Progressive Education Society's

Modern College of Arts, Science and Commerce (Autonomous),

Shivajinagar, Pune 5

(An Autonomous College Affiliated to Savitribai Phule Pune University)

Syllabus

For

M.Sc. (Statistics) Semester I and II

(Based on NEP 2020 framework)

(To be implemented from the Academic Year 2023-24)

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester I

(2023 Course under NEP 2020)

Course Code: 23ScStaP111

Course Name: Basics of Real Analysis and Calculus.

Teaching Scheme: 4 hours/week **Credit: 4**

Examination Scheme: CIA: 50 Marks **End-Sem: 50 Marks**

Prerequisite Courses:

- Basics of Mathematical analysis.
- Basics of Calculus

Course Objectives:

- To understand the basic concepts of Real analysis and Calculus
- To solve the mathematical problems

Course Outcomes:

- Students are enabled to solve problems of real analysis and calculus.

Course Contents:

Chapter 1		10 lectures
	1.1 Review of Set theory, Set of real numbers , Supremum and infimum of sets of real numbers, real field, ordered set and field, Archimedean principal, countable and uncountable sets, countability of rational numbers, uncountability of real numbers. 1.2 Metric space, Properties of metric space Neighborhood points, Exterior and interior points , boundary points , limit points, open , closed and compact sets. Bolzano – Weierstrass and heine- Borel theorem.	
Chapter 2		15 lectures
	2.1 Sequence of real numbers, limit of sequence and its properties, Convergence and divergence of sequence Cauchy sequence and related theorems (Cauchy criteria of convergence), subsequence and their convergence and divergence, convergence of bounded monotone sequence. 2.2 Series of real numbers, convergence and divergence of series of real numbers, test for convergence (root test, ratio test) , absolute convergence (without proof), uniform convergence, power series , radius of convergence of power series (Binomial , Exponential, geometric and log series), term by term differentiation (integration) of absolute convergent series, change of order of summation of series.	
Chapter 3		15 lectures
	3.1 Revision: derivative of function of single variable, Mean value theorem, Taylor's series expansion 3.2 Multivariate calculus: explicit and implicit functions, continuity, partial derivatives, differentiability, partial derivatives of higher order, Hessian matrix, Taylor's theorem, extreme values, differentiation with respect to vector and matrix, jacobian of transformation.	

Chapter 4		20 lectures
	4.1 Riemann and Riemann – Stieltjes integral**:** Partition of interval, norm of partition, finer partition, tagged partition, Upper and Lower Riemann and R-S sums, order relation between Upper and Lower Riemann sums. Effect of finer partition on the difference between Upper and Lower Riemann sums. Necessary and sufficient condition for a function to be R and R-S integrable Properties with proof (i) Continuous bounded function is R and R-S integrable (ii) Monotonic bounded function is R and R-S integrable (iii) Fundamental theorem of integral calculus. 4.2 Improper integrals: Definition, convergence of an integral, P-integral , exponential integral test for convergence (comparison test), convergence of beta and gamma integrals, relationship between beta and gamma functions 4.3 Double integrals: Definition, properties, change of order, iterated integrals, Fubini's theorem, differentiation under integral sign (Leibnitz rule) and transformation of variables.	

Reference Books:

1. Ajitkumar and S. Kumaresan (2014), A Basic Course in Real Analysis (Chapman and Hall)
2. Rudin W. (1985), Principles of Mathematical Analysis (McGraw–Hill)
3. Apostol T. M. (1975) Mathematical Analysis: A Modern Approach to Advanced Calculus (Addison - Wesley)
4. Bartle R. G. (1976), Elements of Real Analysis (Wiley)
5. Malik S. C. &Arora S. (1991), Mathematical Analysis (Wiley Eastern Limited 2nd edition)
6. Goldberg R. R. (1964), Methods of Real Analysis- Blasdell Publishing company, New York, U.S.A.
7. Bartle G.R. &Sherbert D. R. (2000), Introduction to Real Analysis- John Wiley & Sons publications.
8. Royden (1988), Principles of Real Analysis (Macmillian)
9. S R Ghorpade and B V Limaye(2000), A Course in Calculus and Real Analysis, Springer
10. S R Ghorpade and B V Limaye(2009), A Course in Multivariate Calculus and Analysis, Springer

