst Bischof.

Semester – III	Paper – VI
Course Code: MSCST 236 P	Title of the Course: Practical's Based on Computational Tools-III
Credits: 04	Total Lectures: 60 Hrs.

List of Practical All practicals are based on R Coding

Course Outcomes (Cos):

1. On learning different computational tools students will have capacity to compare the features each of them with others.
2. Students can perform statistical analysis by different computational tools so as to know various parts of output.
3. On working with big data students will get confidence and competence.
4. Students will build logical thinking ability to develop statistical utilities.
5. Students will get an exposure of interacting with different platforms.

Practical's based – MSCST-231 T Bayesian Inference Using R

1) Study of properties and plotting of Prior, conjugate priors (exponential and non-exponential family) , Jeffry Prior for beta family, Poisson family and geometric family, posterior density functions..
2) Inference about Mean of Normal distribution, and difference between means of two Normal populations. Both point estimation and confidence intervals. Comparison of these two types of intervals.
3) Study of Sensitivity (Global and local), Model development and model selection criterion, limiting distribution of posterior distribution.
4) On EM algorithm, WLLN, rejection sampling, Gibbs sampling algorithm, Hasting Algorithm.

Practical's based – MSCST-232 T Stochastic Processes

5) Realization of Markov Chain and construction of TPP for An Inventory model, The Ehrenfest Urn model, MC in Genetics, A Discrete Queueing MC. Stationary distribution of Markov chain and classification of its states. Gambler's Ruin problem.
6) Realization of Poisson process (Nonhomogeneous Processes, Cox processes, Spatial Poisson Process, Compound and Marked Poisson Process).

7) Realization of birth and death process. (Pure birth process, the Yule process, Linear Death Process, Cable Failure Under Static Fatigue) (With constant birth and death rates)
8) Realization of Gaussian, Brownian process and Branching Process.

Practical's based – MSCST-233 T Decision Theory

9) Study of decision functions – Loss function, Minimax, Expected Utility.
10) Study of randomized decision rules, hypothesis testing, multiple hypothesis testing.
11) Study of Admissibility, minimax, Bayes, generalized Bayes.

Practical's based – MSCST-234 T Machine Learning

12) Data Cleaning, Imputation, Data transformation & Quality of Data. (Unit I, 1.1 to 1.3)
13) Classification of Data using k-NN, Naïve Bayes Classifier and SVM (Linear & Nonlinear).
14) Classification of Data using Random Forest & Decision Tree.
15) Clustering by different methods & comparison of clustering algorithms. (Unit III, 3.1)
16) Market Basket analysis (Frequent Item Generation, Apriori algorithm & Confidence Based Pruning. Evaluation of Association Rule by different measures.) (Unit III, 3.2)
17) FP Growth Algorithm & FP Tree Representation. (Unit III, 3.2)
18) Reinforcement Learning (Unit-IV, 4.1)
19) Artificial Neural Network – Representation, Perceptron, Perceptron training rules (Unit-IV, 4.2)
20) Artificial Neural Network – Multilayer Networks, Back Propagation algorithm (Unit-IV, 4.2)

The End of Semester Examination of following courses will Computational Tool Intensive Examination. It is based on slips that will be prepared according to the guidelines given by the Subcommittee constituted under the Chairmanship of Prof. S. A. Tarate and as per the minutes of subcommittee meeting held on 09/10/2021.

Semester – III	**Paper – V**
Course Code: MSC-ST 235 T	**Title of the Course: Image Processing**
Credits: 02	**Total Lectures: 30 Hrs.**
Semester – III	**Paper – VI**
Course Code: MSC-ST 236 P	**Title of the Course: Practical's Based on Computational Tools-III**
Credits: 04	**Total Lectures: 60 Hrs.**
Semester – IV	**Paper – V**
Course Code: MSc-ST 245 P	**Title of the Course: Practical's Based on Computational Tools-IV**
Credits: 04	**Total Lectures: 60 Hrs.**

Semester – IV	Paper – I
Course Code: MSCST 241 T	**Title of the Course: Design and Analysis of Experiments using R**
Credits: 04	**Total Lectures: 60 Hrs.**

Course Outcomes (COs):

a) Understand the concept of BIBD, connectedness, balancedness and orthogonality of design.
b) Understand the difference between fixed and random effect models.
c) Compare the pairs of treatment means using different methods. Construct fractional factorial experiments and apply confounding in real life problems. CO4:
d) Apply the response surface methodology on real life examples. Construct the Taguchi design.

