SRI DHARMASTHALA MANJUNATHESHWARA COLLEGE (AUTONOMOUS)
UJIRE – 574 240
DAKSHINA KANNADA, KARATAKA STATE
(Re-Accredited by NAAC at 'A++' Grade)
Ph: 08256-236221, 236101(O), Fax: 236220
e-Mail:sdmcollege@rediffmail.com, sdmcollege@sdmcujire.in, Website: www.sdmcujire.in

DEPARTMENT OF PG STUDIES & RESEARCH IN STATISTICS

Syllabus of
Masters' Degree Programme in Statistics

(CHOICE BASED CREDIT SYSTEM (CBCS) SEMESTER SCHEME)

2016-17 ONWARDS

(EFFECTIVE FROM ACADEMIC YEAR 2024-25)

Approved by the BOS meeting held on 24-02-2024

Approved in Academic Council meeting held on 23-03-2024

INDEX

PREAMBLE

Revision of the Syllabus for the Two years Master Degree (Choice Based Credit System – Semester Scheme) Programme in Statistics.

The PG BOS in Statistics has prepared the revised Syllabus for M.Sc. Statistics (CBCS based) in its meeting held on 3rd December 2022, as per the guidelines suggested by Mangalore University and University Grants Commission, New Delhi. It was resolved to implement this new syllabus from the academic year 2022-23.

In the present revised syllabus, the suggested course pattern includes Hard Core, Soft Core and Open Elective courses with 92 credits for the entire programme. The syllabus consists of 14 Hard Core courses (4 credits each) including 11 theory (3 in I, II, III, and 2 in IV semesters), 2 practicals (in I semester) and one Project work (in IV semester), with a total of **56 credits**. It also consists of 10 Soft Core courses (3 credits each) including 5 theory (1 in I, II, III, and 2 in IV semesters) and 5 practicals (2 in II, III, and 1 in IV semesters), with a total of **30 credits**. The BOS has also proposed 6 Open Elective courses (3 each in II and III semesters) with 3 credits each (with a total of **6 credits**), to be offered to non-Statistics students. Student shall opt any 1 course each in II and III semesters respectively. But the credits of Open Elective courses are not considered for CGPA. All together **total credits** come to **92** (including the credits for Open Elective courses), otherwise, a **total** of **86 credits**.

The revised syllabus is designed to impart quality education by incorporating skill components with practical knowledge. The syllabus has been prepared in a participatory manner, after discussions with a panel of members consisting of subject experts, industrial experts and meritorious alumnus of the department by referring the existing syllabi, U.G.C. model curriculum and the syllabi of other Universities and National Institutes.

The syllabus is structured in a view to prepare the students for higher studies and employability. The curriculum imparts knowledge to the students on the skills required for contributing to the industry and academic institutions. Statistics being an applied science, sufficient emphasis is given in the syllabus for training in data analysis and interpretation. The units of the syllabus are well defined. The number of contact hours required for each unit is also given. A list of reference books is provided at the end of each course syllabus.

ELIGIBILITY FOR ADMISSION

B.Sc. Degree from recognized University, with Statistics/Mathematics as major/ optional subject at the undergraduate level with 45% aggregate excluding languages (40% for SC/ST/Category-I candidates).

PROGRAMME OBJECTIVES

The syllabus aims to:

- Gain knowledge on theoretical foundations for the development of various statistical concepts and find solutions to real world problems by applying quantitative modelling and data analysis techniques.
- Learn about various computational and statistical softwares used for the development and execution of various statistical techniques.
- Broaden computing and statistical skills by giving an exposure towards data science concepts and techniques in solving real-world problems.
- Prepare the younger generation tackle the emerging problems through applications of statistics.
- Develop the department as a hub for teaching, research and extension activities in the field of modern statistical theory and methods.

PROGRAMME SPECIFIC OUTCOMES

The learner will be able to:

PSO1: Gain knowledge on theoretical foundations for the development of various statistical concepts and procedures.

PSO2: Develop technical skills in probability modelling and statistical inference for the practical application of statistical methods in future employment.

PSO3: Find solutions to real world problems by applying quantitative modelling and data analysis techniques.

PSO4: Use computational and statistical softwares to develop and execute various statistical techniques and statistical computing algorithms.

PSO5: Demonstrate theoretical knowledge and applications of parametric, semi-parametric and non-parametric testing procedures.

PSO6: Understand how to design experiments and surveys with a view of providing solutions to real life problems.

PSO7: Use statistical reasoning, formulate a problem in statistical terms, perform exploratory analysis of data, and carry out a variety of advanced inferential procedures.

PSO8: Tackle emerging problems through applications of statistics.

COURSE/ CREDIT PATTERN

Semester	Credits (C)						Total
	Theory (T)			Practical (P)		Project (Pr)	
	Hard Core (H)	Soft Core (S)	Open Elective (E)	Hard Core (H)	Soft Core (S)	Hard Core (H)	
First	3T×4C=12	1T×3C=3	---	2P×4C=8	---	---	23
Second	3T×4C=12	1T×3C=3	1T×3C=3	---	2P×3C=6	---	21+3*
Third	3T×4C=12	1T×3C=3	1T×3C=3	---	2P×3C=6	---	21+3*
Fourth	2T×4C=8	2T×3C=6	---	---	1P×3C=3	1Pr×4C=4	21
Total	44	15	6*	8	15	4	**86+6***

Total Credits from all the Four Semesters = 92

Total Hard Core Credits = 44 (T) + 12 (P) = 56 = 60.9%

Total Soft Core Credits = 15 (T) + 15 (P) = 30 = 32.6%

*Open Elective Credits = 6 = 6.5% (Not to be considered for calculating the CGPA)

COURSE PATTERN AND SCHEME OF EXAMINATION

I Semester

Course Code	Course Title	No. of Units	Evaluation IA + Exam	Teaching Hrs/Week	Exam Hrs	Credits

STHT 501	Real Analysis	4	30 + 70	4	3	4
STHT 502	Probability and Distributions - I	4	30 + 70	4	3	4
STHT 503	Theory of Sampling	4	30 + 70	4	3	4
STST 501	Linear Algebra	4	30 + 70	3	3	3
STST 502	Elements of Statistical Computing**	4	30 + 70	3	3	3**
STHP 501	R programming and Theory of Sampling Practical	---	30 + 70	8	3	4
STHP 502	Programming in Python Practical	---	30 + 70	8	3	4
	Total					**23**

****Optional**

II Semester

Course Code	Course Title	No. of Units	Evaluation IA + Exam	Teaching Hrs/Week	Exam Hrs	Credits
STHT 551	Probability and Distributions - II	4	30 + 70	4	3	4
STHT 552	Design and Analysis of Experiments	4	30 + 70	4	3	4
STHT 553	Theory of Estimation	4	30 + 70	4	3	4
STST 551	Data Management and Statistical Computing with Python	3	30 + 70	3	3	3
STST 552	Actuarial Methods**	4	30 + 70	3	3	3**
STOE 551	Introductory Statistics and Data Analysis	3	30 + 70	3	3	3*
STOE 552	Questionnaire Design and Sample Selection	3	30 + 70	3	3	3*
STOE 553	Data Visualization	3	30 + 70	3	3	3*
STSP 551	Design & Analysis of Experiments and Theory of Estimation Practical	---	30 + 70	6	3	3
STSP 552	Data Management and Statistical Computing with Python Practical	---	30 + 70	6	3	3
	Total					**21+3***

***This credit is not included for CGPA.**
****Optional**
III Semester

Course Code	Course Title	No. of Units	Evaluation IA + Exam	Teaching Hrs/Week	Exam Hrs	Credits
STHT 601	Testing of Hypotheses	4	30 + 70	4	3	4

STHT 602	Regression Analysis	4	30 + 70	4	3	4
STHT 603	Multivariate Analysis	4	30 + 70	4	3	4
STST 601	Stochastic Processes	4	30 + 70	3	3	3
STST 602	Stochastic Finance**	4	30 + 70	3	3	3**
STOE 601	Inferential Statistics and Data Analysis	3	30 + 70	3	3	3*
STOE 602	Categorical Data Analysis	3	30 + 70	3	3	3*
STOE 603	Demographic Methods and Analysis	3	30 + 70	3	3	3*
STSP 601	Testing of Hypotheses and Regression Analysis Practical	---	30 + 70	6	3	3
STSP 602	Machine Learning with Python Practical	---	30 + 70	6	3	3
	Total					**21+3***

***This credit is not included for CGPA.**
****Optional**

IV Semester

Course Code	Course Title	No. of Units	Evaluation IA + Exam	Teaching Hrs/Week	Exam Hrs	Credits
STHT 651	Time Series Analysis	4	30 + 70	4	3	4
STHT 652	Reliability and Survival Analysis	4	30 + 70	4	3	4
STST 651	Statistical Modelling	3	30 + 70	3	3	3
STST 652	Operations Research**	4	30 + 70	3	3	3**
STST 653	Big Data Analytics	4	30 + 70	3	3	3
STST 654	Artificial Intelligence**	4	30 + 70	3	3	3**
STSP 651	Time Series Analysis, Reliability and Survival Analysis and Statistical Modelling Practical	---	30 + 70	6	3	3
STPD 651	Project Work	---	30 + 70	8	---	4
	Total					**21**

****Optional**

SCHEME OF INTERNAL ASSESSMENT EVALUATION

The scheme of evaluation for internal assessment marks shall be as follows:

i. Two Internal Tests — 20 marks
ii. Seminar/Assignments/Classroom Activities etc. — 10 marks

Total: 30 marks

THEORY QUESTION PAPER PATTERN

The pattern of question paper in theory examinations shall be as follows:

i. There shall be totally 8 questions in which Q. No. 1 is compulsory. Students have to answer any 4 questions from the remaining 7 questions.
ii. Q. No. 1 will contain 8 questions of short answer type, each question carrying 3 marks. Students will have to answer any 6 questions. Thus Q. No. 1 carries 18 marks.
iii. Q. No. 2 to Q. No. 8 will be of long answer type, each question carrying 13 marks.

The distribution of marks will be as follows:

Q. No. 1	3×6 = 18
Any 4 questions out of remaining 7 questions	13×4 = 52

Total = 70

M.Sc. Statistics

Time: 3 hrs **Max. Marks: 70**

<u>Instruction</u>: Question No. 1 is compulsory. Answer any FOUR from the remaining seven questions.

1. **Answer any SIX questions.** **6 × 3 = 18**
   a)
   b)
   c)
   d)
   e)
   f)
   g)
   h)

2. **a)**
   b)
   c)
   (13)
3. **a)**
   b)
   c)
   (13)
4. **a)**
   b)
   c)
   (13)
5. **a)**
   b)
   c)
   (13)
6. **a)**
   b)
   c)
   (13)
7. **a)**
   b)
   c)
   (13)
8. **a)**
   b)
   c)
   (13)

PRACTICAL EXAMINATION PATTERN

In the practical examination, out of 70 marks, 10 marks shall be allotted for practical record and 60 marks for practical proper. In the IV semester, there shall be project work of 70 marks. The project work shall be conducted in the department. Project presentation and report shall be valued for 70 marks.

I SEMESTER

STHT 501: REAL ANALYSIS

Teaching Hours: 4 Hrs per week

Learning Objectives:

LO1: To gain knowledge on the fundamental aspects involved in the theory of real analysis

LO2: To explore the concept of limits and its usage in sequences and series of real numbers

LO3: To learn the concept of convergence and its implementation in sequences and series of functions.

LO4: To understand the theory involved in the concept of functions of two variables.

Course Outcomes:

CO1: Ability to understand the fundamental properties of real numbers that lead to the formal development of real analysis.

CO2: Understand the concept of limits and how they are used in sequences and series of real numbers.

CO3: Understand the concept of convergence and how they are implemented in sequences and series of functions.

CO4: Understand the implementation of theoretical aspects involved in functions of two variables.