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester I

(2023 Course under NEP 2020)

Course Code: 23ScStaP112 **Course Name: Linear Algebra**

Teaching Scheme: 4 hours/week **Credit: 4**

Examination Scheme: CIA: 50 Marks **End-Sem : 50 Marks**

Prerequisite Courses:

- Basics of Matrix Algebra
- Basics of vector analysis**.**

Course Objectives:

- To understand the concepts and applications of matrix algebra and Numerical methods.

Course Outcomes:

- Student is able to solve the problems on matrices.
- Student is able to use iterative methods.

Course Contents:

Unit 1		15 lectures
	• Vector Space, Subspace, Linear dependence and independence, Basis of a vector space, dimension of vector space, orthogonal and orthonormal basis, Gram –Schmidt orthogonalization, • Matrix algebra, special types of matrices, rank, inverse and determinant of a matrix and their properties, Orthogonal and idempotent matrix and their properties, Projection theorem, linear transformation, linear equations Solution space and null Space.	
Unit 2		15 lectures
	• Generalized (g) inverse and Moore-Penrose g- inverse (MP g- inverse), its properties and examples. • System of homogeneous and non- homogeneous linear equations, solution space and null Space.	
Unit 3		15 lectures

	• Characteristic roots (eigen values) of real matrices, right and left characteristics vectors (eigen vectors), Independence of characteristics vectors corresponding to distinct Characteristic roots, Algebraic and geometric multiplicity • Spectral decomposition, power of a matrix, Cayley-Hamilton theorem, singular value decomposition.	
Unit 4		15 lectures
	• Introduction of quadratic forms, reduction and classification of a quadratic form, simultaneous reduction of two quadratic forms, Maxima and minima of a quadratic form • Properties of a quadratic form for orthogonal and idempotent matrices.	

Reference Books:

1. Hohn, F. E. (1973). Elements of Matrix Algebra, Macmillan
2. Kollo, T. and Rosen, D. von (2005). Advanced Multivariate Statistics with Matrices, Springer, New York.
3. Kumaresan, S. (2000). Linear Algebra: A Geometric Approach, Prentice Hall
4. Lay, D. C. Lay, S. R. and Mc Donald, J. J. (2016) .Linear Algebra and Its Applications, Fifth Edition, Pearson, Boston.
5. Ramachandra Rao, A. and Bhimasankaram, P. (2000). Linear Algebra. Hindustan Book Agency
6. Rao, C. R. (2009) : Linear Statistical Inference and Its Applications, second edition, Wiley
7. Searle S. R. (2017). Matrix Algebra Useful for Statistics, second edition , Andre I. Khuri.

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester I

(2023 Course under NEP 2020)

Course Code: 23ScStaP113 | **Course Name: Probability Distributions**

Teaching Scheme: 4 hours/week | **Credit: 04**

Examination Scheme: CIA: 50 Marks | **End-Sem: 50 Marks**

Prerequisite Courses:

- Basic probability distributions and related theory.

Course Objectives:

- To introduce students to various univariate, bivariate and multivariate distributions and related theoretical results.

Course outcomes:

- Students can derive theoretical results of various probability distributions which is a prerequisite for inference course.

