M.Sc. Programme in Statistics

Department of Statistics, West Bengal State University, Barasat.

Revised Course Structure & Regulations
for two years M.Sc. programme
in
Statistics

2022

[With effect from the academic year 2022-24]

M. Sc. programme in statistics equips a graduate student to build a career in various disciplines; e.g., data sciences, health sciences, environmental sciences, medical research, pharmaceutical research, finance, public policy making, data analytics etc. In today's fast growing complex world Statistics becomes an important discipline both for researchers and job seekers. Students will learn an optimal blending of various statistical techniques and its' applications. These will help the students to acquire sufficient knowledge in Statistics so that the students can opt for an academic career or can prepare themselves for jobs in different private and public sectors. Moreover, this programme is structured in such a way that a B.Sc. student in Mathematics will be able to pursue it. After successful completion of the programme in Statistics, a student will:

- have the versatility to work effectively in a broad range of analytic, scientific, government, financial, health, technical and other positions,
- recognize the importance and value of statistical thinking, training, and approach to problem solving, on a diverse variety of disciplines,
- be familiar with a variety of examples where statistics helps accurately explain abstract or physical phenomena,
- be able to independently read statistical literature of various types, including survey articles, scholarly books, and online sources; and others,
- be life-long learners who are able to independently expand their statistical expertise when needed, or for interest's sake.

Course Structure

Semester I			Credits	Marks	Total
STSPCOR01T	Departmental 1	Real Analysis & Linear Algebra	4	50	Total mark= 300 Total Credits=22
STSPCOR02T	Departmental 2	**Probability I**	4	50	
STSPCOR03T	Departmental 3	Descriptive Statistics & Regression Analysis I	4	50	
STSPCOR04T	Departmental 4	Distribution Theory & Statistical Inference I	4	50	
STSPCOR05P	Departmental 5	Lab I	4	50	
STSPAEC01M	AECC	Python Programming	2	50	
Semester II			Credits	Marks	Total
STSPCOR06T	Departmental 6	**Probability II** & Statistical Inference II	4	50	Total mark= 300 Total Credits=22
STSPCOR07T	Departmental 7	Linear models & Regression Analysis II	4	50	
STSPCOR08T	Departmental 8	Survey Methodologies I & Design of Experiments I	4	50	
STSPCOR09M	Departmental 9	R Programming & Statistical Inference III	4	50	
STSPCOR10P	Departmental 10	Lab II	4	50	
STSPSEC01M	SEC	Advanced Data Analysis	2	50	
Semester III			Credits	Marks	Total
STSPCOR11T	Departmental 11	Stochastic Process & Time Series Analysis	4	50	Total mark= 300 Total Credits=24
STSPDSE01T	Departmental 12 DSE1	Elective I	4	50	
STSPCOR12T	Departmental 13	Statistical Inference IV & Design of Experiments II	4	50	
STSPCOR13T	Departmental 14	Applied Multivariate Analysis	4	50	
STSPCOR14P	Departmental 15	Lab III	4	50	
STSPGEC01T	GEC	Basic Statistics	4	50	
Semester IV			Credits	Marks	Total
STSPCOR15T	Departmental 16	Advanced Bayesian Methods & Nonparametric Methods	4	50	Total mark= 300 Total Credits=24
STSPCOR16T	Departmental 17	Regression Analysis III & Categorical Data Analysis	4	50	
STSPDSE02T	Departmental 18 DSE 2	**Elective II**	4	50	
STSPCOR17P	Departmental 19	Lab IV	4	50	
STSPCOR18M	Departmental 20	Project Work & Project Presentation	8	100	

DSE1-Elective I: Any one out of 1 and 2

1. Advanced Sample Survey. 2. Econometrics & Financial Statistics.

DSE2-Elective II: Any one out of 3 and 4

3. **Biostatistics**. 4. Industrial Statistics.

Regulations for Two Years M.Sc. Programme in Statistics

In general the regulations for the two year (four semesters) M.Sc. programme in Statistics applicable from the academic year 2019-20 will be same as the comprehensive and uniform regulations of West Bengal State University.

Some particular points in the regulations for M.Sc. programme and Examinations in each of the semesters are as follows:

1. The candidate who has passed three years B.Sc. programme in Statistics/ Mathematics can apply for admission to M.Sc. course. Admission for candidates from other Universities/ Colleges will be governed according to University rules.

2. The examinations for M.Sc. course will be held in 4 semesters. At the end of each semester, an examination of the papers covered in semester would be held. This end semester examination will be referred to as M.Sc. examination of that semester.

3.1. Examination of theoretical papers is of 2 (1) hours (hour) duration and will be usually carry 40 (20) marks. 10 (5) marks of each paper will be set aside for continuous assessments / mid-term examination.

3.2. For theoretical papers paper setters will be appointed as per as university norm.

3.3. Evaluation of performance in a practical paper will be based examination at the end of the semester and on end-semester viva-voce and on practical works performed throughout the semester in the laboratory. The distribution of marks in each practical paper is as follows:

i). 10% for practical works performed throughout the semester in the laboratory –be evaluated by the teachers assigned for that course.

ii). 10% for viva-voce to be conducted by a Board consisting of the faculty members and / or external examiners.

iii). 80% for end-semester practical examination.

Only the total mark is to be shown in the mark-sheet.

3.4. In order to pass a semester examination, a candidate will have to score minimum of 40% of the total marks of theoretical papers and 50% in each practical paper. Pass marks for project work and project presentation will be 50% as in practical paper.

3.5. Each student will have to undertake a project work at the beginning of the 3rd Semester. The project work would have to be completed under the supervision of faculty member(s) at the end of the 4th Semester. He/she will be required to submit a written report and also make a formal presentation at the end of the 4th semester. The project work and its presentation will be assessed by a Board of Examiners consisting of Faculty members of the Department & External Examiner(s).

3.6. AECC and SEC Courses: These are 2 cp courses and the examination of 50 marks will be taken. Out of 50, 10 marks are assigned for continuous internal assessment (based on assignments), and the end term examination will be taken in two parts: 20 marks theory and 20 marks practical.

4.1. A candidate shall be eligible to appear at the semester examinations provided he/she is present in regular course of studies with proper attendance as per University rules.

4.2. All candidates who have completed a semester examination will join the next semester classes. Candidates failing to qualify in a Semester examination shall automatically revert back to the respective semester in the next academic session immediately after publication of the result. However the candidate failing in a paper in the previous examination has to sit for supplementary examination in the concerned course. A candidate will get a maximum of three consecutive chances.

4.3. Others details of supplementary examinations and pass/fail in a semester will be according to University Rules/ Departmental Committee.

5. Students will be required to select 1 elective paper in the 3rd semester and 1 elective paper in the 4th semester. These are detailed in the syllabus of the respective semester.

Syllabus for M.Sc. Programme in Statistics

Semester I

STSPCOR01T: Real Analysis & Linear Algebra [50 marks] (Core Theory Course)

Objectives & Outcomes: This course is comprised of two important mathematical tools necessary for in depth understanding of theoretical statistics. One is a foundation course on real analysis and another is an introductory course in linear algebra with emphasis on matrix algebra.

Real Analysis

Real number system, cluster points of sets, closed and open sets, compact sets, Bolzano-Weierstrass property, Heine-Borel property & its applications. [14]

Sequences and Series of functions: pointwise convergence, uniform convergence, absolute convergence. Some tests of convergence. Continuity, uniform continuity, differentiability of univariate and multivariate functions. Mean value Theorem. [14]

Reimann integral and its properties. Reimann-Stieltjes integral. Review of sequence and series of functions. Uniform convergence: term by term differentiation and integration Power series. Taylor series expansion. [10]

Linear Algebra

Vector spaces with real field. Basis dimension of vector space. Orthogonal vectors, Gram-Schmidt orthogonalization. [2]

Linear transformation of matrices. Matrix operations. Elementary matrices and their uses. Rank of a matrix and related results. Determinants. Inverse of a matrix. [3]

System of linear equations: homogenous and non-homogenous system. Generalized inverse: Moore-Penrose. [3]

Idempotent matrices and its properties. Characteristic roots and vectors. Quadratic forms and canonical reduction. Singular value decomposition. [4]

Reference Books

T.M. Apostal. : Mathematical analysis.
R. Rudin. : Principles of Mathematical analysis.
R. R. Goldberg. : Methods of real analysis.
J. C. Burkill. : First course of mathematical analysis.
J. C. Burkill, & H.Burkill : Second course of mathematical analysis.
R. G. Bartle, & D.R. Sherbert. : Introduction to real analysis.
S. K. Mapa. : Introduction to real analysis.
A. M. Goon. : Vectors and matrices.
G. Hadley. : Linear algebra.
F. E. Graybill. : Introduction to Matrices with Applications in Statistics.
C. R. Rao. : Linear Statistical Inference and Its Applications.
S. R. Searle. : Matrix Algebra Useful for Statistics.