Detailed Syllabus:

Unit I		Factorial Designs	15 L
	1.1	Introduction, Classical One at a Time versus Factorial Plans, Interpreting Interactions, Creating a Two-Factor Factorial Plan in R, Analysis of a Two-Factor Factorial in R (Matrix Representation of Model and Analysis, Determining the Number of Replicates, Analysis with an Unequal Number of Replicates per Cell, Testing for Interaction with One Replicate per Cell).	
	1.2	Full factorial experiments: Review of 2^2 and 2^3 factorial experiment. Geometric view of 2^2 and 2^3 factorial experiment. 2^K factorial experiments. main effects and interaction effect ANOVA, model analysis using ANOVA, total confounding of 2^k design in 2^p blocks $p \geq 2$, methods to from blocks I) Linear contrast method and method using principal, partial confounding in 2p blocks; p = 2, 3. Factorial Designs with Multiple Factors—CRFD, Two-Level Factorials (Main Effects and Regression Slopes, Interactions, Shortcut Formula for Determining the Number of Replicates, Analysis with One Replicate per Cell), Verifying Assumptions of the Model.	
	1.3	Random effect model for the two factor, estimation of variance components in two factor random effect model, two factor mixed effect model, Estimation of variance component in the model, sample size determination with random effect model.	
	1.4	Creating and analyzing various designs in 1.1 and 1.3 using R code.	
		Self-Learning Activity: 1. Write a summary of five web pages of Two Factor Factorial Design 2. Write a summary on determining number of replicates. 3. Prepare a PPT on Testing for interaction on one replicate per cell. 4. Write a summary of three videos on factorial design with multiple factors.	

		5. Write a note on verification assumption of a model.	
Unit II		Fractional Factorial Designs	15 L
	2.1	Introduction, Half-Fractions of 2^k Designs, Quarter and Higher Fractions of 2^k Designs, Quarter fractions of 2^{k-2} experiments. Smaller fractions of 2^k experiments, Criteria for Choosing Generators for 2^{k-p} Designs, Augmenting Fractional Factorials (Augmenting by Foldover or Mirror Image Designs, Augmenting by Optimal Design), Plackett-Burman (PB) and Model Robust Screening Designs, Mixed Level Factorials and Orthogonal Arrays (OAs), Definitive Screening Designs.	
	2.2	Fractions from Block Designs; Factors with 3 Levels, One-Third Fractions of 3^p Experiments; 3^{p-1} Experiments, One-Ninth Fractions of 3^p Experiments; 3^{p-2} Experiments. Fractions from Orthogonal Arrays, 2^p Orthogonal Arrays, Saturated Designs, $2^p \times 4^q$ Orthogonal Arrays ,3^p Orthogonal Arrays.	
	2.3	Creating and analyzing various designs in 2.1 using R code.	
		Self-Learning Activity: 1. Prepare a PPT on Half-Fractions of 2^k Designs. 2. Write summary of five web pages on Plackett-Burman (PB) Design. 3. Write a summary of three videos on Model Robust Screening Designs, 4. Prepare PPT on Definitive Screening Designs.	
Unit III		Incomplete and Confounded Block Designs	15 L
	3.1	Concept of contrast, orthogonal contrast, Tuckey's method of pairwise comparison, Scheffe's method of comparing all possible contrast, Dunnet test, Bartlet Test, Levenge test, Duncan Multiple Range Test, Friedman test.	
	3.2	Balanced Incomplete block design: Condition under which BIBD is used, Incomplete Block design, Balanced Incomplete Block Design, parameters of BIBD(v, r, b, k and λ), conditions for existence of BIBD ($vr = bk,\ \lambda(v-1) = r\,(k-1), b \geq v$), incidence matrix, properties of incidence Matrix, symmetric BIBD: Definition, necessary condition for symmetric BIBD with v as even, properties of symmetric BIBD, resolvable BIBD and affine resolvable BIBD, results on affine resolvable design, BTIB and PBIB Designs, Row Column Designs,	
	3.3	Intra class analysis of BIBD, Comparison of two treatments in BIBD, Efficiency of BIBD relative to RBD.	
	3.4	Creating and analyzing various designs in 3.1 and 3.2 using R code.	
		Self-Learning Activity: 1. Prepare PPT on BIBD. 2. Write comparison of BIBD, BTIB & PBIBD.	

		3. Write a summary on five web pages of PCBF Design. 4. Write a summary of application on Confounding desing.	
Unit IV		Response Surface Designs and Taguchi Method	15 L
	4.1	Introduction, Fundamentals of Response Surface Methodology (Empirical Quadratic Model, Design Considerations), Standard Designs for Second Order Models (Central Composite Design (CCD), Location of stationary point, Characterizing response surface, optimization of response, linear and quadratic model, stationary point, canonical analysis, Box-Behnken Design, Small Composite Design, Hybrid Design), Creating Standard Response Surface Designs in R, Non-Standard Response Surface Designs.	
	4.2	Fitting the Response Surface Model (RSM) with R (Fitting a Linear Model and Checking for Curvature, Fitting the General Quadratic Model, Fitting a Nonlinear Mechanistic Model), Determining Optimum Operating Conditions (Contour Plots, Canonical Analysis, Ridge Analysis, Nonlinear Optimization, Multi-Response Optimization), Blocked Response Surface (BRS) Designs.	
	4.3	Taguchi methods: Basic idea behind robust design, concept of loss function, overview of Taguchi Parameter design method, inner and outer array, orthogonal arrays, construction of inner and outer arrays, signal to noise ratio, split plot designs.	
	4.3	Creating and analyzing various designs in 4.1 and 4.3 using R code.	
		Self-Learning Activity: 1. Prepare PPT on one or more designs in 4.1. 2. Prepare PPT on one or more optimum operating conditions in 4.2. 3. Write a summary of five web pages on Blocked Response Surface (BRS) Designs.	