Course Contents:

Chapter 1		15 lectures
	1.1 Random Variables: Axiomatic definition of probability, elementary properties of probability measures, Continuity of probability measure (without proof), Borel field and random variables, probability distributions and probability distribution functions (p.d.f.) of random variables. (4L) **1.2 Distribution function:** Properties of p.d.f. Discrete, continuous and mixture type p.d.f. Standard distribution functions. Decomposition theorem. Construction of new distributions from the given p.d.f. - products and powers of distribution functions, truncated distributions. Symmetric p.d.f. and its properties. Probability mass function (p.m.f.) and density function and relation to distribution functions. **1.3 Expectations and Moments of Random Variables:** Expectation of a random variable. Moments and moment inequalities. Moment Problem. **1.4 Generating functions**-Probability generating function (p.g.f.) and its properties, PGF of compound distribution. Moment generating function (M.G.F.) Characteristic function and properties, conjugate pairs of distributions, Parseval relation, uniqueness theorem	
Chapter 2		15 lectures

	2.1 Random vectors: Bivariate and multivariate probability distributions and their characteristic properties. Joint, marginal and conditional distributions. Independent random variables.(4L) **2.2 Standard multivariate distributions:** Compound distributions and Wald's Identity. Bivariate Poisson distribution. Dirichlet distribution and its properties.	
Chapter 3		15 lectures
	3.1 Bivariate exponential distributions: Decomposition theorem of Bivariate distributions. Marshal-Olkin, Freund, Moran, and Gumbel Bivariate exponential distributions and their properties. **3.2 Transformations of random variables:** Functions of univariate random variables and their distributions. Quantile functions (probability integral transformation) and their properties, relationship with uniform distribution. **3.3 Transformations of random vectors:** Distribution of sums of random variables and convolutions, Order statistics and related distributions. Distribution of spacings.	
Chapter 4		15 lectures
	4.1 Quadratic forms: Quadratic forms and their classification and related distributions. Fisher-Cochran theorem. **4.2 Non-central distributions:** Chi-square, t and F distributions and their properties. **4.3 Distribution free statistics:** Some distribution free statistics and their distributions.	

Reference Books:

1. Berger, R. and Casella G. (2002). Statistical Inference, Duxbury Resource Center, Second Edition.
2. Dasgupta, A. (2010) Fundamentals of Probability: A First Course, Springer, NewYork.
3. Hogg, R. V. , McKean, J. W. and Craig, T. T. (2005). Introduction to Mathematical Statistics, Sixth Edition, Pearson Prentice Hall, New Jersey.
4. Rao, C. R. (2002). Linear Statistical Inference and Its Applications, Wiley
5. Rohatgi, V. K. & A. K. M. E Saleh (2001). Introduction to Probability and Statistics, Wiley,New York.

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester I

(2023 Course under NEP 2020)

Course Code: 23ScStaP114

Course Name: Lab Course on 23ScStaP111, 23ScStaP112 and 23ScStaP113

Teaching Scheme: 4 hours/week **Credit: 2**

Examination Scheme: CIA: 25 Marks **End-Sem: 25 Marks**

Prerequisite Courses:

- Knowledge of MS-Office, R software .
- Basic concepts of linear algebra, probability distributions and numerical analysis

Course Objectives:

- To acquaint students with data analysis techniques with the help of software

Course Outcomes:

- Students can solve problems on probability distributions, linear algebra etc. using statistical software.

Course Contents:

Title and Content
1. Introduction to Statistical Software (R software & MATLAB software) : Classification, tabulation and frequency tables (univariate and bivariate data). • Graphical representation of data. • Summary statistics. • Scatter diagram, correlation, Regression. 2. Matrices: Properties of Matrices (rank, inverse, transpose, determinant etc), Getting vectors in row/column space and null space of the given matrix. 3. Definiteness of Matrix, eigen values and eigen vectors of a matrix, algebraic and geometric multiplicity of an eigen value, etc. Computing power of a given matrix using spectral decomposition. 4. Inverse of a square matrix (by direct method and partitioning method), g-inverse, MP g-inverse. 5. Gram-Schmidt orthonormalization: Forming an orthogonal matrix of specified order using Gram-Schmidt orthogonalization.