UNIT-I **13 Hrs**

Elements of set theory, sets in Euclidean space of k-dimensional R^k rectangles. Metric spaces, neighbourhood, interior point and limit point, open and closed sets, Bolzano-Weierstrass theorem in R^2, compact set, real-valued functions, Heine-Borel theorem (Statement only), continuity and uniform continuity.

UNIT-II **13 Hrs**

Sequences and Series of real numbers - Cauchy sequence, convergence of bounded monotone sequence. Limit superior, limit inferior and limit properties. Series of positive terms - tests for convergence, divergence. Series of arbitrary terms - absolute and conditional convergence.

UNIT-III **13 Hrs**

Sequences of functions - uniform convergence and point wise convergence, series of functions - uniform convergence, Weierstrass M test. Power series and radius of convergence. Riemann-Stieltjes integration continuous integrand and monotonic/ differentiable integrator.

UNIT-IV **13 Hrs**

Functions of two variables - partial and directional derivatives. Maxima and minima of functions, maxima-minima under constraints (Lagrange's multipliers). Improper integrals.

References:

1. Apostol, T. M. (2002). *Mathematical Analysis*. Narosa India Ltd.
2. Bartle, R. G. (1975). *The Elements of Real Analysis* (2nd ed.). Wiley.
3. Courant, R. and John, F. (1974). *Introduction to Calculus and Analysis*. John Wiley $ Sons Inc.
4. Goldberg, R. R. (2020). *Methods of Real Analysis*. Oxford Publishing Co.
5. Khuri, A. I. (2002). *Advanced Calculus with Applications in Statistics*. WileyInterscience.
6. Rudin, W. (2017). *Principles of Mathematical Analysis* (3rded). McGraw Hill Education.

STST 501: LINEAR ALGEBRA
Teaching Hours: 3 Hrs per week

Learning Objectives:

LO1: To gain knowledge on the theoretical foundations involved in the concepts of field and vector geometry.
LO2: To study about the computational techniques involved in the theory of matrix algebra, linear transformations, and systems of linear equations.
LO3: To explore the concepts of eigen values and eigen vectors and its importance in matrix theory.
LO4: To understand the theory and applications involved in the concepts of quadratic forms, vector and matrix differentiation.

Course Outcomes:

CO1: Awareness of necessary theoretical foundations on field and vector geometry, which will help them better understand linear models and multivariate analysis.
CO2: Develop algebraic skills and knowledge on computational techniques essential for the study of matrix algebra, linear transformations, and systems of linear equations.
CO3: Understand the concepts of eigen values and eigen vectors and their implementation in matrix theory.
CO4: Understand the theoretical aspects of quadratic forms, vector and matrix differentiation, with their practical implementation.

UNIT-I **10 Hrs**

Fields, vector spaces, subspaces, linear dependence and independence, basis and dimension of a vector space, finite dimensional vector spaces, completion theorem. Examples of vector spaces over real fields. Vector spaces with an inner product, Gram-Schmidt orthogonalization process, orthonormal basis.

UNIT-II **10 Hrs**

Row and column spaces of a matrix. Rank and inverse of a matrix, properties of inverse. Rank of a product of matrices, partitioned submatrices, rank factorization of a matrix, rank of a sum, inverse of a partitioned matrix. General linear system of equations, generalized inverse, Moore-Penrose inverse, idempotent matrices. Solutions of matrix equations.

UNIT-III **10 Hrs**

Characteristic roots and vectors, Cayley-Hamilton theorem, minimal polynomial, similar matrices. Algebraic and geometric multiplicity of characteristic roots, spectral decomposition of a real symmetric matrix, reduction of a pair of real symmetric matrices, singular value decomposition.

UNIT-IV **10 Hrs**

Real quadratic forms, reduction and classification of quadratic forms, index and signature. Extrema of quadratic forms. Vector and matrix differentiation.

References:

1. Biswas, S. (1984). *Topics in Algebra of Matrices*. Academic Publications.
2. Hadley, G. (2002). *Linear Algebra*. Narosa.
3. Graybill, F. A. (1983). *Matrices with Applications in Statistics*.
4. Rao, A. R. and Bhimasankaran, P. (2000). *Linear Algebra* (2^{nd}ed).Hindustan book agency.
5. Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley.
6. Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications*. Wiley.
7. Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. Wiley.

STHT 502: PROBABILITY AND DISTRIBUTIONDS I
Teaching Hours: 4 Hrs per week

Learning Objectives:
LO1: To understand the necessary theoretical foundations required for the development of probability theory.
LO2: To gain knowledge on the fundamentals and principles involved in probability theory.
LO3: To study about the characteristics of various discrete and continuous distributions, and also learn about various transformation techniques.
LO4: To learn about the concepts of order statistics, truncated and mixed distributions.

Course Outcomes:
CO1: Familiarity with necessary theoretical foundations on the developments of probability theory.
CO2: Ability to understand the fundamental aspects and principles of probability theory.
CO3: Exhibit learning about the standard discrete and continuous distributions and its characteristics, and various transformation techniques.
CO4: Show improved knowledge on order statistics, and truncated and mixed distributions.

UNIT-1 **13 Hrs**
Algebra of sets, sequence of sets and limits, fields and sigma-fields, minimal sigma-field. Events, sample space. Probability measure, probability space, property of probability measure, properties related to sequences of events, independent events, conditional Probability.

UNIT-II **12 Hrs**
Measurable functions, random variables, probability induced by a random variable. Definition of simple random variables. Integration of measurable functions with respect to measures. Expectation, properties of expectation, moments, inequalities. Standard discrete distributions:

Bernouli, Binomial, Poisson, Negative Binomial, Geometric, Hyper Geometric, Discrete Uniform distributions

UNIT-III **16 Hrs**

Continuous distributions: Normal, Gamma, Beta (Two kinds), Exponential (Univariate and bi-variate), Cauchy, Weibul, Logistic, Lognormal, Laplace and Pareto distributions and their properties. Probability generating function and moment generating function. Bivariate normal and multinomial distributions. Transformation techniques. Distribution of functions of random variables.

UNIT-IV **11 Hrs**

Order Statistics - their distributions and properties, joint and marginal distributions. Distribution of range and median. Truncated and mixture distributions.

References:

1. Ash, R. B. and Catherine Doleans-Dade (2000). *Probability and Measure Theory*. Academic Press.
2. Bhat, B. R. (2018). *Modern Probability Theory* (4rd ed.). New Academic Science Ltd.
3. Johnson, S. and Kotz (1972). *Distributions in Statistics*. Vols. I, II and III, Houghton and Miffin.
4. Mukhopadhyaya, P. (2016). *Mathematical Statistics*. Books and Allied.
5. Pitman, J. (1993). *Probability*. Narosa.
6. Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley Eastern.
7. Rohatgi, V. K. and Saleh, A. K. Md. E. (2015). *An Introduction to Probability Theory and Mathematical Statistics*. Wiley Eastern.
8. Laha, R. G. and Rohatgi, V. K. (1979). *Probability Theory*. Wiley Eastern.
9. Ross, S. M. (1993). *First Course in Probability*. Academic Press.
10. Billingsley, P. (1986). *Probability and Measure*. John Wiley and Sons.

STHT 503 - THEORY OF SAMPLING

Teaching Hours: 4 Hrs per week

Learning Objectives:

LO1: To explore theoretical foundations involved in the development of sampling theory and PPSWR sampling technique.

LO2: To gain knowledge on the theoretical concepts involved in the development of PPSWOR sampling technique.

LO3: To learn the theory and applications of single stage cluster sampling and two stage sampling.

LO4: To study about two phase sampling and also regarding randomized response techniques.

Course Outcomes:
CO1: Ability to understand the theoretical aspects involved in the development of sampling theory and PPSWR sampling technique.
CO2: Exhibit theoretical knowledge on various concepts involved in the development of PPSWOR sampling technique.
CO3: Ability to understand the theory of single stage cluster sampling and two stage sampling with their real life applications.
CO4: Understand the theory involved in two phase sampling and explore about randomized response techniques.

UNIT-I **13 Hrs**
Basic Concepts - sampling design, sampling scheme, sampling strategy, interpenetrating subsampling, concept of non-random sampling. Probability proportional to size with replacement (PPSWR) sampling - selection of PPSWR sample, estimation of population mean, total and their sampling variances. Hansen-Hurwitz strategy, estimation of sampling variance. Comparison with SRSWR, estimation of gain due to PPSWR sampling.

UNIT-II **13 Hrs**
Varying probability without replacement (PPSWOR) sampling - some properties of sampling design. Horwitz-Thomson estimator, sampling variance of population total and its unbiased estimator. Sen-Midzuno sampling scheme, Des-Raj's ordered estimator (general case), Murthy's unordering principle (sample of size two).

UNIT-III **13 Hrs**
Single stage cluster sampling - concepts, estimation of efficiency of cluster sampling, clusters of varying sizes. Two-stage sampling - notions, estimation of population total and its variance, efficiency of two-stage sampling relative to cluster and uni-stage sampling.

UNIT-IV **13 Hrs**
Ratio and regression estimators based on SRSWOR, method of sampling, bias and mean square errors, comparison with mean per unit estimator. Two-phase sampling - notion, double sampling for ratio estimation, double sampling for regression estimation. Randomized response techniques - Warner's model, related and unrelated questionnaire methods, non-sampling errors.

References:

1. Cochran, W. G. (2007). *Sampling Techniques* (3rd ed.). Wiley.
2. Des Raj and Chandok (1998). *Sampling Theory*. Narosa Publication.
3. Mukhopadhyay, P. (2013). *Theory and Methods of Survey Sampling* (2nded). Prentice Hall of India.
4. Murthy, M. N. (1977). *Sampling Theory and Methods*. Calcutta: Statistical Publishing Society.

5. Sampath, S. (2001). *Sampling Theory and Methods*. Narosa Publishers.
6. Singh, D. and Chaudhary, F. S. (1986). *Theory and Analysis of Sample Survey Designs*. New Age International Publishers.
7. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Ashok (1984). *Sampling Theory of Surveys with Applications*. ICAR Publication.
8. Banett, V. (2002). *Sample Survey: Methods and Principles*. Arnold Publishers.

Self-Study Concepts:

National Development - Official Statistics for National Development - NSO, CSO, MOSPI, Human Development Index. Measuring inequality in income - Lorenz Curve, Gini coefficient.

References:

1. Sen, A. (1997). *Poverty and Inequality*.

STHP 501 - R PROGRAMMING AND THEORY OF SAMPLING PRACTICAL

1. Fundamentals of R
2. Handling datasets using R
3. Graphical analysis using R
4. Descriptive statistics using R
5. Functions in R
6. Gram-Schmidt orthogonalization process using R
7. One sample and two sample parametric tests using R
8. Nonparametric tests for independence of attributes using R
9. Determination of sample size
10. Probability proportional to size with replacement (PPSWR)
11. Probability proportional to size without replacement (PPSWOR)
12. Des Raj's ordered estimator and Murthy's unordered estimator
13. Cluster sampling with clusters of equal and unequal sizes
14. Two-stage sampling
15. Ratio and regression method of estimation

STHP 502 - PROGRAMMING IN PYTHON PRACTICAL

1. Program to perform arithmetic operations and display the output using placeholders.
2. Programs based on decision making statements.
3. Loop programming exercises.
4. Programs based on probability and statistics.
5. Programs to perform mathematical operations.
6. Programs for matrix operations.

II SEMESTER

STOE 551: INTRODUCTORY STATISTICS AND DATA ANALYSIS
Teaching Hours: 3 Hrs per week

Learning Objectives:
LO1: To explore different techniques involved in descriptive statistics and their practical applications.
LO2: To study about the characteristics and applications of binomial, Poisson and normal distributions.
LO3: To understand the real life applications of various sampling techniques.