Suggested Readings Books:

1. **Lawson_Design-and-Analysis-of-Experiments-with-R:** http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/502_07_00_Lawson_Design-and-Analysis-of-Experiments-with-R-2017.pdf
2. Montgomery, D.C. (2001). Design and Analysis of Experiments, Wiley.
3. Dean, A. and Voss, D. (1999). Design and Analysis of Experiments, Springer.
4. S.C Gupta and V.K Kapoor (2007) Fundamentals of applied Statistics, Sultan Chand and Sons.
5. George E. P. Box, Draper N.R. (1987). Empirical Model-Building and Response Surfaces, Wiley.
6. Hicks, C.R., Kenneth V. and Turner, Jr. (1999). Fundamental Concepts in the Design of Experiments, Oxford University Press.
7. John P.W.M. (1971). Linear Models, Wiley.
8. Kshirsagar A.M. (1983). Linear Models, Marcel Dekker

Semester – IV	Paper – II
Course Code: MSCST 242 T	Title of the Course: Survival Analysis
Credits: 04	Total Lectures: 60 Hrs.

Course Outcomes (COs):

a) The objectives of this course are to study the different models from Survival Analysis, to understand different types of censoring, learn to estimate and interpret survival characteristics.
b) To provide the construction of parametric and non-parametric estimators of survival distributions, and probability density functions based on censored data.
c) To get general idea about the model to perform covariate analysis.

Detailed Syllabus:

Unit-I		Properties of Survival data and Life distributions	15L
	1.1	Concept of Survival Analysis, examples of failure or life time situations, points to follow while measuring survival time, concept of staggered and aligned entries in survival analysis, concept of censoring and truncation, types of censoring: Type I censoring with examples, type II censoring with examples, difference between type I and type II censoring, Undesigned censoring with examples, Right and left random censoring with examples and differences between them, type I censoring as a special case of Type I censoring, situations in which both right and left censoring occur, interval censoring with example.	
	1.2	Survival function, probability density function, hazard function, cumulative hazard rate, mean residual life function, equilibrium residual life function, interrelation between all these function, Total time on test transform (TTT), Concept of No-ageing and Characteristics of No-ageing, different classes of Positive ageing and Negative ageing, bathtub failure rate, concept of spacing and normalized spacing, distribution of spacing and normalized spacing.	
	1.3	IFRA closure property, bounds on reliability function of an IFRA distribution, Classification Parametric families of life distributions: Exponential distribution: Hazard rate for the exponential, plot of hazard rate for the exponential family using r software. Weibull Distribution: Hazard rate for Weibull distribution with scale parameter λ and shape parameter γ, classification of Weibull distribution based on different values of shape parameter (such as $0 < \gamma < 1$, $\gamma > 1$, $\gamma = 2$, $3 < \gamma < 4$), plotting hazard rate for different values of scale and shape parameter using r software. Gamma Family: Hazard rate for the Gamma distribution with scale parameter λ and shape parameter γ, plot of hazard rate for the Weibull family using R software, real life applications of these family in survival analysis.	