STSPCOR02T: Probability I [50 marks]

Objectives & Outcomes: Probability Theory I will offer mathematical foundation required for explaining and analyzing random phenomena. Here introductory classical probability theory will be taught. This is a basic course specially designed for students with Mathematics background in B.Sc. In Probability Theory II, rigorous measure theoretic probability theory and related theoretical developments will be taught. This will help students to understand rigorous theory of statistics.

Probability I

Random Variables: Definition of discrete and continuous random variables. Cumulative distribution function and its properties, probability mass function and probability density function. Expectation and moments, Dispersion, Skewness, Kurtosis and Quantiles. [5]
Bivariate probability distributions. Marginal and conditional distributions. Independence. Conditional moments. [7]
Correlation and Regression. Generating Functions : probability generating function and moment generating function in univariate and bivariate cases. [5]
Probability Inequalities : Chebyshev's lemma, Markov's & Chebyshev's inequalities. Some common univariate distributions. Bivariate Normal distributions and its properties. [6]
Limit Theorems: Convergence in distribution: De-Moivre-Laplace limit theorem and Normal approximation to the Poisson distribution. [2]
Sigma fields in probabily. Borel sigma field. Measures and its properties. Probability as a measure. Measurable functions and its properties. Random variable as a measurable function . Integration of a measurable functions. [10]
Sequence of measurable functions: Monotone convergence theorem, Fatou's lemma and Dominated convergence theorem and their probabilistic aspects. Radon-Nikodym theorem and its applications. Distribution functions: application of Lebesgue-Stieltje's measure. Expectation and inequalities. [15]

Reference Books

K. L. Chung. : Elementary probability theory with stochastic process.
W. Feller. : An introduction to probability theory & its application (Vol. 1).
A. M. Goon, M. K. Gupta, & B. DasGupta. : An outline of statistical theory (Vol. 1).
V. K. Rohatgi, & Md. E. Saleh. : An introduction to probability and statistics.
P. J. Hoel, S.C. Port, & C.J. Stone. : Introduction to probability theory.
J. V. Uspensky. : Introduction to mathematical probability and its application.
T. Cacoullos. : Exercises in probability.
N. A. Rahman. : Practical Exercises in probability and statistics.
J. Pitman. : Probability.
D. Stirzaker. : Elementary probability
T. K. Chandra, & D. Chatterjee. : A first course in probability.

R. R. Bhat. : Modern probability theory.
A. Gut. : An intermediate course in probability.
J. Jacod, & P. Protter. : Probability essentials.
K. B. Athereya, & S. N. Lahiri. : Measure theory.
K. B. Athereya, & S. N. Lahiri. : Probability theory.
A. K. Basu. : Measure theory and probability.
P. Billingsley. : Probability and measure.
J. F. C. Kingman, & S. J. Taylor. : Introduction to measure and probability.
R. Ash. : Real analysis and probability.
G. Grimmett, & D. Stirzaker. : One thousand exercises in probability.

STSPCOR03T: Descriptive Statistics & Regression Analysis I [50 marks]

Objectives & Outcomes: The Descriptive Statistics is useful for describing the basic features of data and its graphical displays. This is the first step to exploratory data analysis extensively used when statistics is applied in real life scenarios. In Regression Analysis I, students will learn the basic concepts of simple and multiple regressions are related theories. Importantly, here students will study how to handle a prediction problem based on real data. This course in statistics is primarily aimed towards the students with Mathematics background in B.Sc.

Descriptive Statistics

Introduction to Descriptive statistics. Different types of data. Graphical representations. Univariate data : Different measures of location, dispersion, Relative dispersions, Skewness and kurtosis, Moments and Quantiles. Stem-and-leaf plot. Histogram and density plot. Box plot. [10]

Gini's coefficient and Lorenz curve. [3]

Bivariate data : Scatter diagram, Correlation coefficient and its properties. Correlation ratio, Correlation Index. Intraclass correlation with equal and unequal group sizes. Some multivariate plots and its' interpretations: Time plot and Star plot. [12]

Regression Analysis I

Simple linear regression and related results. Multiple linear regression. Multiple correlation and partial correlation and their properties. [7]

Heterogeneity and analysis of variance (ANOVA) and covariance (ANCOVA). Linear hypothesis and orthogonal splitting of total variance. Applications of the ANOVA technique: one-way classified data, two-way classified data with number of observations per cell. [8]

Testing simple regression coefficients, correlation ratio, linearity of simple regression, and multiple correlation and partial correlation coefficients. [10]

Reference Books

A. M. Goon, M. K. Gupta, & B. DasGupta. : Fundamentals of statistics (Vol.1).

G. U. Yule, & M. G. Kendall. : An introduction to the theory of statistics.
M. G. Kendall, & A. Stuart. : Advanced theory of statistics (Vols. 1-2)
G. W. Snedecor, & W. G. Cochran : Statistical methods.
F. F. Croxton, D. J. Cowden, & S. Klein. : Applied general statistics.
F. E. Wallis, & H.V. Roberts : Statistics: a new approach.
J. W. Tukey : Exploratory data analysis.
M. S. Lewis Beck (ed.) : Regression analysis.
N. R. Draper, & H. Smith. : Applied regression analysis.
H. Sheffe : The analysis of variance.
H. Sahai : The analysis of variance: fixed, random, and mixed models.
D. G. Montgomery, E. A. Peck, & G. G. Vining. : Introduction to linear regression analysis.
Debasis Sengupta & S. Rao Jammalamadaka (2019). Linear Models and Regression With R: An Integrated Approach, World Scientific, Singapore.

STSPCOR04T: Distribution Theory & Statistical Inference I [50 marks]

Objectives & Outcomes: In Distribution Theory students will learn the basic concepts of random sampling and distributions of some standard statistics under a postulated model (e.g., normal population) and will be familiar with the concepts of multivariate normal distribution and the sampling distributions of some standard statistics under multivariate normal populations. This is an extremely important prerequisite for understanding of statistical inferences and multivariate theories/ applications. The Statistical Inference I will provide basic ideas of inference from data; in particular, detail theory for point estimation. This course will provide students an introductory idea of how statistical theories are used to draw conclusions on unknown.

Distribution Theory

Concepts of random sampling. Sampling distribution of a statistics and standard errors. Some standard sampling distributions: Chi-square distribution, Derivation of the sampling distribution of sample mean and variance for a normal population, Student's and Fisher's t-distributions, Snedecor's F-distribution, Relationship between t, F and Chi-square distributions. Order Statistics: Introduction, distribution of the rth order statistic, smallest and largest order statistics. Joint distribution of order statistics, distribution of sample median and sample range. [10]
General discussion on multivariate distributions. Multivariate normal distribution and related results. Distribution of quadratic forms, Cochran's theorem. [7]
Random sampling from a multivariate normal distribution: Wishart matrix and its distribution and properties. Hotelling T^2 statistics: its distribution and related results. Mahalanobis D^2 statistics. [8]

Statistical Inference I

Concepts of Statistical Inference. A general overview: Estimation, Testing, and Confidence interval estimation. [5]

Data reduction: Sufficiency and minimal sufficiency, Completeness, Bounded completeness and Ancillary Statistic. Exponential family of distribution. [10]

Point Estimation: Concepts of point estimation. Requirement of a good estimator: notion of mean square errors, Unbiasedness and minimum variance. Methods for Minimum variance unbiased estimators. Cramer-Rao and Bhattacharaya system of lower bounds. Rao-Blackwell and Lehmann-Scheffe theorems. [10]

Reference Books

C. R. Rao. : Linear Statistical Inference and Its Applications.
T. W. Anderson. : Introduction to Multivariate Analysis.
S. S. Wilks. : Mathematical Statistics.
A. M. Khirsagar. : Multivariate Analysis.
C.G. Khatri, & M.S. Srivastava. : Introduction to Multivariate Statistics.
R. J. Muirhead. : Aspects of Multivariate Statistical Theory.
G. A. F. Seber. : Multivariate Observations.
M. Bioodeau, & D. Brenner. : Theory of Multivariate Statistics.
A. M. Goon, M. K. Gupta, & B. Dasgupta. : An Outline of Statistical Theory (Vol. I).
V. K. Rohatgi, & A. K. Saleh. : An Introduction to Probability and Statistics.
R. V. Hogg, & E. A. : A Brief Course in Mathematical Statistics.
R. A. Johnson, & G. K. Bhattacharya : Statistics-Principles and Methods.
A.M. Mood, F.A. Graybill, & D.C. : Introduction to the Theory of Statistics.
R. V. Hogg, & A.T. Craig. : Introduction to Mathematical Statistics.
G. Casella , & R. L. Berger. : Statistical Inference.