Course Outcomes:

CO1: Ability to explain the various techniques involved in descriptive statistics and their implementation in real life problems.
CO2: Show an improved knowledge about standard univariate distributions such as binomial, Poisson and normal, and understand its applications.
CO3: Ability to apply various sampling techniques in real life.

UNIT-I **13 Hrs**

Statistics - introduction, meaning, definition and scope of the subject as a science of decision making against uncertainty. Data types, methods of collection, presentation in the form of tables and graphs. Descriptive Statistics - measures of central tendency, positional averages, measures of dispersion, skewness and kurtosis. Methods of summarizing categorical data - univariate and bivariate contingency tables. Box plots - construction and interpretations.

UNIT-II **15 Hrs**

The concept of permutation and combination, random experiment, simple events, sample space, types of events, probability of an event, rules of probability, conditional probability, Baye's rule, exercises on computation of probabilities using these rules to fix the ideas. The concept of random variables - discrete and continuous type, Binomial, Poisson and Normal distributions - their use in practical applications, computing probabilities using these distributions.

UNIT-III **12 Hrs**

Sampling methods - population and sample, parameter and statistic, concepts of random sample, simple random sampling, stratified sampling, systematic, sampling, sample size determination. The concept of sampling distribution of sample mean.

References:

1. Campbell, R. C. (1989). *Statistics for Biologists*(3rd ed). Cambridge University Press.
2. Chatfield, C. (1981). *Statistics for Technology*. Chapman and Hall.
3. Frank, H. and Athoen, S. C. (1997). *Statistics: Concepts & Applications*. Cambridge University Press.
4. Medhi, J. (2006). *Statistical Methods: An Introductory Text*. New Age International Publisher.
5. Ross, S. M. (2017). *Introductory Statistics*. Academic Press.
6. Rice, J. A. (2006). *Mathematical Statistics and Data Analysis*. Singapore: Thomson-Duxbury.

STOE 552 - QUESTIONNAIRE DESIGN AND SAMPLE SELECTION

Teaching Hours: 3 Hrs per week

Learning Objectives:
LO1: To design questionnaires in a way that make them attractive towards target respondents.
LO2: To explore different methods of primary data collection.
LO3: To understand the practical implementation of various sampling techniques.

Course Outcomes:
CO1: Ability to formulate a relevant questionnaire for real life problems.

CO2: Able to understand and implement different methods of data collection through questionnaire.
CO3: Ability to implement relevant sampling techniques in real life.

UNIT-I **14 Hrs**
Introduction, qualities of a good questionnaire, types of questionnaires: exploratory questionnaire (qualitative) and formal standardized questionnaire (quantitative). Questionnaire question types: open-ended questions, multiple choice questions, dichotomous questions, scaled questions, and pictorial questions; questions to avoid in a questionnaire. Steps involved in the development of a questionnaire, checking the validity and reliability of questionnaire.

UNIT-II **13 Hrs**
Methods of reaching target respondents: personal interviews, group or focus interviews, mailed questionnaires, and telephone interviews. Advantages and disadvantages of questionnaires. Examples of questionnaires. Introduction to pilot surveys and its use in questionnaire development and modification.

UNIT-III **13 Hrs**
Sampling: introduction, techniques: probability sampling - simple random, systematic, stratified, and cluster sampling; non-probability sampling - convenience, quota, judgement, and snowball sampling. Sampling and non-sampling errors. Applications of sampling. Sample size determination.

References:

1. http://www.fao.org/3/w3241e/w3241e05.htm
2. https://www.kyleads.com/blog/questionnaire/
3. https://www.digitalvidya.com/blog/sampling-techniques/

STOE 553 - DATA VISUALIZATION

Teaching Hours: 3 Hrs per week

Learning Objectives:
LO1: To learn about how to develop simple summaries and exploratory graphs for univariate data.
LO2: To explore different visualization techniques for bivariate data.
LO3: To gain knowledge on multivariate data visualization for real life problems.

Course Outcomes:
CO1: Ability to formulate and visualize univariate data.
CO2: Able to implement various visualization techniques for bivariate data in real life.
CO3: Capable to obtain multidimensional graphs for real life applications.

UNIT-I **12 Hrs**

Tables and Univariate Graphs: Introduction, Tables, Q-Q Plot, Categorical (Bar chart, Pie chart, Tree map), Quantitative (Histogram, Frequency polygon, Frequency Curve, Ogives, Stem and leaf plot, Dot chart). Applications and examples.

UNIT-II **14 Hrs**

Bivariate Graphs: Categorical vs. Categorical (Stacked bar chart, Grouped bar chart, Segmented bar chart), Quantitative vs. Quantitative (Scatterplot, Line plot, Area chart), Categorical vs. Quantitative (Bar chart on summary statistics, Box plots). Applications and examples.

UNIT-III **14 Hrs**

Multivariate Graphs and Statistical Models: Grouping, Faceting, Correlation plots, Linear Regression, Scatter Plot Matrix, Parallel coordinates plot, Star plot, Chernoff faces, regression model with two variables (simple, quadratic and exponential), Growth curve. Applications and examples.

References:

1. https://rkabacoff.github.io/datavis/
2. https://www.analyticsvidhya.com/blog/2015/07/guide-data-visualization-r/
3. https://towardsdatascience.com/a-guide-to-data-visualisation-in-r-for-beginners-ef6d41a34174
4. https://www.tutorialspoint.com/excel_data_analysis/excel_data_analysis_visualization.htm

STHT 551 - PROBABILITY AND DISTRIBUTIONS - II

Teaching Hours: 4 Hrs per week

Learning Objectives:

LO1: To learn necessary theoretical foundations of probability and measure theory.

LO2: To understand the importance of weak law of large numbers.

LO3: To understand the concepts of characteristic function and central limit theorem, and its importance.

LO4: To study about the theoretical aspects and applications of various sampling distributions.

Course Outcomes:

CO1: Ability to understand the conceptual basis for the probability and measure theory.

CO2: Ability to understand the theory of weak law of large numbers.

CO3: Gain knowledge about the theoretical concepts of characteristic function and central limit theorem.

CO4: Acquire knowledge on the theoretical aspects of various sampling distributions.

UNIT-I **16 Hrs**

Measure, probability measure, properties of a measure and probability. Lebesgue and Lebesgue-Stieltjes measure on the real line. Absolute continuity. Monotone convergence theorem, Fatou's lemma and dominated convergence theorem. Borel-Cantelli lemma, convergence in probability, convergence almost surely, convergence in distribution, convergence in r^{th} mean, convergence theorem for expectations. Slutsky's theorem.

UNIT-II **13 Hrs**

Weak law of large numbers - Kolmogorov's generalized WLLN (proof of sufficient condition only), Khintchine's WLLN as special case, Chebyshev's WLLN. Kolmogorov's strong law of large number sequence of independent and iid random variables. Kolmogorov's inequality.

UNIT-III **13 Hrs**

Characteristic function - properties, inversion theorem (statement only and proof for density version), uniqueness theorem, continuity theorem (statement only). Central limit theorem, Lindeberg-Levy and Liapounov central limit theorems. Statement of Lindeberg-Feller form (statement only). Application of these theorems.

UNIT-IV **10 Hrs**

Sampling distributions, chi-square, t, F and non-central chi-square, mgf of non-central chi-square distribution, reproductive property. Non-central t and non-central F – derivation of probability density function and mean.

References:

1. Bhat, B. R. (2018). *Modern Probability Theory* (4rd ed.). New Academic Science Ltd.
2. Mukhopadhyaya, P. (2016). *Mathematical Statistics*. Books and Allied.
3. Pitman, J. (1993). *Probability*. Narosa.
4. Billingsley, P. (1986). *Probability and Measure*. John Wiley and Sons.
5. Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.
6. Ash, R. B. and Catherine Doleans-Dade (2000). *Probability and Measure Theory*. Academic Press.
7. Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*.
8. Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley Eastern.
9. Rohatgi, V. K. and Saleh, A. K. Md. E. (2015). *An Introduction to Probability Theory and Mathematical Statistics*. Wiley Eastern.

STHT 552 - DESIGN AND ANALYSIS OF EXPERIMENTS
Teaching Hours: 4 Hrs per week

Learning Objectives:

LO1: To study about the fundamentals and principles involved in designed experiments.

LO2: To gain knowledge on the theory and applications of complete and incomplete block designs.

LO3: To learn about balanced incomplete block design, fixed, mixed and random effects models.

LO4: To explore the real life applications of factorial experiments, complete and partial confounding, nested designs, split-plot designs, and strip-plot designs.

Course Outcomes:

CO1: Demonstrate necessary theoretical foundations on the fundamentals and principles involved in designed experiments.
CO2: Exhibit theoretical knowledge and real life applications of complete and incomplete block designs.
CO3: Ability to understand the theory and applications of balanced incomplete block design, fixed, mixed and random effects models.
CO4: Ability to understand the importance and applications of various experimental designs such as factorial experiments, complete and partial confounding, nested designs, split-plot designs, and strip-plot designs, in analyzing real life problems.

UNIT-I **13 Hrs**

Gauss-Markov setup, normal equations and least squares estimates, estimable function and estimation space, variance and covariance of least squares estimates, estimation of error variance, estimation with correlated observations, simultaneous estimates of linear parametric functions. Tests of hypothesis for one and more than one linear parametric functions. Confidence intervals and regions, analysis of variance, power of F-test, multiple comparison tests - Tukey and Bonferroni, simultaneous confidence interval.

UNIT-II **13 Hrs**

Introduction to designed experiments, general block design - complete block design, incomplete block design and its information matrix, criteria for connectedness, balance and orthogonality. Intra-block analysis - estimability, best point estimates/interval estimates of estimable linear parametric functions and testing of linear hypotheses, estimation of parameters.

UNIT-III **13 Hrs**

BIBD - definition, concept of connectedness, balancing, properties, estimability, recovery of inter-block information. Analysis of covariance in a general Gauss-Markov model, applications to CRD and RCBD. Fixed, mixed and random effects models, variance components estimation.

UNIT-IV **13 Hrs**

General factorial experiments, factorial effects - best estimates and testing the significance of factorial effects, study of 2^n and 3^n factorial experiments in randomized blocks. Complete and partial confounding. Nested designs. Split-plot, strip plot designs.

References:

1. Bapat, R. B. (2012). *Linear Algebra and Linear Models*. Hindustan Book Agency.
2. Rao, C. R. (2009). *Linear Statistical Inference and its Applications* (2^{nd}ed). Wiley.
3. AlokeDey (1986). *Theory of Block Designs*. Wiley Eastern.
4. Dean, A. and Voss, D. (1999). *Design and Analysis of Experiments*. Springer.
5. Chakrabarti, M. C. (1962). *Mathematics of Design and Analysis of Experiments*. Asia.
6. Cochran and Cox, D. R. (1957). *Experimental Designs*. John Wiley.
7. Das, M. N. and Giri, N. (1979). *Design and Analysis of Experiments*. Wiley Eastern.

8. Giri, N. (1986). *Analysis of Variance*. South Asian Publishers.
9. John, P. W. M. (1911). *Statistical Design and Analysis of Experiments*. Macmillan.
10. Joshi, D. D. (1987). *Linear Estimation and Design of Experiments*. Wiley Eastern.
11. Montgomery, C. D. (2013). *Design and Analysis of Experiments*(8thed). New York: Wiley.
12. Mukhopadhyay, P. (1998). *Applied Statistics*. Books and Allied (P) Ltd.
13. Pearce, S. C. (1984). *Design of Experiments*. New York: Wiley.
14. Rao, C. R. and Kleffu, J. (1988). *Estimation of Variance Components and Applications*. North Holland.
15. Searle, S. R., Casella, G. and McCullugh, C. E. (1992). *Variance Components*. Wiley.
16. N.R. Mohan Madhyast, Ravi and Praveena (2020). A First course in linear models in design of experiments. Springer.

STHT 553 - THEORY OF ESTIMATION

Teaching Hours: 4 Hrs per week

Learning Objectives:

LO1: To study about the fundamentals of estimation theory.