		Lognormal Distribution: Hazard rate for the Lognormal family, plot of hazard rate for the Lognormal family using R software, real life applications of this family in survival analysis. Hazard rate and classification of Linear failure rate family, Makeham family, Pareto family and plot of hazard rates for these families using R.	
	1.4	R coding for different computations and data visualization in 1.1 to 1.3	
		Self-Learning Activities: 1) Prepare a list of real-life examples where the type I, type II, Right and left random censoring. 2) See the video on the different classes of life distribution and prepare short summary. 3) Draw Hazard rate for the different life distribution using R software and classify the families accordingly. 4) Prepare PPT on application of different life distribution in survival Analysis.	
Unit II		Parametric Estimation	15L
	2.1	Parametric estimation of complete data: Concept of Score function, Fisher information matrix, Newton-Raphson and Fisher Method of Scoring to obtain MLE by using iterative method. 1. Exponential distribution: MLE of parameter λ, asymptotic distribution of $\hat{\lambda}$, exact and asymptotic confidence interval for λ, point and interval estimation using R software, Graphical methods of testing exponentiality. 2. Weibull: Obtaining MLE of scale and shape parameter of Weibull distribution using iterative methods (Newton Raphson and Fisher Method of Scoring), graphical method to find initial solution of parameters and sample information matrix, R code to find MLE of parameters of Weibull distribution. 3. Gamma: Obtaining MLE of scale and shape parameter of Gamma distribution and sample information matrix, R code to find MLE of parameters of Weibull distribution. 4. Lognormal distribution: Obtaining MLE of parameters (μ and σ^2) of Lognormal distribution, unbiased estimators of μ and σ^2, confidence interval for μ and σ^2, estimation of mean time to failure, R code to find MLE of parameters of Weibull distribution.	
	2.2	Parametric estimation of type I and type II censored data: 1. Type I Censoring Likelihood function for the type I censored data, finding MLE of parameter (λ) of exponential distribution in type I censored data, interpretation of estimator, estimate of mean (μ), point estimate of λ, testing of hypothesis problem Ho: $\lambda=\lambda_0$ Vs $\lambda < \lambda_0$ (or $\lambda > \lambda_0$), R commands for the above estimation problem. 2. Type II censoring: Likelihood function for the type II censored data Exponential distribution: finding MLE of parameter (λ) of exponential distribution in type II censored data, interpretation of	

		estimator, confidence interval for λ and R command for these estimation problems. Gamma distribution and Lognormal distribution: Obtaining MLE of scale and shape parameter of Gamma distribution, Obtaining MLE of parameters of Lognormal distribution using method of Cohen, R code to find MLE's	
	2.3	Parametric estimation of Right Random censored data: Likelihood function for the Right random censored data. Exponential distribution: finding MLE of parameter (λ) of exponential distribution in right random censored data, interpretation of estimator, R commands to solve the estimation problem. Weibull distribution: Obtaining MLE of scale and shape parameter of Weibull distribution using Newton Raphson, graphical method to find initial solution of parameters and sample information matrix, R code to find MLE's Proportional Hazard rate family: distribution function for proportional hazard rate family, hazard rate for the proportional hazard rate family, MLE of parameter of proportional hazard rate family, result to construct the confidence interval for the parameter of proportional hazard rate family.	
	2.4	R coding for different computations and data visualization in 2.1 to 2.3	
		Self-Learning Activities: 1) See the videos on the parametric estimation of Exponential, Weibull, gamma and lognormal distribution and share it on the student WhatsApp group. 2) See the videos on the parametric estimation type I and Type II censored data and prepare PPT of it. 3) See the videos on the parametric estimation Right randoms censored data and prepare summary of it.	
Unit III		Non-Parametric Estimation of Survival function and Test for exponentiality	15L
	3.1	Non parametric estimation of survival function for complete data: Empirical distribution function and Empirical survival function, Nonparametric estimation of Survival Function, Product-Limit Estimates of Survivorship functions, Relative, Five year and corrected survival rates, Finding Median Survival and Confidence Interval for the Median, Median Follow-up time, Empirical distribution function as unbiased estimator of distribution function, asymptotic distribution of empirical distribution function, unbiased estimator of survival function, plot of empirical survival function, asymptotic distribution of estimator survival function, confidence interval for survival function (by Using Kolmogorov – Smirnov statistics). Comparison of two groups of survival times, stratified tests.	

	3.2	Notations in life tables, Cohort life tables and current life table, reduce sample method, actuarial method of survival function for complete and censored data, delta method, variance of actuarial estimator, Kaplan Meier estimator of survival function, variance of Kaplan Meier estimator, redistribution to right Algorithm, R command to find the actuarial and Kaplan Meier estimator, plot of Kaplan Meir Estimator using R.	
	3.3	Test for exponentiality: need of tests for exponentiality, estimable parametric function, kernel, degree of Kernel, symmetric kernel, U-statistic, examples to compute U-statistics, Variance of U- statistics, One sample U-statistics theorem. Hallander and Proschan test for testing exponentiality against NBU class of distribution, R code for the test, Proof of the following properties of TTT transform 1) Total time on test transform $H^{-1}{}_F(t)$ is increasing function of t 2) Largest value of $H^{-1}{}_F(t)$ is μ mean of random variable T. 3) $\frac{d}{dt}\ H^{-1}{}_F(t) = \frac{1}{r(F^-(t))}$, where $r\left(F^-(t)\right)$ is hazard rate at $F^-(t)$. 4) F IFR if and only if $H^{-1}{}_F(t)$ is concave function of T 5) F is NBUE iff Ψ_F (t) $\geq t,\quad$ o $\leq t \leq 1,\ \Psi_F$ (t) is scale TTT transform. 6) $If\ F$ is exponential then Ψ_F (t)= t Tests for exponentiality against positive-ageing based on sample spacing a) Graphical test for testing exponentiality against IFR class b) Graphical test for testing exponentiality against NBUE class.	
	3.4	R coding for different computations and data visualization in 3.1 to 3.3	
		Self-Learning Activities: 1) Prepare R code to find estimator of survival function, distribution function and Confidence band on survival function. 2) See the video on the actuarial estimator of survival function and share on the students WhatsApp group. 3) See the videos on the Kaplan Meir estimator of Survival function and prepare the PPT on it. 4) See video on Hollander and Proschan Test and prepare summary of it. 5) Prepare a PPT on graphical test for exponentiality.	
Unit IV		One, two sample test to study Survival Data and test for Covariate models.	15L
	4.1	Estimation of TTT transform and scale TTT transform, Analytical test for testing exponentiality against NBUE class of alternative, computation of test statistic based on scale TTT transform, Moore's theorem, theorem on consistency of the test, consistency of the test. The Klefsijo test for exponentiality against IFRA alternative based on scaled TTT transform, exact distribution of test and asymptotic distribution of test.	