STSPCOR05P: Lab I [50 marks]

Objectives & Outcomes: In this course necessary practical exercises related to theoretical courses will be taught and will help students to learn implementations of important statistical tools in real life data sets.

Practical exercises based on STSC 1001 (Linear Algebra), STSC 1002 (Probability Theory I), STSC 1003 (Descriptive Statistics & Regression Analysis I) and STSC 1004 (Distribution Theory & Statistical Inference I).

STSPAEC01M: Python Programming [50 marks]

Objectives & Outcomes: In the last decade Python has become a popular environment for data science. Python is a general-purpose language with statistics modules. When it comes to building complex analysis pipelines that mix statistics with image analysis, text mining, control of a physical experiment, etc., the richness of Python is an invaluable asset. This course introduces both basic Python and the most important packages in a hands-on way with sufficient exercises. The objective of the course is that students will be capable of doing data management, visualization and analysis in Python on their own. This will enhance student's ability in statistical computing which is currently high in demand in industry.

Python Programming:

Part I: Introduction to Python programming

Machine set up: Python and the Anaconda distribution, Required Customisation, os.path, basic I/O, etc.; Variables: Local, Global-their scope and binding etc.; Data Types: Arrays, tuple, set, dictionary, user defined etc.; Loops: Conditionals, Comparisons, Inerrable and Iterators etc.; Functions: Defining and calling, argument Passing, self, recursive etc.; Classes: Methods, Instance variable, initialiser, etc.; Module: export, import, etc.; Style Guide. : docstrings, preceding , Trailing, etc.

Part II: Statistics using Python

Data representation, file handling, tables & plots; Linear regression, multiple factors, and analysis of variance; Probability distributions and random sampling.

Part III: Some important packages

Managing data with NumPy and pandas; Graphs with matplotlib and seaborn; Statistical analysis with statsmodels.

Reference Books:

1. Thomas Haslwanter: An introduction to statistics with Python.
2. Jose Unpingco. Python for probability, statistics, and machine learning (2nd editiuon).
3. Jesus Rogel-Salazar. Data science and analytics with Python.
4. Kent D. Lee: Python programming fundamentals (2nd edition).

Semester II

STSPCOR06T: Probability II & Statistical Inference II [50 marks]

Objectives & Outcomes: Probability III is a continuation of Probability II course in STSC 1002. This course will help students to understand theoretical statistics; specially, large sample theories of statistics. Statistical Inference II will impart students the theory of statistical hypothesis testing and confidence interval estimation. Testing and interval estimation have immense applications in different fields; e.g., medical, pharmaceutical, biological fields.

Probability II

Generating functions. Characteristic function: Inversion theorem and Continuity Theorem. [8]

Independence: Borel-Cantelli Lemma. Sequence of random variables. Different modes of convergence of a sequence of random variables-inter-relations. Weak and Strong laws of large numbers: related results. Kolmogorov's inequality. [10]

Central limit theorems & its applications. [5]

Concept of Martingles. [2]

Statistical Inference II

Testing of hypotheses: Nonrandomized and Randomized tests, Critical function, Power function. Most Powerful (MP) Tests: Neyman-Pearson Lemma (Existence, Sufficiency

and Necessity). Uniformly most powerful (UMP) tests: simple problems for exponential and pitman families of distributions. Enlargement technique for testing a composite null against a composite alternative. UMP tests for monotone likelihood ratio (MLR) families. Generalised Neyman-Pearson Lemma: uniformly most powerful unbiased (UMPU) tests for one parameter exponential families. Ideas of similar tests. [15]
Sequential analysis: Sequential Probability ratio test (SPRT). Wald's equation. Optimality of SPRT. Wald's fundamental identity. SPRT of one sided test. [7]
Confidence interval estimation: Relation with hypothesis testing. Optimum parametric confidence intervals. [3]

Reference Books

J. Jacod, & P. Protter. : Probability essentials.
K. B. Athereya, & S.N. Lahiri. : Measure theory.
K. B. Athereya, & S.N. Lahiri. : Probability theory.
A. K. Basu. : Measure Theory and probability.
B. R. Bhat. : Modern probability theory.
P. Billingsley. : Probability and measure.
R. G. Laha, & V. K. Rohatgi. : Probability theory.
R. Ash. : Real analysis and probability.
A. Gut : Probability : A graduate course.
M. Loeve. : Probability theory I.
Y. S. Chow, & H. Teicher. : Probability: Independence, Interchangeability, Martingles.
V.V. Petrov. : Limit theorems of probability theory.
E. L. Lehmann, & J. P. Romano. : Testing statistical hypotheses.
G. C. Casella, & R. L. Berger. Statistical inference.
S. Zacks. : Theory of statistical inference.
V. K. Rohatgi, & Md. E. Saleh. : An Introduction to probability and statistics.
R. W. Keener. : Theoretical statistics: Topics for a core course.
P. J. Bickel, & K. A. Doksum. : Mathematical statistics: basic ideas and selected topics (vol.1).

STSPCOR07T: Linear Models & Regression Analysis II [50 marks]

Objectives & Outcomes: In Linear Models students will learn detailed theoretical development for Gauss-Markov models and its extensions with related inference problems. Regression Analysis II consists of theoretical developments for inferences under deviations from Gauss-Markov models. Here concepts of outliers, robust transformations and dummy variables will be introduced to students. Students will be able to apply and investigate validity of the regression models in practical situations.

Linear Models

Gauss-Markov model. Estimable functions. Best linear unbiased estimator (BLUE). Gauss Markov Theorem. Estimation space and error space. Sum of squares due to a set

of linear functions. Estimation with correlated observations. Least Square estimation with linear restriction on the parameters. [10]
General linear hypothesis: *F* test for general linear hypothesis and associated confidence sets. Multiple comparison procedures of Scheffe and Tukey. Applications of general linear hypothesis to regression. [8]
Analysis of variance and covariance (ANOVA & ANCOVA). Introduction to random and mixed effect models (balanced case). [7]

Regression Analysis II
Weighted least squares, Box-Cox and other transformation methods. Dummy variables and its applications. [7]
Tests of fit of a model. Detection of outliers. Residual analysis: residuals and their plots. Departures from the usual assumptions: heteroscedasticity, autocorrelation, multicollinearity, Non-normality- detection and remedies. [15]
Model selection. [3]

Reference Books
H. Scheffe. : Analysis of variance.
S. R. Searle. : Linear models.
G. A. F. Seber, & A. J. Lee. : Linear regression analysis.
N. Giri. : Analysis of variance.
D. D. Joshi. : Linear estimation & design of experiments.
C. R. Rao. : Linear statistical inference and its applications.
R. B. Bapat. : Linear algebra and linear models.
R. W. Keener. : Theoretical statistics: Topics for a core course.
A. Sen, & M. Srivastava. : Regression analysis: Theory, methods, and applications.
S. Chatterjee, & A. S. Hadi. : Sensitivity analysis in linear regression.
S. Chatterjee, & A. S. Hadi. : Regression analysis by example.
S. Weisberg. : Applied linear regression.
J. Fox. : Regression diagnostics: An introduction.
D. A. Belsley, E. Kuh, & R. E. Welsch. : regression Diagnostics: Identifying influential data and Sources of collinearity.
N. Giri. : Linear Estimation & design of experiments.
D. N. Gujarati. : Basic econometrics.
D. N. Gujrai, & D. Porter. : Basic econometrics.
J. M. Wooldeidge.: Introductory econometrics: A modern approach.
Debasis Sengupta & S. Rao Jammalamadaka (2003). Linear Models: An Integrated Approach, World Scientific, Singapore.