LO2: To learn the theory and practical implementation of minimum variance unbiased estimation and confidence interval construction.

LO3: To explore the concept of consistency of estimators.

LO4: To understand the theory and applications of maximum likelihood estimation.

Course Outcomes:

CO1: Develop necessary theoretical foundations on the fundamental aspects involved in estimation theory.

CO2: Able to understand and apply the concepts associated with minimum variance unbiased estimation and also explore about confidence interval construction.

CO3: Ability to apply the concepts associated with consistency of estimators.

CO4: Gain knowledge about the theory and applications of maximum likelihood estimation.

UNIT-I **13 Hrs**

Parametric models, likelihood function, example from standard discrete and continuous models. Sufficiency, Neyman factorization criterion, Fisher information for single and several parameters. Minimal sufficient statistic, likelihood equivalence. Exponential families and Pitman families. Completeness, Ancillary Statistics, Basu's theorem and applications.

UNIT-II **13 Hrs**

Minimum variance unbiased estimation, unbiasedness, locally unbiased estimators, minimum variance, locally minimum variance, mean squared error, Cramer-Rao lower bound approach. Minimum variance unbiased estimators (MVUE), Rao-Blackwell theorem, completeness, Lehman-Scheffe's theorem, necessary and sufficient condition for MVUE. Bhattacharya bounds (without proof). Introduction to interval estimation, construction of confidence intervals using pivot.

UNIT-III **13 Hrs**

Consistency, estimation of real and vector valued parameters, invariance properties. Consistency of estimators by method of moments and method of percentiles, mean squared error criterion, asymptotic relative efficiency, consistent asymptotic normal (CAN).

UNIT-IV **13 Hrs**

Method of maximum likelihood - notion, MLE in location and scale family, exponential family, Cramer family (statement only). Cramer-Huzurbazar theorem. Solutions to likelihood equations by methods of scoring, Newton-Raphson and other iterative procedures. Fisher lower bound to asymptotic variance, extension to multi-parameter case (without proof).

References:

1. Casella, G. and Berge, R. L. (2002). *Statistical Inference* (2nd ed.). Singapore: Thomson-Duxbury.
2. Kale, B. K. (2000). *A First Course on Parametric Inference*. Narosa Publishing House.
3. Lehman, E. L. and Casella G. (2003). *Theory of Point Estimation*. Springer.
4. Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley Eastern.
5. Rohatgi, V. K. and Saleh, A. K. L. (2001). *An Introduction to Probability and Mathematical Statistics*. Wiley Eastern.
6. Rajagopalan, M. and Dhanavanthan, P. (2012). *Statistical Inference*. Phi Learning Pvt. Ltd.
7. Zacks, S. (1981). *Parametric Statistical Inference*. Pergamon Press.

STST 551 - DATA MANAGEMENT AND STATISTICAL COMPUTING WITH PYTHON

Teaching Hours: 3 Hrs per week

Learning Objectives:

LO1: To learn the fundamental aspects involved in Python programming paradigms.

LO2: To gain knowledge on how to use pandas for data manipulation.

LO3: To study about how to generate data visualizations using Python.

Course Outcomes:

CO1: Be able to gain comprehensive knowledge on the fundamentals involved in Python programming paradigms.

CO2: Exhibit insights on using Pandas in Python required for data manipulation.

CO3: Be able to explore about how to generate powerful data visualizations using Python.

UNIT-I **14 Hrs**

Using Numpy - Basics of NumPy-Computation on NumPy-Aggregations-Computation on Arrays Comparisons, Masks and Boolean Arrays-Fancy Indexing-Sorting Arrays-Structured Data: NumPy’s Structured Array.

UNIT-II **14 Hrs**

Data Manipulation with Pandas - Introduction to Pandas Objects-Data indexing and Selection-Operating on Data in Pandas-Handling Missing Data-Hierarchical Indexing - Combining Data Sets.

UNIT-III **12 Hrs**

Visualization and Matplotlib - Basic functions of matplotlib-Simple Line Plot, Scatter Plot-Density and Contour Plots-Histograms, Binnings and Density-Customizing Plot Legends, Colour Bars-Three-Dimensional Plotting in Matplotlib.

References:

1. VanderPlas, J. (2016). *Python Data Science Handbook - Essential Tools for Working with Data*. O'Reily Media, Inc.
2. Zhang, Y. (2016). *An Introduction to Python and Computer Programming*. Springer Publications.
3. Thareja, R. (2023)*Python Programming using Problem Solving Approach* (2^{nd}ed). Oxford University Press.
4. Grus, J. (2016). *Data Science from Scratch First Principles with Python*. O'Reilly Media.
5. Padmanabhan, T. R. (2016). *Programming with Python*. Springer Publications.

STSP 551 - DESIGN & ANALYSIS OF EXPERIMENTS AND THEORY OF ESTIMATION PRACTICAL

1. One way and two way analysis of variance
2. Least squares estimation
3. Analysis of incomplete block design
4. Analysis of balanced incomplete block design
5. Analysis of covariance
6. Analysis of factorial experiment
7. Analysis of 2^n factorial experiment
8. Analysis of 3^n factorial experiment
9. Confounding in 2^n factorial experiment
10. Split-plot design
11. Estimation by the methods of moments and percentile
12. Uniformly minimum variance unbiased estimator
13. Unbiased estimation
14. Estimation by the method of maximum likelihood - I
15. Estimation by the method of maximum likelihood – II

STSP 552 - DATA MANAGEMENT AND STATISTICAL COMPUTING WITH PYTHON PRACTICAL

1. Programs to perform a number of mathematical operations on arrays such as trigonometric, statistical and algebraic routines.
2. Programs to perform fast vectorized array operations for data munging and cleaning, subsetting and filtering, transformation, and any other kinds of computations.
3. Exercises to handle n-dimensional arrays, broadcasting, performing operations, data generation, etc.
4. Programs to work with arrays, queries, and dataframes.
5. Programs to perform the groupby, merge, and join methods in Pandas.

6. Exercises using matplotlib to generate various plots.
    a. Plot charts, individually and in multiples.
    b. Customize the style and appearance of different plot components.
    c. Choose different chart types based on data type and requirement.
    d. Show the distribution of data.
    e. Customize objects in Matplotlib.
    f. Plot data in 2D and 3D.

III SEMESTER

STOE 601 - INFERENTIAL STATISTICS AND DATA ANALYSIS
Teaching Hours: 3 Hrs per week

Learning Objectives:
LO1: To understand the theory and applications of various parametric and nonparametric hypothesis tests.
LO2: To study the practical implementation of one-way analysis of variance test.
LO3: To explore the real life applications of correlation and regression techniques.

Course Outcomes:
CO1: Able to identify the basics of hypothesis testing and perform parametric and nonparametric hypothesis tests.
CO2: Able to conduct one-way analysis of variance hypothesis test.

CO3: Apply correlation and regression techniques to real life problems.

UNIT-I **16 Hrs**

The concept of hypothesis and tests of hypothesis: null hypothesis, alternate hypothesis, type I and type II errors, test statistic, level of significance, power of the test, p-value. Testing of hypothesis for population means and population proportions, confidence intervals. Nonparametric tests - sign test, Wilcoxon-Mann-Whitney test, Wilcoxon signed rank test. Contingency tables, chi-square test for independence of attributes.

UNIT-II **10 Hrs**

Testing for the equality of several population means. The concept of analysis of variance, one way analysis of variance, its utility in the analysis of survey data and data obtained from designed experiments.

UNIT-III **14 Hrs**

Regression and correlation - bivariate data, correlation, scatter plot, correlation coefficient and its properties, testing for correlation coefficient, rank correlation. Regression - use of simple linear regression model to study the linear relationship between two variables, fitting the simple linear regression model, testing significance of regression coefficient, coefficient of determination.

References:

1. Campbell, R. C. (1989). *Statistics for Biologists* (3rd ed). Cambridge University Press.
2. Chatfield, C. (1981). *Statistics for Technology*. Chapman and Hall.
3. Frank, H. and Athoen, S. C. (1997). *Statistics: Concepts & Applications*. Cambridge University Press.
4. Medhi, J. (2006). *Statistical Methods: An Introductory Text*. New Age International Publisher.
5. Ross, S. M. (2017). *Introductory Statistics*. Academic Press.
6. Rice, J. A. (2006). *Mathematical Statistics and Data Analysis*. Singapore: Thomson-Duxbury.

STOE 602 - CATEGORICAL DATA ANALYSIS
Teaching Hours: 3 Hrs per week

Learning Objectives:

LO1: To gain theoretical knowledge on various statistical procedures used for categorical data analysis.

LO2: To explore different statistical approaches used for analyzing two-way and three-way tables.

LO3: To study about statistical models for analyzing binary, polytomous and multivariate categorical responses.

Course Outcomes:

CO1: Develop conceptual understanding and application of statistical procedures for analyzing categorical data.
CO2: Able to identify designs of contingency tables and recommend appropriate measures of association and statistical tests.
CO3: Ability to develop models for binary, polytomous and multivariate categorical responses, interpret results regardless of model parameterization.

UNIT-I **15 Hrs**
Introduction and Probability distributions: Meaning of categorical data, Scales of measurement, categorical methods, probability distributions for categorical variables. Frequency distribution tables for discrete variables, Bernoulli, binomial, Poisson, hypergeometric and multinomial distributions. Hypothesis testing for a single proportion, Confidence intervals for a single proportion, Goodness-of-fit: comparing distributions for a single discrete variable.

UNIT-II **15 Hrs**
Analyzing Contingency Tables: Probability Structure for Contingency Tables, Comparing Proportions in 2×2 Contingency Tables, The Odds Ratio, Chi-Squared Tests of Independence, Testing Independence for Ordinal Variables. Contingency tables for three categorical variables, Marginal, conditional and independence, analysis of three-way tables.

UNIT-III **10 Hrs**
Generalized Linear Models: Components of a Generalized Linear Model, Generalized Linear Models for Binary Data, Generalized Linear Models for Counts and Rates.

References:

1. Agresti, A. (2019). *An Introduction to Categorical Data Analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons, 2019. Series: Wiley Series in Probability and Statistics, ISBN 9781119405269.
2. Azen, R. and Walker, C. M. (2021). *Categorical Data Analysis for the Behavioral and Social Sciences* (2nd ed). Routledge, Taylor & Francis Group, New York, ISBN 978-1-84872-836-3.

STOE 603 - DEMOGRAPHIC METHODS AND ANALYSIS
Teaching Hours: 3 Hrs per week

Learning Objectives:
LO1: To study about the basic demographic indicators and explore various sources of demographic data.
LO2: To understand the key measures and techniques used in studying population behavior and change.
LO3: To gain knowledge on the theory and applications of life table, and concept of migration and its types.

Course Outcomes:

CO1: Understand the basic demographic indicators and explore the different sources of demographic data.

CO2: Develop conceptual understanding and applications of the key measures and techniques used in studying population behavior and change.

CO3: Understand the theory and applications of life table, along with the concept of migration and its types.

UNIT-I **12 Hrs**

Demography: introduction, purpose, nature of demographic information: births, fertility, fecundity, deaths, mortality, life expectancy. Data collection methods: census, sample surveys, registration of vital events, population registers, and administrative records.

UNIT-II **16 Hrs**

Measurement of population, measures of fertility: crude birth rate, age-specific fertility rate, general fertility rate, and total fertility rate; measures of mortality: crude death rate, age-specific death rate, standardized death rate, infant mortality rate, neo-natal mortality rate, and maternal mortality rate.

UNIT-III **12 Hrs**

Life table and population projection: Life table, components of life table, force of mortality, expectation of life tables, types of life tables, complete and abridged life tables, uses of life tables. Basic concepts and definition of population change, migration. Types of migration-internal and international, factors affecting migration.