		Deshpande's test for testing exponentiality against IFRA alternative, computation of J_b statistic, asymptotic distribution of J_b statistic, proof of $\frac{1}{2} \leq J_b \leq 1$, computation of J_b using R software.	
	4.2	Two sample U-statistic theorem, Mann- Whitney test for complete data, computation of U-statistic, and asymptotic distribution of U-statistic, Gehan's test, Mental-Haenzal test for 2x2 contigency table, Log-Rank test, Torone-Ware class of tests, R code for all the test.	
	4.3	Concept of covariates, baseline model and link function, types of link function, Cox proportional hazard rate model as model for covariate analysis, Likelihood function for proportional hazard model in case of complete and censored data, estimation of parameter in the proportional hazard rate model, Nelson-Aalen estimators, graphical and Analytical test for constant of proportionality in Cox model.	
	4.4	R coding for different computations and data visualization in 4.1 to 4.3	
		Self-Learning Activities: 1) See the video on Klefsijo test and share on the student WhatsApp group. 2) See the video on Deshpande's test and prepare summary of it. 3) See the video Gehan's test and Mental-Haenzal test for 2x2 contingency table and prepare PPT on it. 4) See the video on Log-Rank test and Torone-Ware class of tests prepare summary on it.	

Suggested Readings Books:

1. Deshpande, J.V, Purohit, S.G., (2005), Life Time Data: Statistical Models and Methods
2. Miller, R.G. (1981) Survival Analysis, Wiley and Sons
3. Klein J. P. and Moeschberger M.L. (1997) Survival Analysis: Techniques for censored and truncated data. Springer, New York .
4. Collett D (2003) Modelling Survival Data in Medical research 2nd edition, Chapman and Hall/CRC
5. Cox, D.R. and Oakes, D. (1984) Analysis of Survival Data, Chapman and Hall, New York.
6. Elandt-Johnson, R.E., Johnson N.L. (1980) Survival models and Data Analysis, John Wiley and Sons
7. Gross A.J. and Clark, V. A. (1975) Survival Distributions: Reliability Applications in the Biomedical Sciences, John Wiley and Sons. 7.
8. Therneau T M and Grambsch P M (2000) Modelling Survival data extending the Cox model. Springer, New York.

Semester – IV	Paper – III
Course Code: MSCST 243 T	Title of the Course: Applications of Statistics in Clinical Trials
Credits: 04	Total Lectures: 60 Hrs.

Course Outcomes (COs):

a. Students will become next generation scientific supports to researchers in clinical trials
b. Students will understand the steps in clinical trials
c. Students will understand the ethical and legal issues involved in drug development
d. Students-researcher's interaction will gain more insight about in-situ operations of clinical trials.

Detailed Syllabus:

Unit I		Introduction of Clinical Trial and definition	15 L
	1.1	Clinical investigation before 1940's, limitations of case reports, Genesis of the clinical trials, the requirement for control, the dilemma of randomisation, the early monitoring tools-sequential trials, data monitoring committees. Introduction of Clinical Trials, definition of clinical trial, Bias, random error, Placebo, Treatment, multicentre clinical trial, Regulatory Process and Requirements, Investigation of New Drug Application, New Drug Application, Clinical Development and Practice. need and ethics of clinical trials, Protocol of clinical trials, Good Clinical Practice of Clinical trial, Case report of Clinical trial.	
	1.2	Blinding and types of blinding (Unblinded, Single, Double and Triple), Special problems in Double Blind studies (Matching drugs, coding drugs unbinding trials, assessment of blindness). Data management (Problems in data collection, major types of problems, specific examples of variability). Minimizing Poor Quality Data (Design of protocol and manual, development of forms, training, pretesting, techniques to reduce variability, data entry).	
	1.3	The Randomization Process (Fixed allocation randomization-simple, blocked and stratified randomization). Adoptive randomization procedures (Baseline adaptive randomization procedures, Response adaptive randomization), Mechanics of randomization. Concurrent control treatment. Target Population and Patient Selection, Selection of Controls Phases (I- IV) of Clinical trial, the need of clinical trials, Bioavailability, pharmacokinetics and pharmacodynamics, two compartment model	
		Self Learning 1. See the video of Clinical trial and share with students 2. Prepare summary of need and ethics of clinical trial 3. Prepare a summary of Protocol of clinical trial	