STSPCOR08T: Survey Methodologies I & Design of Experiments I [50 marks]

Objectives & Outcomes: In Survey Methodologies I students can learn the basic concepts of finite populations, drawing of samples by different sampling techniques from the populations, and related estimation of target quantities (that is, finite population parameters) like population mean, variances etc. Classical survey mythologies will be taught here and will help students to understand how surveys are conducted in different public and private sectors. In Design of Experiments I students will be taught the concepts of controlled experiments and related inference problems. It has enormous applications in designing and anglicising of experiments in the physical, chemical, biological, medical, social, psychological, economic, engineering, agricultural, or industrial sciences. The course will examine how to design experiments, carry them out, and analyze the data they yield. Upon completion, students will be able to understand the issues and principles of Designs of experiments.

Survey Methodologies I

Probability sampling from a finite population. Basic sampling schemes–Unequal probability sampling with and without replacement, Related estimators of population total/mean, their variances and variance estimators – Mean per distinct unit in simple random with replacement sampling, Hansen Hurvitz estimator in unequal probability sampling with replacement, Des Raj and Murthy's estimator (for sample of size two) in unequal probability sampling without replacement. [9]

Stratified sampling – The allocation scheme & the basic idea of optimal allocation, construction of strata. Basic ideas of Ratio, Product, Difference and Regression estimators. Unbiased Ratio type estimators. Hartley – Ross estimator in simple random sampling. [6]

Sampling and sub-sampling of clusters. Two-stage sampling with equal/unequal number of second stage units and simple random sampling without replacement / unequal probability sampling with replacement at first stage. Double sampling for stratification. Double sampling ratio and regression estimators. [10]

Design of Experiments I

Basic principles of experimental design: randomization, replication, and local control. Uniformity trials. Shapes and sizes of plots and blocks. [4]

Standard designs and their analyses : completely randomised design (CRD), randomised block design (RBD), latin square design (LSD), split plot design, and strip arrangements. Comparison of efficiencies. Applications of the techniques of analysis of variance to the analysis of the above designs. [9]

General block designs and its information matrix. Concepts of connectedness, orthogonality and balance. Resolvable designs. Properties of BIB designs. Designs derived from BIB designs. Intrablock analysis of orthogonal (CRD, RBD, LSD) and non-orthogonal designs (Balanced incomplete block design and Youden square design). [12]

Reference Books

W. G. Cochran. : Sampling techniques.

Des Raj. : Sampling theory.

A. S. Hedayat & B. K. Sinha. : Design & inference in finite population sampling.

P. Mukhopadhyay. : Theory & methods of survey sampling.

P. S. Levy, & S. Lemeshow. : Sampling of populations: Methods and applications.

S. K. Thompson. : Sampling.

A. Chaudhuri. : Essentials of survey sampling.

A. Chaudhuri, & H. Stenger. : Survey sampling: Theory and methods.

O. Kempthrone. : The design and analysis of experiments.
M. N. Das, & N. C. Giri. : Design and analysis of experiments.
D. C. Montgomery. : Design and analysis of experiments.
W. G. Cochran, G. M. Cox. : Experimental designs.
W. T. Federer. : Experimental designs : Theory and application
A. M. Dean, & D. Voss : Design and analysis of experiments.
M. C. Chakraborty. : Mathematics of design and analysis of experiments.
A. Dey. : Theory of block designs.
G. A. Milliken, & D.A. Johnson. : Analysis of messy data Vol. 1: Designed experiments.

STSPCOR09M: R Programming & Statistical Inference III [50 marks]

Objectives & Outcomes: R is a language and environment for statistical computing and graphics. This course will offer students programming knowledge on R. At the end of the course, the students should be able to parse a real-world data analysis problem into one or more computational components and to apply suitable statistical methods for optimal decision making in different research areas. Statistical computing with R language helps students for getting a job in analytics and any other fields where data analysis is required. The course on R programming is a skill development course. Also learning R becomes necessary for students who opt for a Ph.D. programme in statistics. In Statistical Inference III students will learn classical frequentist approach for parametric inference based on approximate distributions of the statistics in large samples. This enables students to draw inference based on statistics whose exact sampling distributions are very difficult to evaluate analytically. This course is an important prerequisite for learning advanced statistical theories and Ph.D. in statistics.

R Programming

Introduction to R : Installing R and Rstudio. Basic operations. Running R script. R notebook. 'Input from' and 'output to'. Accessing systems. R as a scientific calculator. Scripting: Basic rules, operations and relational. Looping and Iterating. Conditional loops. [4]

Graphics in R: 2D, 3D plots: scatter, line, bar, histogram, pie, etc. plot functions. Legend, text, equation and colour . par function and multiple /multi-paneled plots. [2]

Data types. Data frames, lists, arrays- vector, matrices. String manipulation, sub-strings and/or pattern– searching, inserting and concatenating. Data manipulation : built-in datasets in R, read.table() and write.table() functions, and importing data from external sources. Regular expressions. R functions, binding values to symbols, creating user defined function, lazzy evaluation, scoping rules - lexical and dynamic, etc. Environments in R, pipes and traits in R. [5]

Matrix calculations using R: Vectors and matrices. Basic operations on vector and matrices . Solution of linear equations. Eigenvalues and Eigenvectors of a square symmetric matrix. Spectral Decomposition, LU,QR, Choleski, and Singular Value Decomposition (SVD). Moore-Penrose Generalized inverse. Finding a basis,

orthonormalisation and finding rank. Linear models : the lm function; ANOVA/ANCOVA/regression, models, the summary function, goodness of fit measures, predicted values and residuals; the ANOVA table, confidence intervals and confidence ellipsoids. Multiple testing. [7]
Advanced graphical using R. Optimization technique using R. Numerical Integration and differentiation, Random number generations and Simulation. Monte Carlo methods. [7]

Statistical Inference III
Asymptotic Methods: Large sample properties of estimators: Consistency, Efficiency and Asymptotic Normality-CAN and BAN estimators. [8]
Method of moments estimator, Maximum likelihood estimatimator (MLE) - Large sample properties. Efficient estimator. [8]
Likelihood ratio, Rao and Wald tests for simple and composite hypotheses- properties and asymptotic distribution of test criteria in the simple hypothesis case. [9]

Reference Books
N. Matloff. : The art of R Programming
W. N. Venables, & B. D. Ripley. : Statistics: An introduction using R.
A. de Vries, & J. Meys. : R for dummies.
P. Dalgaard. : Introductory statistics with R.
K. Baclawski. : Introduction to probability with R.
M. J. Crawley. : Statistics: An introduction using R.
J. C. Nash. : Nonlinear parameter optimization Using R tools.
M. Lawrence. : Programming graphical user interfaces in R.
P. Murrell. : R graphics.
R. L. Eubank. : Statistical computing with C++ and R.
J. Fox., & S. Weisberg. : An R Companion to Applied Regression.
J. Maindonald, & W. J. Braun. : Data analysis and graphics using R: An example-based approach.
E. L. Lehmann. : Elements of large-sample theory.
E. L. Lehmann, & G. Cesella. : Theory of point estimation.
T. S. Ferguson. : A course in large sample theory.
P. K. Sen, & Singer. : Large sample methods in statistics: An introduction with applications.
R. J. Serfling. : Approximation theorems in mathematical statistics.
A. C. P. de Lima, J. M. Singer, & P. K. Sen. : From finite sample to asymptotic methods in statistics.
A. Dasgupta. : Asymptotic theory of statistics and probability.
G. C. Casella, & R. L. Berger. : Statistical inference.
S. Zacks. : Theory of statistical inference.
V. K. Rohatgi, & Md. E. Saleh. : An Introduction to probability and statistics.
R. W. Keener. : Theoretical statistics: Topics for a core course.
P. J. Bickel, & K. A. Doksum. : Mathematical statistics: basic ideas and selected topics (vol.1).