References:

1. Yusuf, F., Martins, J. M., and Swanson, D. A. (2016). *Methods of Demographic Analysis*. Springer, New York, London.
2. Carmichael, G. A. (2016). *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*. Springer, New York, London.
3. Ram Kumar R. (2018). *Technical demography.* New Age International.
4. Srinivasan K. (1998). *Basic demographic techniques and applications.* Sage publication, New Delhi.

STHT 601 - TESTING OF HYPOTHESES

Teaching Hours: 4 Hrs per week

Learning Objectives:

LO1: To gain theoretical knowledge on the fundamentals involved in testing of hypothesis theory.

LO2: To understand the theory and practical implementation of the concepts such as uniformly most powerful, uniformly most powerful unbiased, and likelihood ratio tests.

LO3: To study about the theoretical aspects and applications of interval estimation.

LO4: To explore the real life applications of various nonparametric tests.

Course Outcomes:

CO1: Show acquisitions of adequate foundations on the fundamentals involved in testing of hypothesis and understand its importance.

CO2: Show learning about the theoretical aspects of uniformly most powerful, uniformly most powerful unbiased, likelihood ratio tests, and its implementation in practical problems.

CO3: Develop knowledge about the theoretical aspects of interval estimation and its implementation in practical problems.

CO4: Exhibit knowledge on various nonparametric tests and its applications in real life problems.

UNIT-I **13 Hrs**

Framing of null hypothesis, critical region, level of a test, randomized and non-randomized tests, two kinds of error, size of a test, p-value, power function. Most powerful tests in class of size α test, Neyman-Pearson lemma, MP test for simple null against simple alternative hypothesis. UMP tests for one sided null against one sided alternatives, monotone likelihood ratio property. Extension of these results in Pitman family when only upper or lower end points depend on the parameter.

UNIT-II **16 Hrs**

Non-existence of UMP test. Neyman-Pearson generalized lemma (statement only), concept of UMP for simple null against two sided alternatives in one parameter exponential family and UMPU tests with application to one parameter exponential family, UMP for two sided null (statement only). Likelihood ratio test (LRT), asymptotic distribution of LRT statistic, Pearson's chi-square test for goodness of fit, Bartlett's test for homogeneity of variances, large sample tests.

UNIT-III **10 Hrs**

Interval estimation, confidence level, construction of confidence intervals by inverting acceptance region. Shortest expected length confidence interval, evaluating interval estimators using size and coverage probability and test related optimality, uniformly most accurate one-sided confidence interval and its relations to UMP test for one sided null against one sided alternative hypothesis.

UNIT-IV **13 Hrs**

U-statistics, properties and asymptotic distributions (in one and two sample case). Nonparametric tests: One sample test - test based on total number of runs, the ordinary sign test, the Wilcoxon signed rank test, the Kolmogorov-Smirnov one sample goodness of fit test. Two sample tests - the median test, the Wilcoxon-Mann-Whitney test, Kolmogorov-Smirnov two sample test.

References:

1. Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Wadsworth Group.
2. Gibbons, J. D. (1971). *Nonparametic Inference*. McGraw Hill.
3. Kale, B. K. (2000). *A First Course on Parametric Inference*. Narosa Publishing House.
4. Lehmann, E. L. and Romano, J. (2008). *Testing Statistical Hypotheses*. John Wiley.
5. Pratt, T. W. and Gibbons, J. D. (2012). *Concepts of Nonparametric Theory*. Springer.

6. Rao, C. R. (2009). *Linear Statistical Inference and its Applications* (2nd ed). Wiley Eastern.
7. Rohatgi, V. K. and Saleh, A. K. L. (2001). *An Introduction to Probability and Mathematical Statistics*. Wiley Eastern.
8. Rajagopalan, M. and Dhanavanthan, P. (2012). *Statistical Inference*. Phi Learning Pvt. Ltd.

STHT 602 - REGRESSION ANALYSIS
Teaching Hours: 4 Hrs per week

Learning Objectives:
LO1: To explore the theory and real life applications of linear regression techniques and importance of regression diagnostics.
LO2: To understand the consequences of the problem of multicollinearity in linear regression.
LO3: To learn about the problems of heteroscedasticity and autocorrelation, and understand its consequences.
LO4: To gain theoretical knowledge on the concepts of simultaneous equation models and identification problem.

Course Outcomes:
CO1: Show an acquisition of necessary theoretical foundations on linear regression techniques, model adequacy checking and regression diagnostics, and its extensive use in data analysis.
CO2: Able to understand the consequences of multicollinearity and work on its remedies.
CO3: Able to test and understand the consequences of heteroscedasticity and autocorrelation, and work on its remedies.
CO4: Familiarity with the theoretical aspects of simultaneous equation models and identification problem.

UNIT-I **16 Hrs**
Simple linear regression, multiple linear regression, basic assumptions, ordinary least squares (OLS) - estimation and their properties, tests of hypothesis about regression coefficients, likelihood ratio criterion. Dummy variables. Prediction - best linear unbiased predictor. Regression diagnostics and specification tests - residual analysis for identifying influential observations, recursive residuals and their applications, specification tests, subset selection of explanatory variables, Mallows C_p statistic.

UNIT-II **12 Hrs**
Violation of basic ideal conditions - disturbance with non-zero mean, asymptotically uncooperative regressor. Multicollinearity - its consequences and testing. Ridge estimator and its properties, ridge regression. Stochastic regressor, autoregressive models, instrumental variables, errors in variables. Distributed lag models.

UNIT-III **12 Hrs**
Heteroscedasticity - tests for heteroscedasticity. Generalized least squares (GLS) estimators and its properties, feasible generalized least squares estimators. Grouping of observations. Sets

of Regression Equations. Auto correlation - its consequences and testing for autocorrelation, estimation and prediction.

UNIT-IV **12 Hrs**

Simultaneous equation models. Identification problem, identification using linear homogeneous restrictions on structural parameters, rank and order conditions, estimation in simultaneous equation models. Indirect least squares, two stage least squares, structural equation modelling.

References:

1. Cook, R. D. and Weisberg, S. (1982). *Residual and Influence in Regression*. London: Chapman and Hall.
2. Draper, N. R. and Smith, H. (2019). *Applied Regression Analysis* (3rd ed.). Wiley India Exclusive.
3. Gunst, R. F. and Mason, R. L. (1980). *Regression Analysis and its Application - A Data Oriented Approach*. Marcel Dekker.
4. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). *Introduction to Linear Regression Analysis*. John Wiley.
5. Ryan, T. P. (1997). *Modern Regression Methods*. New York: John Wiley.
6. Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis* (2nd ed.). New York: John Wiley.
7. Fomby, T. B., Hill, C. R. and Johnson, S. R. (2012). *Advanced Econometric Methods*. Springer.
8. Greene, W. H. (2002). *Econometric Analysis* (5th ed.). New York: Prentice Hall.
9. Johnston, J. and Dinardo, J. (1996). *Econometric Methods* (4th ed.). McGraw-Hill.
10. Maddala, G. S. (1992). *Introduction to Econometrics* (2nd ed.). New York: Macmillan.
11. Gujarati, D. N. (2004). Basic Econometrics (4th ed.). McGraw-Hill.

STHT 603 - MULTIVARIATE ANALYSIS
Teaching Hours: 4 Hrs per week

Learning Objectives:

LO1: To gain knowledge on the theoretical foundations required for analyzing multivariate data.

LO2: To explore different inferential techniques used for analyzing multivariate data.

LO3: To understand the theory and applications of multivariate techniques such as principal component analysis, canonical correlation analysis, and factor analysis.

LO4: To learn the theoretical aspects of discrimination and classification, and cluster analysis, and understand its applications in real life.

Course Outcomes:

CO1: Show an acquisition of necessary theoretical foundations for analyzing multivariate data.

CO2: Exhibit theoretical knowledge on various inferential techniques used for analyzing multivariate data.
CO3: Develop theoretical knowledge about multivariate techniques such as principal component analysis, canonical correlation analysis, factor analysis, and its applications in real life problems.
CO4: Gain conceptual understanding about multivariate techniques such as discrimination and classification, cluster analysis, and its applications in real life problems.

UNIT-I **13 Hrs**

Nature of a multivariate problem, main types of multivariate problems, objectives of multivariate analysis. Organization of multivariate data, descriptive statistics, visualization techniques. Multivariate normal distribution properties, maximum likelihood estimators of the parameters. Independence of sample mean vector and sample covariance matrix. Assessing the assumptions of normality Q-Q plot, chi-square plot, transformations to near normality.

UNIT-II **13 Hrs**

Inference problems in multivariate normal distribution, Hotellings T^2, Mahalanobis D^2 statistics, likelihood ratio tests for collinearity, q-sample problem. Roy's union and intersection test. Test for symmetry. Confidence regions, simultaneous confidence statements. Independence of subvectors, sphericity test. Wishart matrix, statement of Wishart distribution, its properties and applications.

UNIT-III **13 Hrs**

Principal component analysis (PCA) - definition and properties, graphing the principal components, sample principal components, interpretation of zero, small and repeated eigenvalues, component loadings and component correlations, the problem of scaling, tests of hypotheses. Canonical correlation analysis - canonical variates and canonical correlations, sample canonical variates, sample canonical correlations, inference problems. Factor analysis - orthogonal factor model, factor loadings, estimation of factor loadings, factor scores.

UNIT-IV **13 Hrs**

Classification and discrimination problems - concepts of separation and classification, Bayes and Fisher's criteria, classification rules based on expected cost of misclassification (ECM) and total probability of misclassification (TPM), classification with two multivariate normal populations (equal and unequal covariance matrices), evaluating classification rules, classification with several populations, Fisher's linear discriminant function, tests associated with discriminant functions. Cluster Analysis - distances and similarity measures, hierarchical clustering methods, k-means method.

References:

1. Anderson, T. W. (1984). *An Introduction to Multivariate Analysis* (2nd ed.). John Wiley.
2. Flury, B. (2010). *A First Course in Multivariate Statistics*. Springer Texts in Statistics.
3. Kshirasagar, A. M. (1972). *Multivariate Analysis*. Marcel Dekker.
4. Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivatiate Analysis*. Academic Press.
5. Rao, C. R. (2009). *Linear Statistical Inference and its Applications*. Wiley Eastern.

6. Johnson, R. A. and Wichern, D. W. (1986). *Applied Multivariate Statistical Analysis* (6th ed.). Prentice Hall of India.
7. Rencher, A. C. (2003). *Methods of Multivariate Analysis*. Wiley.

STST 601 - STOCHASTIC PROCESSES
Teaching Hours: 3 Hrs per week

Learning Objectives:
LO1: To gain knowledge on the necessary theoretical foundations of stochastic process.
LO2: To understand the theory and applications of Poisson process.
LO3: To study about the theoretical aspects of renewal theorem.
LO4: To explore the real life applications of branching process and understand its importance.

Course Outcomes:
CO1: Acquire necessary theoretical knowledge on the foundations of stochastic processes.
CO2: Explore the theoretical aspects associated with Poisson process and its real life applications.
CO3: Demonstrate conceptual understanding and applications of renewal theory.
CO4: Gain knowledge about the theoretical aspects of branching process and its applications.

UNIT-I **10 Hrs**
Introduction to stochastic processes - classification according to state space and time domain. Stationary process - weakly stationary and strongly stationary processes. Countable state Markov chains (MCs), Chapman Kolmogorov equations, calculation of n-step transition probability and its limit. Stationary distribution, classification of states, random walk and gamblers ruin problem, estimation of TPM for finite states of MC.

UNIT-II **10 Hrs**
Discrete state space continuous time MC, Kolmogorov-Feller differential equations, Poisson process, birth and death process, applications to queues. Wiener process as a limit of random walk, first passage time and other problems.

UNIT-III **10 Hrs**
Renewal theory - elementary renewal theorem and applications. Statement and uses of key renewal theorem, study of residual life time process.