		4. Prepare PPT on Good Clinical Practice and Case report of Clinical trial 5. Prepare summary report on Goals of clinical trial 6. Prepare PPT on Phases of Clinical Trial 7. Write a summary of FIVE web pages on Randomization 8. Write a summary of THREE videos on Phases in clinical trials	
Unit II		Designs of Clinical Trials.	15 L
	2.1	The study objective, Basic Design Considerations, Randomized Control studies, nonrandomized concurrent control studies, Withdrawal studies, Factorial design, Group Allocation Designs, Hybrid Designs Parallel Designs, Crossover Design, The Balanced Incomplete Block Design, parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, Phase I trials, design of single-stage and multi-stage Phase II trials. Design and monitoring of Phase III trials with sequential stopping.	
	2.2	Sample Size: Dichotomous Response Variables (Two Independent Samples, Sample size tables for independent samples, Power for Independent samples, Confidence Interval approach for independent samples), Paired dichotomous response, adjusting sample size to compensate for nonadherence, sample size determination for "time to failure", sample size for testing "equivalency" of Intervention.	
	2.3	Statistical Inference for effects from standard 2X2 Cross Over Design: Introduction, the carry over effects, the direct Drug Effect, The Period Effect, the analysis of Variance., Tests for inter subject and intra subject variability with illustration.	
		Self Learning 1. See the video of Parallel Designs, Crossover Design and their application and share with students 2. Prepare summary of Phase I trials and Phase II and Phase III, designs 3. Prepare a summary of the direct Drug Effect, The Period Effect, the analysis of Variance 4. Prepare PPT on Statistical Inference for effects from standard 2X2 Cross Over Design 5. Write a summary of FIVE web pages on sample size determination	
Unit III		Statistical inference of clinical trial	15 L
	3.1	Statistical methods for Average Bioavailability : Introduction, Confidence interval approach(The Classical or Shortest confidence interval Westlake's Symmetric Confidence Interval, of Confidence Interval based on Fieller's theorem, Chow and Shao's joint Confidence Region)	
	3.2	The methods of interval hypothesis testing: Interval hypothesis, Schuirmann's Two one-sided Tests Procedure, Anderson and Hauck's Test. Bayesian methods (the Rodda and Davis method,	

		Mandallaz and Mau's method, Griever's method) concept of covariate and Nonparametric methods ,Wilcoxon-Mann-Whitney Two One Sided Test Procedure, Distribution- Free confidence interval based on the Hodges-Lehmann Estimator(Kaplan Meier Estimator, Cox proportional Hazard model, Discussion with other alternatives	
		Self Learning 1.See the video of Parallel Designs, Statistical methods for Average Bioavailability 2.Prepare summary of Different Types Confidence interval approach 3.Prepare a summary of The methods of interval hypothesis testing 4.Prepare PPT Nonparametric methods	
Unit IV		Statistical Analysis of clinical trials	15 L
	4.1	Optimal Cross Design for two formulations: Introduction, Balaam's Design (analysis of average bioavailability, inference for the carry over effect, the assessment of intra subject variabilities), The Two- sequence Dual Design (analysis of average bioavailability, inference for carry over effects, intra subject contrasts, the assessment of intra subject variabilities), Optimal Four Period Designs (Two sequence four period design, four sequence four period designs).	
	4.2	Assessment of average bioavailability with more than two formulations (statistical model & Assumptions, Confidence Interval & two one sided test procedure, log-transformation, variance balanced designs) Analysis of Williams design (Williams designs with three formulation, Williams designs with four formulation, Heterogeneity of intra subject variabilities)	
		Self Learning 1. See the video Optimal Cross Design for two formulations 2. Prepare PPT on of Power and Sample Size Determination	

Suggested Readings Books:

1. Chow S.C. and Liu J.P.(2009). Design and Analysis of Bioavailability and bioequivalence. 3rd Ed. CRC Press.
2. Chow S.C. and Liu J.P. (2004). Design and Analysis of Clinical Trials. 2nd Ed. Marcel Dekkar.
3. Fleiss J. L. (1989). The Design and Analysis of Clinical Experiments, Wiley.
4. Friedman L. M. Furburg C. Demets D. L.(1998). Fundamentals of Clinical Trials, Springer.
5. Jennison .C. and Turnbull B. W. (1999). Group Sequential Methods with Applications to Clinical Trails, CRC Press.
6. Marubeni .E. and Valsecchi M. G. (1994). Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley. ST O18 Statistical Methods in Micro-Array Data Analysis [ST 202]

Semester – IV	Paper – IV
Course Code: MSCST-244 T	Title of the Course: Non-Parametric Inference using R
Credits: 04	Total Lectures: 60 Hrs.