STSPCOR10P: Lab II [50 marks]

Objectives & Outcomes: In this course necessary practical exercises related to theoretical courses will be taught and will help students to learn implementations of important statistical tools in real life data sets.

Practical exercises based on STSC 2006 (Statistical Inference II) and STSC 2007 (Linear Models & Regression Analysis II), STSC 2008 (Survey Methodologies I & Design of Experiments I) and STSC 2009 (Statistical Inference III).

STSPSEC01M: Advanced Data Analysis [50 marks]

Objectives & Outcomes: This **skill development** course comprises of three different topics relevant to the data analytic problems encountered in practice. The first two topics, missing data mechanism and re-sampling techniques will be introduced to students. Missing data is a part of all most all real life surveys. Students will learn how to use missing data sets efficiently for estimation problems. In modern era of high level extensive computing, Re-sampling methods for drawing inference, particularly bootstrap, becomes extremely popular in practice since its implementation does not require theoretical knowledge of statistics. Here students will learn basic theoretical understanding of re-sampling theory and its implementations. The most useful and contemporary part of this course is statistical learning. The goal is to help students learn, understand supervised and unsupervised (machine) learning approaches which include the study of modern computing and scaling up machine learning techniques focusing on industry applications. The students become familiar with regression methods, classification methods, clustering methods, dimensionality reduction techniques. The students will be able to demonstrate their knowledge in various machine learning related Artificial Intelligence Techniques. This knowledge will be useful in applying in various jobs in Governance, Banking, Manufacturing, Healthcare, Robotics etc.

Missing data: Inference for data with missing values. Different types of missing data mechanism. MLE using E-M algorithms. Various imputation techniques.
Re-sampling Techniques: Introduction to Jackknife and Bootstrap – methods for estimating bias, standard error and distribution function based on i.i.d. random variables. Standard examples. Justification for the methods in i.i.d. set-up. Bootstrap confidence intervals. Computational aspects.
Statistical learning: fundamentals ideas of supervised and unsupervised learning. Supervised learning: Cross-validation, Nearest neighbourhood classification, Classification and regression trees, Bootstrap, Random forests and Gradient boosting. Unsupervised learning: Principal component analysis (PCA), probabilistic PCA, k-means clustering.

Reference Books

R. J. A Little, & D. B. Rubin. : Statistical analysis with missing data.
C. K. Enders. : Applied missing data analysis.
T. Raghunathan. : Missing data analysis in practice.
B. Efron & R. J. Tibshirani. : An introduction to the bootstrap.
A. C. Davison, & D.V. Hinkley. : Bootstrap methods and their applications.
J. Shao, & D. Tu. : The jackknife and bootstrap.
M. R. Chernick, & R. A. LaBudde. : An introduction to bootstrap methods with application to R.
T. Hastie, R. Tibshirani, & J. Friedman. : Elements of statistical learning: Data mining, inference & prediction.
G. James, D. Witten, T Hastie, & R. Tibshirani. : An introduction to statistical learning with applications in R.
R. A. Berk. : Statistical learning from a regression perspective.
M. Kuhn, & K. Johnson. : Applied predictive modelling.

Semester III

STSPCOR11T: Stochastic Process & Time Series Analysis [50 marks]

Objectives & Outcomes: Here theory of different dependent processes will be taught to students. These processes are developed to model real life random phenomena from various fields of sciences and social sciences where underlying random variables are dependent in nature. In Time Series course students will experience introductory theoretical developments for particular type of dependency, known as time dependency. This course on is extremely important for elective course on Econometrics and Financial Statistics.

Stochastic Process

Markov chain with finite state space and countable state space, Classification of states, Chapman-Kolmogorov equation, Calculation of n-step transition probability matrix and its limit, Stationary distribution of Markov chain. Random walk and Gambler's ruin problem and reversibility. [14]
Discrete state space continuous time Markov chains, Poisson process. [6]
Renewal theory: Elementary Renewal theorem, Stopping time, Statement and uses of Key Renewal theorem. [10]

Time Series Analysis

Elements of time series: Introduction to time series data and its applications to various fields. Modelling of time series as deterministic functions of iid terms: different components and decomposition. Estimation of trends and seasonal components. Notions of multiplicative models. [5]

Stationary Time Series, linear process, causality and Invertibility, properties of linear stationary process, autocovariance function (ACVF) and partial autocorrelation function (PACF), methods of finding ACVF, ARMA process as particular case. [6]
Forecasting a time series, Best linear unbiased predictor. Yule-walker equations for finding predictors for ARMA process. [5]
Non-stationary process: ARIMA and SARIMA. [4]

Reference Books

S. Karlin, & H. M. Taylor. : A first course in stochastic processes.
S. Ross. : Stochastic process.
J. Medhi. : Stochastic process.
A. K. Basu. : Stochastic process.
C. Chatfield. : The analysis of time series: An introduction.
M. G. Kendall. : Time series.
G. E. P. Box, G. M. Jenkins, & G. C. Reinsel. : Time series analysis: Forecasting and control.
P. J. Brockwell, & R. A. Davis. : Introduction to time series and forecasting.
P. J. Brockwell, & R. A. Davis. : Time series: Theory and methods.
R. H. Shumway, & D. S. Stoffer. : Time series analysis & its applications.
J. D. Cryer, & K.-S. Chan. : Time series analysis with application in R.

STSPDSE01T: Elective I [50 marks]

<u>Elective paper 1</u>: Advanced Sample Survey

Objectives & Outcomes: Advanced survey sampling is an unique advanced survey course. It comprises of important features from classical and modern analytic methods used in survey sampling. This course is divided into two papers: Advanced Sample Survey and Model-dependent Inference in Survey Sampling. Some important features of this course are: (a) randomized response techniques; (b). small area estimation; (c) adaptive cluster sampling for rare populations; and (d) model-based and model-assisted estimators for efficient use of auxiliary information. The students are expected to learn the model-dependent modern survey techniques from a data analytic point of view. In Indian context there is great scope for applications of modern survey methodologies in different surveys (e.g., socio-economic surveys) conducted by public/private sectors. After completion of the course, students will be able to demonstrate their knowledge in various government organisations such as NSSO, CSO and also in non-government agencies. Small area estimation may be used for various national developments in different Government and private surveys and health related issues. Adaptive cluster sampling may be applied in forest surveys, estimation of rare birds, animals, soils etc. It may be used in different rare disease related health surveys as well.

Survey Methodologies II

The basic model, Sampling Design and sampling schemes. Hanurav's unit drawing algorithm: Inclusion probabilities of first two orders, Relation with effective sample size and variance of effective sample size. Data and estimators-linear and linear unbiased estimators of population total, Horvitz –Thompson estimator, Generalized difference and generalized regression estimators, issues in non-negative variance estimation. π-PS sampling schemes of Midzuno-Sen, Brewer, Durbin and JNK Rao, Rao-Hartley-Cochran strategy. Randomised response: The Warner model: unbiased and maximum likelihood estimation. The unrelated question models methods (one and two unrelated characters)-unbiased estimation under the cases where the population in the unrelated group is known/unknown, comparison with the Warner model. Small Area Estimation - the basic estimation method, ratio and regression estimators for domains. Issues in small domain estimation - synthetic estimators. Adaptive sampling for rare and elusive population. [25]

Model-Dependent Inference in Survey Sampling

Inference under fixed population model: sufficiency and likelihood. Choosing good sampling strategy. Nonexistence theorem of Godambe and Joshi. Inference under super population model. Prediction approach. Asymptotic approach- asymptotic design unbiasedness and consistency. Finite population distribution function estimation : Chambers-Dunstan and Rao-Kovar-Mantel estimator. Resampling methods in finite population inferences (Only introduction). Bayesian Estimation in survey sampling – Empirical Bayes & Hierarchical Bayes estimators. Model based inference in small area estimation – Fay - Herriot model. [25]

Reference Books

C. M. Cassel, E. Sarndal, & J. H. Wretman. : Foundations of inference in survey sampling.
A. Chaudhuri & H. Stenger. : Survey **s**ampling—theory and methods.
A. Chaudhuri. : Essentials of survey sampling.
A. Chaudhuri. : Modern survey sampling.
A. Chaudhuri & J. W. E. Vos. : United theory and strategies of survey sampling.
P. Mukhopadhyay. : Inferential problems in survey sampling.
C. E. Sarndal, B. Swensson, & J. Wretman. : Model assisted survey sampling.
S. Thompson & G. Seber. : Adaptive sampling.
W. Fuller. : Sampling statistics.
J. N. K. Rao. : Small area estimation.
P. Mukhopadhyay. : Topics in survey sampling.
R. Chambers & R. Clark. : An introduction to model-based survey sampling with applications.
Z. Mashreghi et al. : A Survey of bootstrap method in finite population sampling (Statistics Surveys, 2016, p.1-52).