UNIT-IV **10 Hrs**
Branching process - Galton-Watson branching process, probability of ultimate extinction, distribution of population size. Martingale in discrete time - definition and elementary properties, convergence theorem, applications.

References:

1. Basu, A. K. (2003). *Introduction to Stochastic Processes*. Narosa Publications.
2. Bhat, B. R. (2000). *Stochastic Models: Analysis and Applications*. New Age International.
3. Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*. Academic Press.
4. Medhi, J. (2020). *Stochastic Processes*. New Age International Publishers.
5. Ross, S. M. (2008). *Stochastic Processes*. Wiley.
6. Lawler, G. F. (2006). *Introduction to Stochastic Processes* (2nd ed.). Chapman and Hall.

STSP 601 - TESTING OF HYPOTHESES AND REGRESSION ANALYSIS PRACTICAL

1. Size and power of the test
2. Most powerful test
3. Uniformly most powerful test for discrete and continuous distributions
4. Large sample tests
5. Uniformly most powerful unbiased test
6. Interval estimation
7. Nonparametric tests
8. Simple linear regression analysis
9. Multiple linear regression analysis
10. Regression diagnostics
11. Backward elimination and forward selection procedures
12. Multicollinearity diagnostics
13. Heteroscedasticity
14. Autocorrelation
15. Simultaneous equation models.

STSP 602 - MACHINE LEARNING WITH PYTHON PRACTICAL

1. Compare the performance of the OLS, ridge and the Lasso models on the movie collection test data. Establish which of these methods is giving the lowest test error.
2. Hyperparameter optimization using random search and grid search.
3. Implementing different types of encoding for categorical data.
4. Implement logistic regression model with scikit-learn.
5. Implementation of logistic regression in python - step by step.
6. Implementation of KNN classifier using scikit-learn.
7. Multiple linear regression from scratch using python.
8. Implement decision tree and random forest models, compare the performance in python using scikit-learn.
9. Build PCA from scratch in python and also using scikit-learn and fit the model for the transformed data.
10. Perform factor analysis using python.
11. Perform linear discriminant analysis (LDA) in python.

12. Perform cluster analysis using python.

IV SEMESTER

STHT 651 - TIME SERIES ANALYSIS

Teaching Hours: 4 Hrs per week

Learning Objectives:
LO1: To study the fundamentals involved in the theory of time series analysis.
LO2: To gain knowledge on the theory and applications of MA, AR, ARMA and ARIMA models.
LO3: To learn about spectral density function and the estimation process involved in MA, AR and ARMA models.
LO4: To explore the practical implementation of Box-Jenkins approach, exponential and Holt Winters smoothing techniques to model and forecast time series data.

Course Outcomes:
CO1: Develop necessary theoretical foundations on the fundamental concepts involved in time series analysis.

CO2: Gain theoretical knowledge on various time series models such as MA, AR, ARMA and ARIMA, and its real life applications.
CO3: Understand the concept of spectral density function and also regarding the estimation process involved in MA, AR and ARMA models.
CO4: Implement Box-Jenkins approach, exponential and Holt Winters smoothing techniques to model and forecast time series data empirically.

UNIT-I **13 Hrs**

Simple descriptive techniques - time series plots, trend, seasonal effect. Tests for trend and seasonality - estimation and elimination of trend and seasonal components. Exponential and moving average smoothing. Time series as discrete parameter stochastic process. Stationarity, autocovariance and autocorrelation function and their properties. Partial autocorrelation function.

UNIT-II **13 Hrs**

Probability models - White noise model, random walk, linear processes, moving average (MA), autoregressive (AR), ARMA and ARIMA models, invertibility, ACF and PACF of these processes. Spectral properties of stationary models - periodogram, spectrum.

UNIT-III **13 Hrs**

Spectral density function - estimation of spectral densities of AR, MA and ARMA models. Sample ACF and PACF for model identification. Model building - estimation of mean, autocovariance function and autocorrelation function. Estimation in AR models, Yule-Walker equations, estimation in MA model and ARMA models. Order selection in AR and MA models.

UNIT-IV **13 Hrs**

Forecasting - forecast mean square error (FMSE), least squares prediction, BLUP, innovation algorithm. Box-Jenkins forecasting for ARMA Models. Forecasting through exponential smoothing and Holt Winters smoothing. Residual analysis and diagnostic checking. Non-stationary time series models and their identification. Introduction to ARCH and GARCH models.

References:

1. Box, G. E. P., Jenkins, G. M., Reinsel G. C. and Ljung G.M. (2015). *Time Series Analysis: Forecasting and Control* (5^{th}ed). Wiley.
2. Brockwell, P. J. and Davis, R. S. (2002). *Introduction to Time series and Forecasting* (2nd ed.). Springer.
3. Chatfield, C. (1996). *The Analysis of Time Series: An Introduction*. Chapman Hall.
4. Kendall, M. G. and Ord, J. K. (1990). *Time Series* (3rd ed.). Edward Arnold.
5. Montgomory, D. C. and Johnson, D. A. (1977). *Forecasting and Time Series Analysis*. McGraw Hill.
6. Tanaka, K. (1996). *Time Series Analysis*. Wiley Series.
7. Tsay, R. S. (2005). *Analysis of Financial Time Series*. John Wiley & Sons.
8. Shumway, R. H. and Stoffer, D. S. (2017). *Time series analysis and its applications.*Springer.

STHT 652 - RELIABILITY AND SURVIVAL ANALYSIS

Teaching Hours: 4 Hrs per week

Learning Objectives:

LO1: To understand the theoretical aspects involved in the concept of coherent structure.

LO2: To learn the fundamental concepts of reliability theory and understand its real life applications.

LO3: To learn the importance and applications of different censoring situations in survival analysis.

LO4: To explore different lifetime models and understand its applications in medical studies.

Course Outcomes:

CO1: Understand the concept of coherent structure, its representation and importance in industrial applications.

CO2: Show an acquisition of necessary theoretical foundations on the fundamentals of reliability theory and its applications in real life problems.

CO3: Understand different censoring situations and its implementation in the theoretical concepts of survival analysis.

CO4: Show an increased learning about statistical lifetime models used in medical sciences and also regarding the estimation of survival function.

UNIT-I **13 Hrs**

Coherent structures, representation of coherent systems in terms of paths and cuts, duals systems, modules of coherent systems. Reliability of system of independent of components, association of random variables, bounds on system reliability, improved bounds on system reliability using modular decompositions, lifetime distribution of k out of n system.

UNIT-II **13 Hrs**

Measures of reliability, survival/failure rate, hazard function, cumulative hazard function, lack of memory property, graphs of the system reliability functions. Notion of aging, life distributions of coherent systems, classes of life distributions - parametric and nonparametric models, mean residual lifetime with survival function. NBU, NBUE, NWU, NWUE classes of life distributions and their implications.

UNIT-III **13 Hrs**

Complete and censored samples, type I, II and random censoring, life distributions - Exponential, Gamma, Weibull, Lognormal, Pareto family. Estimation of parameter for exponential and gamma distribution under various censoring situations. Confidence interval for parameters of Exponential, Weibull, and Lognormal distributions. Wald, Score and LR tests for Exponential against Gamma and Weibull.

UNIT-IV **13 Hrs**

Estimation of survival function - Kaplan-Meier estimator, Nelson-Aalen estimator, Greenwoods formula. Other life table estimators. Actuarial method of estimation of survival function. Semi-parametric regression for failure rate, Cox's proportional hazards model with one and more number of covariates, log likelihood function, log linear hazards, test for regression coefficients with and without ties.

References:

1. Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart and Winston Inc.
2. Barlow, R. E. and Proschan, F. (1996). *Mathematical Theory of Reliability*. John Wiley.
3. Tobias, P. A. and Trindane, D. C. (1995). *Applied Reliability* (2nd ed.). CRC Press.
4. Lawless, J. R. (1982). *Statistical Models and Methods for Lifetime Data*.
5. Bain, L. J. and Engelhardt (1991). *Statistical Analysis of Reliability and Life Testing Data*.
6. Zacks, S. (1992). *Introduction to Reliability Analysis: Probability Models and Statistical Methods*. Springer.
7. Cox, D. R. Oakes, D. (1984). *Analysis of Survival Data*. New York: Chapman and Hall.
8. Kalbfeish, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (2nd ed.). John Wiley & Sons, Inc.
9. Deshpande, J. V. and Purohit, S. G. (2005). *Lifetime Data: Statistical Models and Methods*. World Scientific.

STST 651 - STATISTICAL MODELLING

Teaching Hours: 3 Hrs per week

Learning Objectives:

LO1: To study the fundamentals of Bayesian theory and understand its real life applications.

LO2: To gain theoretical knowledge on statistical concepts such as nonparametric density estimation, nonparametric regression and resampling techniques.

LO3: To explore the real life applications of advanced regression techniques such as logistic, multi-logit, count data, and log linear regression.

Course Outcomes:

CO1: Explain the Bayesian framework for data analysis and demonstrate when the Bayesian approach can be beneficial.

CO2: Understand the importance of some advanced statistical concepts such as nonparametric density estimation, nonparametric regression and resampling techniques.

CO3: Exhibit theoretical knowledge on some advanced regression techniques such as logistic, multi-logit, count data, and log linear regression and understand its applications.

UNIT-I **15 Hrs**

Introduction to Bayesian theory and philosophy - loss function and risk, foundations of optimal decision making, Bayes rule, minimax rule, admissibility. Prior and posterior distributions, conjugate families, non-informative priors - uniform and Jeffrey's prior. Bayesian estimation. Introduction to credible sets, Bayesian hypothesis testing, Bayesian prediction.

UNIT-II **12 Hrs**

Nonparametric density estimation (kernel based), nonparametric regression techniques - kernel, nearest neighbour, local polynomial (LOESS regression) and spline based methods. Concept of resampling techniques - bootstrap and jackknife methods. Bootstrap procedure - hypothesis testing and bootstrap interval estimation.

UNIT-III **13 Hrs**

Introduction to generalized linear model (GLM), logistic regression, multi-logit regression, count data regression, log linear regression.

References:

1. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer-Verlag.
2. Ghosh, J. K., Delampady, M. and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer Texts in Statistics.
3. Box, G. E. P. and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. Wiley Classics Library.
4. Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.
5. Hardle, W. (1991). *Smoothing Techniques*. Springer Science & Business Media.
6. Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer Science & Business Media.
7. Wasserman, L. (2005). *All of Nonparametric Statistics*. Springer Science & Business Media.
8. Dobson, A. J. (1983). *Introduction to Statistical Modelling*. Chapman and Hall.
9. Agresti, A. (1990). *Categorical Data Analysis* (3rd ed.). Wiley.
10. Myers, R. H., Montgomery, D. C., Vining, G. G. and Robinson, T. J. (2010). *Generalized Linear Models: with Applications in Engineering and the Sciences* (2nd ed.). John Wiley & Sons.
11. Davison, A. C. and Hinkley, D. V. (1991). *Bootstrap Methods and their Application*. Cambridge University Press.
12. Higgins, J. J. (2004). *An Introduction to Modern Nonparametric Statistics*. Brooks/Cole.

STST 653 - BIG DATA ANALYTICS

Teaching Hours: 3 Hrs per week

Learning Objectives:

LO1: To impart knowledge in Fundamentals, Big Data Analytics, Technologies and databases.

LO2: Learn the main components of BI platforms, their capabilities, and understand the competitive landscape of BI platforms, and data mining techniques.

LO3: To learn the concepts, principles and architecture of data warehousing and the algorithms used to facilitate interactive and exploratory analysis.

LO4: Enables the students to identify the problems where artificial intelligence is required and the different methods available, and also to build and train neural network architectures and data mining techniques.

Course Outcomes:

CO1: Ability to identify what are and what are not big data problems and be able to recast big data problems as data science questions, and understand major components of the Hadoop ecosystem.

CO2: Ability to gain knowledge and skills on BI concepts and tools, and also the concepts of data mining along with the basic methodologies and applications.