Course Outcomes (COs):

1. On study of the course distinction between parametric and non-parametric tests will be more clear.
2. Opportunities of R coding will excel the logical thinking of the students.
3. Students will get a option if underlying domain dose not satisfy the condones of parametric set up.
4. Very useful to address the real life problems in Data science.

Detailed Syllabus:

Unit-I		One Sample Methods	15 L
	1.1	Concept of nonparametric Test of Hypothesis. Some comments on Nonparametric statistics. and Confidence interval for the Median a) Binomial Test and estimation of p. b) The Quantile test and estimation of x_p. c) Tolerance limits / Confidence Interval d) The sign test and some variations of it.	
	1.2	Estimating the Population cdf and Percentiles a) Confidence Interval for the Population cdf b) Inference for Percentiles	
	1.3	A comparison of Statistical Test Type I errors, Power and its Derivations	
		Self Learning Activity: 5. Write summary of FIVE web pages on Binomial Test. 6. Write summary of THREE videos on sign test 7. Prepare PPT on Confidence Interval for the Population cdf. 8. Prepare PPT on Type I errors, Power and its Derivations.	
Unit-II		Two-Sample Methods	15L
	2.1	A Two-Sample Permutation Test a) The Permutation Test b) Steps Used in a Two-Sample Permutation Test c) Hypothesis for the two –Sample Permutation Test Permutation Test Based on the Median and Trimmed Means. a) A permutation Test Based on Medians. b) Trimmed Means.	
	2.2	Wilcoxon Rank-Sum Test a) Steps in Conducting the Wilcoxon Rank-Sum Test b) Comments on the Use of the Wilcoxon Rank-Sum Test c) A Statistical Table for the Wilcoxon Rank-Sum Test Wilcoxon Rank-Sum Test Adjusted for Ties. Steps in Conducting the Wilcoxon Rank-Sum Test Mann-Whitney Test and Confidence Interval	

		a) The Mann-Whitney Statistic b) Equivalence of Mann Whitney and Wilcoxon Rank-Sum Statistic. c) A confidence interval for a Shift Parameter and the Hodges-Lehmann Estimate	
	2.3	Test for equality of scale parameters and Omnibus test a) Siegel-Tukey and Ansari-Bradley test b) Tests on Deviances c) Kolmogorov-Smirnov Test Selecting among Two sample tests a) The 't' Test b) The Wilcoxon Rank Sum Test Verses T test c) Relative Efficiency d) Power of permutation Test	
		Self Learning Activity: 5. Write summary of FIVE web pages on A Two-Sample Permutation Test 6. Write summary of THREE videos Permutation Test Based on the Median and Trimmed Means. 7. Prepare PPT on Mann-Whitney Test and Confidence Interval 8. Prepare PPT on Test for equality of scale parameters and Omnibus test 9. Prepare PPT on Test for equality of scale parameters and Omnibus test	
Unit-III		K-Sample Methods	15L
	3.1	K-Sample Permutation Test: a) The F Statistic b) Steps in Carrying Out the Permutation F-Test c) Alternative Forms of the Permutation F-Statistic	
	3.2	The Kruskal-Wallis Test a) The Kruskal-Wallis Statistic b) Adjustment for Ties c) An intuitive Derivation of the Chi-Square Approximation for KW test d) Test on General Scores	
	3.3	Multiple Comparisons a) Three Rank –Based Procedure for Controlling for Experiment-Wise Error Rate Assuming No Ties in the Data. b) Multiple Comparisons for General Score Test (Including Ties) c) Multiple Comparison Permutation Tests d) Variance of a Difference of Means When Sampling from a Finite Population	
		Self Learning Activity:	