<u>Elective paper 2</u>: Econometrics & Financial Statistics

Objectives & Outcomes: This course provides an introduction to some standard econometric methods and their applications. The course emphasizes intuitive and conceptual understanding as well as hands on econometric analysis using modern computer software (like R) on data sets from economics and business. Course on Statistics in Finance is a contemporary course in financial statistics. This course covers standard topics like option pricing, portfolio management and value-at-risk. After completion of this advanced course students may opt for research in econometrics and finance. Apart from that finance is one of the important areas where job opportunities are growing day by day.

Econometric Methods

Single-equation linear model – some variations. Nonparametric methods in econometrics. Simultaneous Equations – identification & estimation. Analysis of Panel Data. Bayesian Econometrics. Demand Analysis. Production Function Analysis. Analysis of some special econometric models. [25]

Statistics in Finance

The value of time, bond pricing with a flat term structure. The term structure of interest rates. The Mean Variance Frontier. The global minimum variance portfolio. Efficient portfolio. The zero beta portfolio. Allowing for a riskless asset. Efficient sets with risk free assets. Pricing of futures contract. Binomial option pricing. Multiperiod binomial pricing. Basic Option Pricing. Black-Scholes formula, Extending the Black Scholes formula. Dividends. Risk-free and risky assets. Contracts and options. Continuously compounded interest, present valuation, risk, risk-neutral valuation. Arbitrage: examples, contracts and options under no-arbitrage assumptions. Option Pricing: Cox-Ross-Rubinstein Binomial and Black-Scholes models. Elementary portfolio management, Value-at-risk. [25]

Reference books

J Johnston. : Econometric Methods
G. G. Judge, et.al. : The theory and practice of econometrics.
W. Greene. : Econometric analysis.
A. Zellner. : An introduction to Bayesian inference in econometrics.
E. Malinvaud. : Statistical methods in econometrics.
H. Wold & L. Jureen. : Demand analysis – a study in econometrics.
P. Sankhayan. : An Intro.to the Economics of Agricultural Production
M. Nerlove. : Estimation and identification of Cobb-Douglas models.
A. Pagan & A. Ullah. : Nonparametric econometrics.
D. Ruppert. : Statistics and finance.
D. Reppert. : Statistics and data Analysis for financial engendering.
E. Lindstrom, H. Madsen, & J. N. Nielsen. : Statistics for finance.
J. Franke, W. K. Hardle, & C. M. Hafner. : Statistics for financial markets.

S. M. Ross. : Introduction to mathematical finance: Options and other topics.
V. S. Bawa, S. J. Brown, & R. W. Klein. : Estimation risk and optimal portfolio choice.
R. Carmona. : Statistical analysis of financial data Using R.

STSPCOR12T: Statistical Inference IV & Design of Experiments II [50 marks]

Objectives & Outcomes: Statistical Inference III will impart introductory ideas of decision theory and Bayesian methods. In Decision theory classical inference problems are viewed as a risk minimization problem and optimum rules are evaluated. This is a theoretical course. In today's world for explaining complex phenomena Bayesian inference becomes increasingly popular in statistics. Here basic ideas of Bayesian methods will be introduced to the students. Apart from freuquentist approach, Students will learn another kind of statistical theory for inference on complex random phenomena. The Design of Experiments II is a continuation of the earlier course. Both the course together will provide a comprehensive overview of practical used in the physical, chemical, biological, medical, social, psychological, economic, engineering, cryptography, agricultural, or industrial sciences.

Statistical Inference IV

Decision Theory: Loss, Decision rules and Risk function, Admissibility of decision rules, Bayes and Minimax rules. [10]
Bayesian Methods: Overview and comparison of different paradigms, Relative advantages and disadvantages. Priors & Posteriors : Subjective priors, Conjugate and other Non-subjective priors. Bayesian Inference – estimation, testing, interval estimation and prediction for some common models and common priors. Hierarchical and Empirical Bayes Methods. Bayesian Computation. [15]

Design of Experiments II

Recovery of inter block information in BIB designs. Missing plot techniques. Elementary ideas of Lattice and PBIB designs. [5]
Construction of Mutually Orthogonal Latin Squares (MOLS).Construction of BIBD using MOLS and Boses's fundamental method of difference. [6]
Factorial experiment: Confounding and balancing in symmetric factorial experiments-Analysis. [10]
Response survey designs. [4]

Reference Books
T.S. Ferguson. : Mathematical statistics.
E. L. Lehmann, & G. Cesella. : Theory of point estimation.
J. O. Berger. : Statistical decision theory and Bayesian analysis.
J. Pratt, & H. Raiffa. : Introduction to statistical decision theory.
F. Liese., & K.-J. Miescke. : Statistical decision theory.
P. D. Hoff. : A first course in Bayesian statistical methods.

G. E. P. Box, & G. C. Tiao. : Bayesian inference in statistical analysis.
J.-M. Marin, & C.P. Robert. : Bayesian Essentials with R
J. Albert. : Bayesian Computation with R
M. C. Chakraborty. : Mathematics of design and analysis of experiments.
A. Dey. : Theory of block designs.
D. Raghavarao. : Constructions & combinatorial problems in design of experiments.
D. Raghavarao, & L.V.Padgett. : Block design: Analysis, combinatorics and applications.
R. C. Bose. : Mathematical theory of symmetric factorial design (Sankhya – Vol. 8).
D. G. Kabe, and A. K. Gupta. : Experimental designs: Exercises and solutions.
G. Casella. : Statistical design.
T. P. Ryan. : Modern experimental design.
C. F. J. Wu, & M. S. Hamada. : Experiments: planning, analysis and optimization.
D. C. Montogomery. : Design and analysis of experiments.

STSPCOR13T: Applied Multivariate Analysis [50 marks]

Objectives & Outcomes: This is extremely useful course consists of important multivariate real life problems like Principal components analysis, classification and discrimination problems, factor analysis and cluster analysis. In practice survey data are generally multivariate in nature and these techniques will help students to perform exploratory data analysis to survey data to revel important conclusions about the underlying populations. These techniques are very common in any field where data based conclusions are to be drawn; particularly in Analytics. This course develops important **skills** to handle multivariate problems.

Multivariate linear regression model: estimation of parameters, tests of linear hypotheses, different test criteria, Multivariate Analysis of variance of one and two way classified data. Multivariate Analysis of Covariance. [15]
Hierarchical and non-hierarchical clustering methods. [8]
Classification and discrimination procedures for discrimination between two known populations – Bayes, Minimax and Likelihood Ratio procedures. Discrimination between two multivariate normal populations. Sample discriminant function. Likelihood ratio rule. Tests associated with discriminant function, Probabilities of misclassification and their estimation. Classification of several populations. Fisher's method for discriminating among several populations. [10]
Population and sample principal components and their uses and related large sample inference. [5]
The orthogonal factor model, Estimation of factor loading, Factor rotation, Estimation of Factor scores, Interpretation of Factor Analysis. [7]
Canonical variables and canonical correlations (population & sample) and their interpretations. Large sample inferences. [5]

Reference Books

R. A. Johnson, & D. W. Wichern. : Applied multivariate statistical analysis.
N. H. Timm. : Applied multivariate analysis.
A. V. Rencher, & W. F. Christensen. : Methods of multivariate analysis.
W. K. Hardle, & L. Simar. : Applied multivariate statistical analysis.
T. W. Anderson.: An introduction to multivariate statistical analysis.
G. A. F. Seber.: Multivariate observations.
R. J. Muirhead. : Aspects of multivariate statistical theory.
M. Bilodeau, & D. Brenner. : Theory of multivariate statistics.
P. Mukhopadhyay. : Multivariate statistical analysis.
A.V. Rencher. : Multivariate statistical inference and applications.
B. Everitt & T. Hothorn. : An introduction applied multivariate analysis with R.