CO3: Ability to learn the fundamental concepts of data warehousing and also tools and techniques used to explore and analyze data, and gain knowledge on various algorithms and methods such as clustering, classification and pattern mining.

CO4: Ability to use standard neural networks and natural language processing techniques, apply various optimization algorithms, and implement the same using python.

UNIT-I **10 Hrs**

Introduction to Big Data - Classification of Digital Data, Characteristics of Data, Evolution of Big Data, Challenges with Big Data, Business Intelligence Vs Big Data, A Typical Data Warehouse Environment. Big Data Analytics - Introduction to Big Data Analytics, Classification of Analytics, Importance of Big Data Analytics, Data Science, Terminology used in Big Data Environment, Hadoop - Hadoop Distributed File System Basics, Hadoop MapReduce Framework, MapReduce Programming, Hadoop Essential Tools - Apache HIVE, Apache PIG, Sqoop, Apache Flume.

UNIT-II **10 Hrs**

Business Intelligence Concepts and Applications: Types of BI Tools and Software, Benefits of Business Intelligence, Applications in an Enterprise. Data Mining: Data Pre-processing – Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization. Data Visualization Tools and Techniques.

UNIT-III **10 Hrs**

Data Warehousing and Online Analytical Processing: Data Warehouse Basic Concepts, Data Warehouse Modeling, Data Warehouse Design and Usage. Outlier Detection Methods and Statistical Approaches. Classification Methods: Decision Tree Induction, Bayes Classification Methods, Model Evaluation and Selection, Techniques to Improve Classification Accuracy.

UNIT-IV **10 Hrs**

Artificial Neural Networks: Bayesian Belief Networks, Classification by Backpropagation, Support Vector Machines. Text Mining: Natural Language Processing, Sentiment Analysis.

References:

1. Acharya, S. and Chellappan, S. (2015). *Big Data and Analytics*. Wiley Publications.
2. Eadline, D. (2016). *Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem* (1st ed.). Pearson Education. ISBN-13: 978-9332570351.
3. Maheshwari, A. (2017). *Data Analytics* (1st ed.). McGraw Hill Education. ISBN-13: 978-9352604180.

4. White, T. (2015). *Hadoop: The Definitive Guide*. (4th ed.). O'Reilly Media. ISBN-13: 978-9352130672.
5. Lam, C. *Hadoop in Action*. Manning. ISBN 9781935182191.
6. Jiawei Han, Elsevier (2012). *Data Mining Concepts and Techniques*'

STSP 651 - TIME SERIES ANALYSIS, RELIABILITY AND SURVIVAL ANALYSIS AND STATISTICAL MODELLING PRACTICAL

1. Time series plots and elimination of trend and seasonality.
2. Estimation of ACF and PACF (with and without inbuilt functions)
3. Modelling Time Series – I
4. Modelling Time Series – II
5. Confidence interval for the parameter of exponential distribution under random censoring
6. Confidence interval for the parameters of Weibull distribution under random censoring
7. Plotting empirical distribution and empirical survival functions
8. Nonparametric estimation of survival function using Kaplan-Meier estimator
9. Cox proportional hazards (PH) model
10. Bootstrap and Jackknife resampling
11. Logistic regression analysis
12. Multinomial logistic regression analysis
13. Poisson regression analysis
14. Log-linear model for two-way contingency table
15. Log-linear model for three-way contingency table

STST 654 - ARTIFICIAL INTELLIGENCE (3 Credits)

Learning Objectives:
This course enables the students to identify the problems where artificial intelligence is required and the different methods available.

UNIT-I **10 Hrs**
What is artificial intelligence? Problems, Problem Spaces and search, Heuristic search technique.

UNIT-II **10 Hrs**
Knowledge Representation Issues, Using Predicate Logic, representing knowledge using Rules, Symbolic Reasoning under Uncertainty.

UNIT-III **10 Hrs**
Statistical reasoning, Weak Slot and Filter Structures, Strong slot-and-filler structures.

UNIT-IV **10 Hrs**

Game Playing, Natural Language Processing, Learning.

References:

1. Rich, E., Knight, K. and Nair, S. B. (2017).*Artificial Intelligence* (3rd ed.). McGraw Hill.
2. Rusell, S. and Norvig, P. (2022).*Artificial Intelligence: A Modern Approach* (2nd ed.). Pearson Education.
3. Patterson, D. W. (2015). *Introduction to Artificial Intelligence and Expert Systems*. Pearson Education India.
4. Luger, G. (2002). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (4th ed.). Pearson Education.
5. Rolston, D. W. *Artificial Intelligence and Expert Systems Development*. McGraw Hill.
6. Padhy, N. P. (2015). *Artificial Intelligence and Intelligent Systems*. Oxford University Press.

STST 502 - ELEMENTS OF STATISTICAL COMPUTING (3 Credits)

Learning Objectives:

LO1: To learn about the concept of random number generation and explore different techniques used for generating random numbers.

LO2: To explore the different optimization techniques and its implementation in real-life.

LO3: To gain knowledge on different methods used for computing integrals.

LO4: To provide foundations for statistical simulation and validation of models.

Course Outcomes:

CO1: Ability to execute different methods used for generating random numbers.

CO2: Gain knowledge regarding the implementation of different optimization techniques in real-life.

CO3: Demonstrate conceptual understanding and execution of different techniques for computing integrals.

CO4: Develop knowledge about simulation and validation of statistical models.

UNIT-I **10 Hrs**

Random number generation, requisites of a good random number generator, methods of random number generation such as linear congruential, mixed congruential and multiplicative congruential. Testing of random number generator, run test, Kolmogrov-Smirnov test, sign test, rank test, gap test, digit frequency test and serial correlation, selection of a random number generator. Methods of generating random observations such as inverse transforms, composition, convolution and acceptance-rejection.

UNIT-II **10 Hrs**

Simple optimization method, direct search, grid search, interpolatory search, gradient search. Newton-Raphson method, Muller's method, Aitken's extrapolation, simple problems and applications.

UNIT-III **10 Hrs**

Methods to compute integrals - quadrature formula, double integration, singularity, Gaussian integration. Monte Carlo Methods - Monte Carlo integration and simple case studies, applications of Monte Carlo methods to compute expected values of functions of random variables such as Laplace transform, fourier transform etc., some case studies.

UNIT-IV **10 Hrs**

Approximating probabilities and percentage points in selected probability distribution, verification of WLLN and CLT using random number generator, simulating null distribution of various test statistics, simple applications and case studies.

References:

1. Kennedy, W. J. and Gentle, J. E. (2021). *Statistical Computing*. Routledge.
2. Sen, K. V. (1993). *Numerical Algorithm Computation in Science and Engineering* (2nd ed.). Affiliated East West Press.
3. Law, A. M. and Kelton, W. D. (2000). *Simulation, Modeling and Analysis* (3rd ed.). Tata McGraw Hill.
4. Rajaraman, V. (1993). *Computer Oriented Numerical Methods* (4th ed.). Prentice Hall.
5. Ripley, B. D. (1987). *Stochastic Simulation*. John Wiley.
6. Ross, S. M. (2000). *Introduction to Probability Models*. Academic Press.
7. Ross, S. M. (2013). *Simulation*. Academic Press.
8. Thisted, R. A.(1988). *Elements of Statistical Computing*. Chapman and Hall.

Practicals based on Elements of Statistical Computing

1. Generation of random numbers by acceptance rejection method.
2. Generation of random numbers by linear congruential method.
3. Solution of the equations using iterative methods.
4. Numerical integration.
5. Monte-Carlo integration.
6. Empirical distributions of the test statistics.
7. Applications of CLT.

STST 602 - STOCHASTIC FINANCE (3 Credits)

Learning Objectives:

LO1: To provide foundations on the fundamentals of financial markets and stocks.

LO2: To explore the theory of portfolio and its associated conceptual aspects.

LO3: To understand the concepts of derivatives options and pricing, and its applications in finance.

LO4: To learn about hedging and pricing of derivative instruments and its implementation in finance.

Course Outcomes:

CO1: Understand the fundaments of financial markets and stocks with real-life applications.

CO2: Explain the implementation of portfolio theory in the field of finance.

CO3: Demonstrate conceptual understanding and applications of derivatives options and pricing.

CO4: Develop knowledge about the theoretical aspects of hedging and pricing of derivative instruments and its applications.

UNIT-I **10 Hrs**

Basic concepts of financial markets and stocks, types of traders, forward contracts and futures, call and put options, European option and American options. Interest rates, continuous compounding, present value analysis, bond pricing, risk free interest rates. Returns, gross returns, log returns.

UNIT-II **10 Hrs**

Portfolio theory, mean variance portfolio theory. One risky asset and one risk free asset, two risky assets. Sharpes ratio, tangency portfolio, optional mix of portfolio. Market portfolio, beta, security market line, and capital asset pricing model (CAPM) and their assumption. Value at risk (VAR), nonparametric and parametric estimation of VAR, VAR for a derivative and for a portfolio of assets, delta normal method, simulation of VAR models.

UNIT-III **10 Hrs**

Financial derivatives, options, pricing via arbitrage, law of one price. Risk neutral valuation, arbitrage theorem. Convexity of cost of call option, binomial model single and multiperiod binomial model. Modeling returns - lognormal model, random walk model, modeling through geometric Brownian motion process. Ito lemma (without proof). Arbitrage theorem. The Black Scholes formula and assumptions, properties of the Black Scholes option cost.

UNIT-IV **10 Hrs**

Black Scholes Merton differential equations and assumptions, the delta hedging arbitrage strategy, volatility and estimating the volatility parameter, implied volatility. Pricing American options, pricing of an European option using Monte Carlo and pricing an American option using finite difference methods. Call options on dividend paying securities.

References:

1. Ross, S. M. (2003). *An Elementary Introduction to Mathematical Finance*. Cambridge University Press.
2. Ruppert, D. (2004). *Statistics and Finance: An Introduction*. Springer International Edition.
3. Hull, J. C. (2008). *Options, Futures and Other Derivatives*. India: Pearson Education.
4. Cuthbertson, K. and Nitzsche, D. (2001). *Financial Engineering: Derivatives and Risk Management*. John Wiley & Sons Ltd.
5. Leuenberger, D. G. (1998). *Investment Science*. Oxford University Press.
6. Wilmott, P. (2000). *Quantitative Finance*. John Wiley & Sons.
7. Tsay, R. S. (2005). *Analysis of Time Series*. John Wiley & Sons.

STST 552 - ACTUARIAL METHODS (3 Credits)

Learning Objectives:

LO1: To provide theoretical knowledge on the fundamental aspects of decision theory and actuarial applications.

LO2: To explore about risk models and its real-life applications in the field of actuarial science.

LO3: To understand the theoretical aspects of Bayesian statistics and its applications in actuarial science.

LO4: To gain knowledge on theoretical aspects of actuarial methods and its applications in real-life.

Course Outcomes:

CO1: Develop knowledge on the fundamental aspects of decision theory and actuarial applications.

CO2: Demonstrate conceptual understanding about risk models and its applications in actuarial science.

CO3: Able to understand about the applications of Bayesian statistics in actuarial science.

CO4: Able to explore about some more theoretical aspects of actuarial methods and its applications in real-life.

UNIT-I **10 Hrs**

Review of decision theory and actuarial applications. Loss distributions - modeling of individual and aggregate losses, moments, fitting distributions to claims data, deductibles and retention limits, proportional and excess of loss reinsurance, share of claim amounts, parametric estimation with incomplete information.

UNIT-II **10 Hrs**

Risk models - models for claim number and claim amount in short term contracts, moments, compound distributions, moments of insurer's and reinsurer's share of aggregate claims.

UNIT-III **10 Hrs**

Review of Bayesian statistics - estimation and application. Experience rating - rating methods in insurance and banking, claim probability calculation, stationary distribution of proportion of policyholders in various levels of discount.