		5. Write summary of FIVE web pages on K-Sample Permutation Test 6. Write summary of THREE videos on The Kruskal-Wallis Test 7. Prepare PPT on Multiple Comparison Permutation Tests	
Unit-IV		Paired Comparisons and Blocked Designs	15 L
	4.1	Paired-Comparison Permutation Test a) Steps for Paired –Comparison Permutation Test b) Randomly Selected Permutations c) Large Sample Approximation d) A Test for the Median of a Symmetric Population Sign Rank Test a) The Wilcoxon Signed Rank Test without Ties. b) Large Sample Approximation c) Adjustment for Ties	
	4.2	Other Paired Comparison Tests a) Sign Test b) General Scoring System c) Selecting Among Paired-Comparison Test A Permutation Test for Randomized Complete Block Design a) F-Statistic for Randomized Complete Block Designs b) Permutation F-Statistics for Randomized Complete Block Designs c) Multiple Comparisons	
	4.3	Friedman's Test (FM) for Randomized Complete Block Design a) Friedman's Test without Ties b) Adjustment for Ties (FM_{ties}) c) Cochran's Q and Kendall's W d) Chi-Square Approximation for FM and FM_{ties}	
		Self Learning Activity: 9. Write summary of FIVE web pages on Paired-Comparison Permutation Test 10. Write summary of THREE videos on Sign Test 11. Prepare PPT on Permutation Test for Randomized Complete Block Design 12. Prepare PPT on Friedman's Test (FM) for Randomized Complete Block Design	

Suggested Readings Books:

1. **Introduction to Modern Nonparametric Statistics by James J. Higgins**
2. Nonparametric Statistical Inference By Jean Dickinson Gibbons, Subhabrata Chakraborti
3. Nonparametric Statistical Methods by Bhattacharya, Medtech
4. https://spu.fem.uniag.sk/cvicenia/ksov/prokeinova/Business%20Statistics%20and%20Econometrics/Literature/20089702653110.pdf
5. https://faculty.ksu.edu.sa/sites/default/files/nonparametric_statistics_a_step-by-step_approach.pdf

Semester – IV	Paper – V
Course Code: MSCST 245 P	Title of the Course: Practical's Based on Computational Tools-IV
Credits: 04	Total Lectures: 60 Hrs.

List of Practical All practicals are based on R Coding

Course Outcomes (Cos):

1. On learning different computational tools students will have capacity to compare the features each of them with others.
2. Students can perform statistical analysis by different computational tools so as to know various parts of output.
3. On working with big data students will get confidence and competence.
4. Students will build logical thinking ability to develop statistical utilities.
5. Students will get an exposure of interacting with different platforms.

Practical's based – MSCST-241 T Designs of Experiments Using R

1) Creating and analyzing On factorial, full factorial and factorial designs with multiple factors.
2) Creating and analyzing fractional factorial designs, fractions from block design.
3) Creating and analyzing BIBD, BTIB and PBIB.
4) Analysis of various standard designs for second order models. Fitting of RSM. Taguchi method.

Practical's based – MSCST-242 T Survival Analysis

5) Parametric analysis of complete data using Exponential, Weibull, Gamma and Lognormal models and censored data of Type I and Type II censoring.
6) Computation of Actuarial estimator of survival function and PL –estimator and their variances.
7) Tests for exponentiality against different classes of life distribution and different kernels. Graphical test against IFR and NBUE class.
8) Two Sample Tests : Two sample Mann- Whitney test for complete data, Gehan's test, Mental-Haenzal test for 2x2 contingency table, Log-Rank test, Torone-Ware class of tests. Nelson- Aalen estimator.

Practical's based – MSCST-243 T Applications of Statistics in Clinical Trials

9) Sample size determination under different cases (as in 2.2)
10) Statistical inference for effect from standard 2X2 cross over design and cross over design for two formulations
11) Testing of hypothesis for various type of clinical trial, Power of the test and sample size determination of clinical trials.
12) Hypothesis testing and estimation of confidences interval for bioequivalence study.
13) Analysis of one-way classification, two-way classification, Multiple comparison tests.
14) Study of pharmacokinetics and pharmacodynamics, two compartment model

Practical's based – MSCST-244 T Non-Parametric Inference Using R

15) Test for Median, Binomial test, Quantile test, Sign Test – Type I error and power of the test. Two sample permutation test, permutation test for medians and trimmed means, Wilcoxon Rank-Sum Test, Mann-Whitney Test. Power of the tests.
16) Test for equality of scale parameters and Omnibus test, Siegel-Tukey and Ansari-Bradley test, Tests on Deviances, Kolmogorov-Smirnov Test. Selecting among Two sample tests - The 't' Test.
17) K-Sample Permutation Test, The Kruskal-Wallis Test including adjustments for ties, test on general scores.
18) Multiple Comparisons - Three Rank –Based Procedure for Controlling for Experiment-Wise Error Rate Assuming No Ties in the Data, Multiple Comparisons for General Score Test (Including Ties), Multiple Comparison Permutation Tests
19) Paired-Comparison Permutation Test, Large Sample Approximation, A Test for the Median of a Symmetric Population. Sign Rank Test-The Wilcoxon Signed Rank Test without Ties and with ties, Large Sample Approximation.
20) Friedman's Test (FM) for Randomized Complete Block Design- Friedman's Test without Ties and with ties, Cochran's Q and Kendall's W, Chi-Square Approximation for FM and FM_{ties1}