STSPCOR14P: Lab III [50 marks]

Objectives & Outcomes: In this course necessary practical exercises related to theoretical courses will be taught and will help students to learn implementations of important statistical tools in real life data sets.

Practical exercises based on STSC 3011 (Time Series Analysis) and STSC 3012 (Elective I), STSC 3013 (Statistical Inference IV & Design of Experiments II) and STSC 3014 (Applied Multivariate Analysis).

STSPGEC01T: Basic Statistics [50 marks]

Objectives & Outcomes: This is an unique course intended for students from other M.Sc. programmes. After completion of the course students are expected to understand fundamental ideas of statistics and can successfully apply basic statistics to analyze real data sets.

Study design. Graphical representation of data. Features of frequency distribution, summary measures. Problems with outliers and extremes. Association, dependence, causality. Correlation and regression in bivariate andmultivariate setups. Discrete data analysis. [17]
Probability. Basic results. Conditional probability and Bayes theorem. Random variables, expectation and variance. Probability models for discrete and continuous variables. Computation of probability in various applied research. [15]
Basics of Statistical inference. Estimation and Hypothesis testing problems in special setups. Applications of statistical inference in applied research. [18]

Reference Books

A. M. Goon, M.K. Gupta, & B. Dasgupta. : Fundamentals of statistics (Vols. 1 & 2).
A. M. Goon, M.K. Gupta, & B. Dasgupta. : Outlines of statistical theory (Vols. 1 & 2).
V. K. Rohatgi, & A. K. Saleh. : An Introduction to probability and statistics.

A.M. Mood, F.A. Graybill, & D.C. : Introduction to the theory of statistics.
G. Casella , & R. L. Berger. : Statistical inference.
S. Ross. : A first course in probability.
G. W. Snedecor, & W. G. Cochran. : Statistical methods.
A. Agresti. : Analysis of ordinal categorical data.

Semester IV

STSPCOR15T: Advanced Bayesian Methods & Nonparametric Methods [50 marks]

Objectives & Outcomes: In this course on Bayesian methodologies student will learn Bayesian linear regressions, Bayesian estimation and testing problems and Bayesian computation & simulations (MCMC, Gibbs sampling and MH algorithm). This course on Bayesian inference will enable students to perform Bayesian data analysis in many real life complex problems. In the second part of the course on Nonparametric methods students will learn to draw distribution-free inference in classical testing problems along with an introductory discussion on nonparametric density estimation problem. In real life any assumed parametric model always has chance to be misspecified and hence inferences based on assumed model would be wrong. The course on non-parametric methods helps students to draw inferences without any model assumption

Advanced Bayesian Methods
Bayesian linear Model and Regression. Hierarchical and Empirical Bayes estimation. parametric empirical Bayes estimator and its computation. Bayesian computation and simulation: Markov Chain Monte Carlo (MCMC), Gibbs Sampling, Metropolis Hastings (MH) algorithm. Testing and Model selection. [25]

Nonparametric Methods
U-statistics- Definition and Asymptotic properties. [4]
Nonparametric tests: Single sample Problems: Location, Location-cum-symmetry, and Goodness-of-fit problems. Two-sample Problems: Location, Scale and Homogeneity problems. Multi-sample location problem. Friedman Two-way Analysis of variance problem. Bivariate association problem, Cochran Q-test for dependent samples. Nonparametric Interval Estimation. Concept of Asymptotic Relative Efficiency. [18]
Nonparametric Density Estimation. [3]

Reference Books

P. D. Hoff. : A first Course in Bayesian statistical methods.
J.K. Ghosh, M. Delampady, & T. Samanta. : An introduction to Bayesian analysis: Theory and methods.
B. P. Carlin, & T. A. Louis. : Bayesian methods for data analysis.
C.P. Robert. : The Bayesian choice.
J.-M. Marin, & C.P. Robert. : Bayesian core: A practical approach to computational Bayesian statistics.

A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin: Bayesian data analysis.
J.-M. Marin, & C.P. Robert. : Bayesian essentials with R.
J. Albert. : Bayesian computation with R.
T.S. Ferguson. : Mathematical statistics.
E. L. Lehmann, & G. Cesella. : Theory of point estimation.
J. O. Berger. : Statistical decision theory and Bayesian analysis.
J. Pratt, & H. Raiffa. : Introduction to statistical decision theory.
F. Liese., & K.-J. Miescke. : Statistical decision theory.
P. D. Hoff. : A first course in Bayesian statistical methods.
G. E. P. Box, & G. C. Tiao. : Bayesian inference in statistical analysis.
J.-M. Marin, & C.P. Robert. : Bayesian essentials with R
J. Albert. : Bayesian Computation with R
J. D. Gibbons. : Nonparametric inference.
J. D. Gibbons, & S. Chakraborty. : Nonparametric statistical inference.
T. P. Hettmansperger. : Statistical inference based on ranks.
E. L. Lehmann. : Statistical methods based on ranks.
M. Hollander, D. A. Wolfe, & E. Chicken. : Applied nonparametric statistics.
L. Wasserman. : All in nonparametric statistics.

STSPCOR16T: Regression analysis III & Categorical Data Analysis [50 marks]

Objectives & Outcomes: In previous two courses on regression analysis students will learn to deal with regression problems when responses are continuous in nature. But here students will primarily learn to develop regression when responses are binary, count or categorical in nature. Regression Analysis III will give theoretical treatments to regression when responses are assumed to belong to an exponential family of distributions (called GLM). Categorical Data Analysis course will teach students regression theories for nominal/ordinal responses. This course has many practical applications in different fields of studies, e.g., medical studies, social studies.

Regression analysis III

Generalized Linear Models (GLM): Introduction, components, goodness of fit measures-residuals and deviance. [5]

Inference for GLM. Applications to binary, count and ordinal/nominal data. Over dispersion. Marginal, conditional and quasi likelihood functions. [10]

Robust regression: Effects of outliers, LAD regression, M-estimation, LTS regression, S-estimation, and data analytic examples. [10]

Categorical Data Analysis

Categorical Response Data: Nominal/ Ordinal Distinction, Probability Distributions for Categorical Data-Binomial & Multinomial Distributions. [5]

Inference for a proportion: Wald, Score, and Likelihood-Ratio Inference for Binomial Parameter. Contingency Tables**:** Probability Structure for Contingency Tables: Joint, Marginal, and Conditional Probabilities, Relative risk and odds ratio-properties. Measures of association and tests for independence in contingency tables: Nominal-Nominal, Ordinal – Ordinal and Nominal–Ordinal Tables. Exact Inference for Small Samples- Fisher's Exact Test for 2 × 2 Tables. [13]

Association in more than two-way classified data: Partial association, Conditional Versus Marginal Associations- Simpson's Paradox, Conditional and Marginal Odds, log-odds ratio and its distribution, Independence-Conditional Versus Marginal. [7]

Reference Books

C. E. McCullagh, S. R. Searle, & J. M. Neuhaus.: Generalized, linear, and mixed models.
J. J. Faraway. : Extending the linear models with R.
P. McCullagh, & J. L. Nelder. : Generalized linear models.
A. Dobson, & A. Barnett. : An introduction to generalized linear models.
J. K. Lindasy. : Applying generalized linear models.
A. Agresti. : Foundations of linear and generalized linear models.
D. W. Hosmer, S. Lemeshow, and R. Sturdivant. : Applied logistic regression.
R. A. Marona, R. D. Martin, & V. Yohai. : Robust statistics: Theory and methods.
F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, & W. A. Stahel. : Robust statistics: The approach based on influence functions.
P. J. Huber. : Robust Statistics.
A. M. Leroy, & P. Rousseeuw. : Robust regression and outlier detection.
R. A. Marona, D. Martin, V. J. Yohai., & M. Salibian-Barrera : Robust statistics: Theory and methods (with R).
A. Agresti. : Categorical Data Analysis.
A. Agresti. : Analysis of Ordinal Categorical Data.
T. J. Santner and D. E. Duffy: The statistical analysis of discrete data.
C. R. Bilder and T. M. Loughin: Analysis of categorical data with R.
Debasis Sengupta & S. Rao Jammalamadaka (2019). Linear Models and Regression With R: An Integrated Approach, World Scientific, Singapore.