UNIT-IV **10 Hrs**

Delay/run-off triangle - development factor, basic and inflation-adjusted chain-ladder method, alternative methods, average cost per claim and Bornhuetter Ferguson methods for outstanding claim amounts, statistical models.

References:

1. Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A. and Nesbitt, C. J. (1997). *Actuarial Mathematics* (2nd ed.). Society of Actuaries.
2. Klugman, S. A., Panjer, H. H., Willmotand, G. E. and Venter, G. G. (1998). *Loss Models: From Data to Decisions*. John Wiley \& Sons.
3. Daykin, C. D., Pentikainen, T. and Pesonen, M. (1994). *Practical Risk Theory for Actuaries*. Chapman Hall.

STST 652 - OPERATIONS RESEARCH (3 Credits)

Learning Objectives:

This course provides theoretical foundations on optimization techniques for managerial decision making process.

UNIT-I **10 Hrs**

Linear programming problem (LPP) - definition, formulation. Simplex method - canonical form, improving non-optimal basic feasible solution (B.F.S), conditions for optimality, conditions for unboundedness. Convex sets, geometry of simplex method extreme point and B.F.S, existence of B.F.S, existence of optimal B.F.S. Two phase method, big M method.

UNIT-II **10 Hrs**

Duality theory of LPP - weak duality theorem and its properties, the fundamental duality theorem, complementary slackness theorem. Dual simplex method. Sensitivity analysis. Integer programming cutting plane technique, Gomory's algorithm for pure integer program. Dynamic Programming, multistage decision making problems, Bellman's principle of optimality, recursive nature of computation, application, applications of dynamic programming.

UNIT-III **10 Hrs**

Inventory theory - nature of inventory problem, motives for carrying inventory, deterministic inventory model with decay. Probabilistic inventory models, continuous review and periodic review systems, (s, S) policy, heuristic solution of lot size reorder point model ((Q, r) policy).

UNIT-IV **10 Hrs**

Queuing theory - characteristics of queues, M/M/1 system, steady state solution, measures of effectiveness, waiting time distributions, Little's formula, M/M/1/K system, M/M/C system, machine interference problem, M/G/1 system, PollaczekKhintchine formula.

References:

1. Gross, D. and Harris, C. M. (1985). *Fundamentals of Queuing Theory* (2nd ed.). John Wiley.
2. Hadley, G. (1975). *Linear and Combinatorial Programming*. John Wiley & Sons.
3. Murty, K. G. (1976). *Linear and Combinatorial Programming*. John Wiley & Sons.
4. Kambo, N. S. (1991). *Mathematical Programming Techniques*. Affiliated East West Press.
5. Taha, H. A. (2001). *Operations Research: An Introduction* (6th ed.). India: Prentice Hall.
6. Sivazlian, B. D. and Stanfel, L. E. (1975). *Analysis of Systems in Operations Research*. Prentice Hall.
7. Daellenbach, H. G. and George, J. A. (1978). *Introduction to Operations Research Techniques*. Allyn and Bacon Inc.

Practicals based on Operations Research

1. Linear Programming Problem
2. Duality of LPP
3. Probabilistic Inventory
4. Queuing Theory

VALUE ADDED COURSES

CERTIFICATE COURSE ON R AND EXCEL FOR DATA SCIENCE

This Course helps you to master in R and Excel, and learn the powerful functions and features to analyze and visualize the data effectively.

Learning Objectives:
1: Learn R language fundamentals and basic syntax.
2: Learn how to program in R, How to use R for effective data analysis.
3: Explore R syntax, functions & packages.
4: Analyze real world challenges in data management; explore general practices of data science.
5: Understand the practicality of excel.
6: Knowledge of formatting, functions & formulas.
7: Learn to use advanced features, graphs & presentation techniques to maximize impact.
8: Perform data cleaning, processing & manipulation techniques using superpower functions & formulas.
9: Build a dashboard / summary report with dynamic charts & tables.

Learning Outcome:
1: Become familiar with the major R data structures.
2: Create your own functions & visualizations.
3: Learn to write your own project syntax ranging from importing data into R to apply standard and more advanced statistical analysis methods.
4: Apply visual elements and advanced formulas to a worksheet to display data in various formats.
5: Learn to use advanced functions & features of excel to improve productivity, enhance spreadsheets with templates, charts, graphics, and formulas and streamline the operational work.

Course Syllabus: R Programming and Excel

Section 1: Getting Started with R and Programming Language Basics **Hours: 7**
a) History of R, Installation of R & R studio, Loading Add-on packages, choosing repositories, Accessing data in packages. b) Help & Documentation – Help for Functions/Packages/Data Sets. c) Data Types - Vectors, Lists, Matrices, Arrays, Factors, Data Frames. d) Variables - Variable Assignment, Finding & Deleting Variables, Data Type of a Variable. e) Operators - Arithmetic, Relational, Logical, Assignment & Miscellaneous Operators. f) Simple Manipulations: Numbers & Vectors, Vectors & Assignment, Vector Arithmetic, Logical Vectors, Character Vectors. g) Generating Sequences & Missing Values. h) Index Vectors: Selecting & Modifying subsets of a data set. i) Objects, their modes & attributes: Intrinsic attributes: mode and length, changing

the length of an object, class of an object.

j) Ordered and unordered factors: A specific example, The function tapply () and ragged arrays.
k) Arrays and matrices: Arrays, Array indexing - Subsections of an array, Index matrices.
l) The array () function: Mixed vector and array arithmetic. The recycling rule, The outer product of two arrays, Generalized transpose of an array.
m) Matrix facilities: Matrix multiplication, Linear equations and inversion, Forming partitioned matrices cbind () and rbind (), The concatenation function c() with arrays, Frequency tables from factors.
n) Lists and Data frames: Lists, Constructing and modifying lists, Concatenating lists.
    a. Data Frames, Making data frames, working with data frames.

Section 2: Functions and Advanced Data Management

Hours: 11

a) Built-in functions (Numeric/Character/Statistical/Other), User define function, Calling function, Defining new binary operators, Assignments within functions, more advanced examples, Applying functions to matrices & data frames
b) Data Input & Output: Changing directories, Managing files & workspace, Reading data from files, writing data from R, Connection to External data sources.
c) View / Edit Data, Objects / Variable types, Converting Objects / Variables, Selecting Variables / Observations.
d) Applying Functions: lappy, sapply, tapply, apply, mapply
e) Combining Variables with c, cbind, rbind functions
f) Combining data with Vector / Matrix / Data Frame / List Function
g) Working with Date & Time
h) Finding NA / NaN& Replacing
i) Conditional Transformation / Decision Making
j) Control Flow (Repetition & Looping / Conditional Execution)
k) Variables: Renaming Variables / Observations, Creating New / Recoding Variables, Keeping & Dropping Variables
l) Generating Random Numbers
m) Data Sets : Stacking/Concatenating/Adding Datasets, Joining / Merging Data Frames
n) Summary: Creating Summarized / Aggregated Datasets (dplyr)
o) Reshaping the data (Reshape Package)
p) Removing Duplicates, Sorting the Data Frames
q) Value Labels or Formats (and Measurement Level)

Section 3: Data Analysis & Data Visualization

Hours: 7

a) Traditional Graphics, Graphics with ggplot2 & Advanced Graph Types.
b) Basic Statistics: Descriptive Statistics, Frequency & Contingency tables, T – tests, Non-parametric tests of group differences
c) Predictive Modeling: Correlation Analysis, Splitting data into training & validation, OLS Regression - Simple / Multiple Linear Regression.
d) Regression diagnostics: Non-Normality, Multicolinearity, Non-linearity, Non-constant error variance.
e) Unusual Observations & Corrective measures: Outliers, High-leverage points & Influential Observations.
f) Choosing Best Regression Model: Comparing Models, Variable selection
g) Model Validation: Cross-Validation
h) Assessment of Regressors: Relative Importance.

Section 4: Introduction and Data Handling in Excel and Data Analysis

Hours: 15

a) Purpose & application of Excel, Understanding the Excel interface - Menu Options, Create & Save Spreadsheets, Save As Formats, Limitations, Insert & delete rows / columns, Printing.
b) Navigation & Editing: Moving around the spreadsheets, Entering information into cells, Types of data, Clipboard, Transformation, Hide rows/columns.
c) Protecting & Sharing: Protect sheet / workbook - Locking cells.
d) Sorting, Filters & Advanced Filters, Remove duplicates, Text to columns, Cell reference.
e) Presentation: Formatting - Cell (Alignment, Height & width, Wrapping, Merging), Numbers (Currency, %, Decimal, negative), Custom Format.
f) Conditional Formatting - Changing the format of the values depends upon the cell value, conditional format formulas
g) Data Cleaning - Extracting / Combining text, for typos & bugs – LEFT (), RIGHT (), LEN (), FIND ()
h) Performing Math with Date & Time: TODAY (), NOW (), DATEVALUE (), YEAR (), MONTH (), DAY (), TEXT ()
i) Lookup & Reference: VLOOKUP, HLOOKUP, INDEX, MATCH
j) Logical Functions: Automatic decision making - IF ELSE, AND, OR, NOT, NESTED IF ELSE.
k) Information Functions : ISERROR, ISBLANK, CELL, ISTEXT
l) Text Functions: TRIM, MID, LOWER, UPPER, PROPER, REPT, TRUNC, CONCATENATE etc…
m) Summarizing Data - SUM () Family, COUNT () Family, AVERAGE, MEDIAN, MIN, MAX, STDEV etc…
n) Array Formulas – Perform multiple calculations in one cell - SUMPRODUCT
o) Pivot Tables & Pivot Charts, Adding Slicers - Value Field Settings, Filtering, Grouping, Sorting, Changing layout & format etc…

p) Data Visualization - Charts & Sparkline's: Static & Dynamic charts, formatting & Designing.
q) Analysis Tools - Apply various statistical methods to analyze the data.

- Correlation Analysis
- ANOVA : Single / Two Factor
- Random Number Generation
- T - test (Paired / Two samples)
- Understanding & Interpretation of statistical results.

Learning Method:

- Classroom (70%)
- Homework (10%)
- Project (20%)

Assessment:

- Practical Exam: 100 Marks

Total Hours Required: 40

CERTIFICATE COURSE ON R FOR ADVANCED STATISTICAL METHODS & MACHINE LEARNING

This course will offer you to learn how to apply advanced statistical methods & machine learning algorithms.

Learning Objectives:
Apply advanced statistical methods which include discovery & exploration of complex multivariate relationships among variables.

Learning Outcome:
1: Become familiar with the major R data structures.
2: Create your own functions & visualizations.
3: Learn to write your own project syntax ranging from importing data into R to apply standard and more advanced statistical analysis methods.

Course Syllabus:

Section 1: Advanced Statistical Methods Using R	Hours: 20
a) Generalized Linear Models: Logistic Regression, Multinomial Logistic Regression, Poisson Regression	
b) Ridge Regression	
c) Forecasting: Time Series Analysis – ARIMA	
d) Cluster Analysis: Hierarchical & Portioning cluster analysis (K- Means)	
e) Classification: Decision Tree & CHAID	
f) Text Mining: Word Cloud	
g) Dimensionality Reduction: PCA & Factor Analysis	

Section 2: Machine Learning with R	Hours: 20
a) Classification: Support Vector Machine, Random Forest Method, Navie Bayes	
b) Gradient Boosting Model	
c) Artificial Neural Network – Single Layer & Multiple Layer	

Learning Method:

- Classroom (70%)
- Homework (10%)
- Project (20%)

Assessment:

- Practical Exam: 100 Marks

Total Hours Required: 40

Changes made in the Syllabus

Sl. No.	Code No.,	Existing	Details of	%	Justification

	Paper Title & Unit	**Topic/Unit**	**changes made/added Topic/Unit**	**change**	
1.	STOE 602, Categorical Data Analysis, Unit – III	Statistical inference and model checking, fitting generalized linear models	Removed	7%	Difficult to understand