STSPDSE02T: Elective II [50 marks]

<u>Elective paper 3</u>: Biostatistics

Objectives & Outcomes: This is a Biostatistics module comprises of four topics: Survival Analysis, Clinical Trials, ROC Curve Analysis & Statistics in Epidemiology. This course Survival Analysis has more weightage compared to other three since it is most important among the four topics. The Survival Analysis course will focus on basic concepts of survival (time-to-event) data analysis. Some of the important features that students are expected to learn in this course are: (a) understanding of different types of censoring, and learn to estimate and interpret survival characteristics; (b) assess the relationship of risk factors and survival times using the Cox regression model, and assess the appropriateness and adequacy of the model; and (c) development of analytic skills through the analysis of data sets taken from the fields of medicine and public health. The goal of three introductory courses on Clinical Trials, ROC Curve Analysis and Epidemiological Statistics is to provide an exposure to the statistical analysis used heavily in medical science as well as biostatistics industry. It will definitely expand the exposure of the students who will opt for either research in Biostatistics & related fields or company job as Biostatistician after his/her post-graduation.

Survival Analysis: Concepts of lifetime. Various schemes of censoring and associated likelihood functions. Estimation of survival function: Parametric procedure: Point estimation, Scores and likelihood ratio tests for selected parametric models and confidence intervals. Distribution free procedures: Actuarial estimator, Kaplan-Meier and Nelson – Aalen estimators. Regression models: Estimation in parametric and Semi-parametric models-Cox's proportional hazard model, Time dependent covariates, Rank test. Competing risk analysis and multivariate models. [25]

Clinical Trials: Ethical issues in Clinical Trials and its different phases, Sample size determination, Brief introduction to Simple randomized design, stratified randomized crossover design and Sequential designs. [8]

ROC Curve Analysis: ROC curve for validity Diagnostic test-Examples from medical field, Population ROC curves-it's properties, Slope of ROC curve and optimality results, Area under ROC (AUC) measure, Bi-normal model and Binary regression model-estimation of ROC curve. [9]

Statistics in Epidemiology: Definition of Epidemiology. Measures of disease frequency-incidence, prevalence and relative risk, Epidemiological study designs – Cohort study and Case control study designs and analysis. [8]

Reference Books

R. G. Miller. : Survival analysis.
D. J. Kleinbaum & M. Klein. : Survival analysis: A self-learning text.
D. W. Hosmer, S. Lemeshow, & S. May. : Applied survival analysis: Regression modelling of time to event data.
J. P. Klein, & M. L. Moeschberger. : Survival analysis: Techniques for censored and truncated data.
J. D. Kalbfleisch, & R. L. Prentiice. : The statistical analysis of failure time data.
D. F. Moore. : Applied survival analysis using R.
S. Piantadosi. : Clinical Trials - A Methodologic Perspective.
B.S. Everitt & A. Pickles. : Statistical Aspects of Design & Analysis of Clinical Trials.
S.J. Pocock. : Clinical Trials.
J. Whitehead. : The Design and Analysis of Sequential Clinical Trials.
W. F. Rosenberger & J.M. Lachin. : Randomization in Clinical Trials- Theory and Practice.
W. J. Krzanowski & D. J. Hand.: ROC Curves for Continuous Data.
M. S. Pepe.: The statistical evaluation of medical tests for classification and prediction.
X.-H. Zhou, N. A. Obuchowski, & D. K. McClish. : Statistical Methods in Diagnostic Medicine.
K. J. Rothman & S. Geenland. : Modern Epidemiology.
S. Selvin. : Statistical Analysis of Epidemiologic Data.
D. McNeil. : Epidemiological Research Methods.

S. C. Newman.: Biostatistical Methods in Epidemiology.
J.F. Jekel, J.G. Elmore, & D.L. Katz. : Epidemiology, Biostatistics and Preventive Medicine.
Mark Woodward.: Epidemiology: Study Design and Data Analysis.

Elective paper 4: Industrial Statistics

Objectives & Outcomes: The course on Operations Research is specially designed for applications of statistics in industry. The course on operation Research will introduce students how to implement quantitative methods and techniques for effective decisions–making; model formulation and applications that are used in solving business decision problems in management engineering. This course on Optimization is specially designed for applications of statistics in industry. The course on Optimizations will introduce the theory of optimization methods and algorithms developed for solving various types of optimization problems in engineering and technology. After successful completion of the course, student will be able to: (a) identify and develop operational research models from the verbal description of the real system; and (b) understand the mathematical tools that are needed to solve optimization problems.

Operations Research

Definition and Scope of Operations Research: phases in Operation Research, models and their solutions, decision-making under uncertainty and risk, use of different criteria, sensitivity analysis. Decision-making in the face of competition, two-person games, pure and mixed strategies, existence of solution and uniqueness of value in zero-sum games, finding solutions in mixed strategy games. Analytical structure of inventory problems, EOQ formula of Harris, its sensitivity analysis and extensions allowing quantity discounts and shortages. Multi-item inventory subject to constraints. Models with random demand, the static risk model. P- and Q- systems with constant and random lead times. Queuing models – specification and effectiveness measures. Steady-state solutions of M/M/1 and M/M/c models with associated distributions of queuelength and waiting time. M/G/1 queue and Pollazcek-Khinchine result. Network Flow Models, minimum spanning tree, shortest path, mincut-maxflow, CPM and PERT using network flow. Traveling salesman Problem. Replacement theory, sequencing. [25]

Optimization

Generalized L.P.P. Bounded variables, decomposition principle of Dantizg and Wolfe. Transportation problem. Unconstrained Optimization, Optimality Conditions, first-order, second-order necessary sufficiency under convexity. Algorithms for Univariate Optimization: Bisection, Newton, Safeguarded Newton, Golden section search, Fibonacci rates of convergence. Integer programming – integer linear and mixed integer linear programming problems, Gomery's cutting plane method, Branch and Bound method. Binary Programming – Bala's algorithm ISI. Non-linear programming –optimization with equality & inequality constraints: Details of Karush-Kuhn-Tucker theory, Quadratic Programming – Wolfe's algorithm and Beale's algorithm. [25]

Reference Books

H. A. Taha. : Operational research.
F. S. Hillier, & G. J. Leiberman. : Introduction to operations research.
K. Swarup, P. K. Gupta, & M. M. Singh. : Operations research.
D. T. Philips, A. Ravindran, & J. Solberg. : Operations research.
C. W. Churchman, R. L. Ackoff, & E. L .Arnoff. : Introduction to operations research.
T. M. Starr, & D. W. Miller. : Inventory control - Theory & practice.
L. Kleinrock. : Queueing systems.
M. Sasieni, A. Yaspan, & L. Friedman. : Operations research.
R. L. Ackoff, & M. W. Sasieni : Fundamentals of operation research.
G. Hadley. : Non-linear and dynamic programming.
K. G. Murthy. : Linear and combinatorial programming.
P. Whittle. : Optimization under constraints - Theory and applications of non-linear programming.
S. S. Vajda. : Probabilistic programming.
N. S. Kambo. : Mathematical programming techniques
S. S. Rao. : Optimization - Theory and applications.
K. V. Mittal. : Optimization methods.

STSPCOR17P: Lab IV [50 marks]

Objectives & Outcomes: In this course necessary practical exercises related to theoretical courses will be taught and will help students to learn implementations of important statistical tools in real life data sets.

Practical exercises based on STSC 4016 (Advanced Bayesian Methods & Nonparametric Methods) and STSC 4017 (Regression Analysis III & Categorical Data Analysis) and STSC 4018 (Elective II).

STSPCOR18M: Project Work & Project Presentation. [100 marks]

Objectives & Outcomes: Students will independently solve theoretical/ empirical statistical problems. This will help students to better understanding and implementations of different statistical methodologies.