6. Solution of System of Linear Equations using Gauss elimination, Gauss Jorden, Gauss Seidal and Gauss Jacobbi methods.
7. Classification and Reduction of Quadratic forms, Verification of Cayley-Hamilton theorem
8. Model sampling from discrete, continuous distributions. (Use inversion method if necessary).
9. Model sampling from mixture distributions.
10. Computation of probability of events related to bivariate probability distributions.
11. Model sampling from bivariate probability distributions.
12. Computation of probability of non-central χ^2, t, F-distributions.
13. Solution to Simultaneous Bivariate equations by Newton Raphson method
14. Newton's Interrpolation for bivariate functions
15. Evalution of double integral by Trapezoidal rule and Simpson"s rule

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester I

(2023 Course under NEP 2020)

Course Code: 23ScStaP121 **Course Name: Sampling theory**

Teaching Scheme: 4 hours/week **Credit: 4**

Examination Scheme: CIA: 50 Marks **End-Sem: 50 Marks**

Prerequisite Courses:

- Basic statistical methods.

Course Objectives:

- To acquaint students with various sampling techniques.

Course Outcomes:

- Student can apply various sampling techniques.

Course Contents:

Chapter 1		15 lectures
	1.1 Review of finite population sampling techniques (random and non-random), SRSWR and SRSWOR: Inclusion probabilities, related results on estimation of population total 1.2 Determination of sample size for pre-specified variance, pre-specified error in the estimation, pre-specified width of the confidence interval, pre-specified relative error in the estimation, Simple random sampling for the proportion**,** Estimation of proportion for the more than two classes**,** Inverse Sampling (Sampling for the rare attribute) and estimator of the population mean and its variance**.** 1.3 Probability Proportional to Size with Replacement (PPSWR) methods, cumulative total method and Lahiri's method for estimation problem, estimation of finite population mean and total. Sampling with varying probability without replacement, Ordered and Unordered estimates**,** Horwitz–Thompson estimator, its variance and properties, midzuno scheme of sampling, Yates-Grundy sampling estimate, Murthy's estimate.	
Chapter 2		15 lectures
	2.1 Stratified sampling, comparison of allocation problem of allocation in stratified sampling, construction of strata, deep stratification, The method of Collapsed strata, post stratification, estimator of population mean and variance of estimator of population mean under post stratification, Stratified random sampling for the proportion, comparison of stratified random sampling with simple random sampling. 2.2 Use of supplementary information for estimation, ratio estimator of population mean, its bias and mean square error, unbiased ratio type estimators of population mean, variance of estimator of population mean under it, , ratio estimator for the stratified random	

	Sampling, regression method of estimation, estimator of population mean, its bias and mean square error of the Estimator, comparison of estimator of population mean under ratio, regression and simple random sampling.	
Chapter 3		15 lectures
	3.1 Systematic sampling, sample mean and its variance, Comparison of systematic sampling under linear trend, Yates corrected estimator, Centered systematic sampling, Balanced systematic sampling and Modified systematic sampling, circular systematic sampling, two dimensional systematic sampling (Aligned and Unaligned Systematic sampling), comparison of systematic sampling with random sampling and stratified sampling, PPS systematic sampling. 3.2 Cluster sampling with clusters of equal sizes, estimation of populationmean and itsstandarderror, Relative efficiency of cluster sampling w.r.t. SRSWOR, Effect of cluster size on relative efficiency, unbiased estimator of relative efficiency, cluster sampling as a one way ANOVA, Optimum value of the cluster size, cluster sampling for the proportion, Cluster sampling with cluster of unequal sizes ,bias in estimator of population mean , bias in the estimator and its MSE, unbiased estimator and relative efficiency of unequal cluster sampling, PPS cluster sampling, estimation of population mean.	
Chapter 4		15 lectures
	4.1 Two stage sampling with equal and unequal second stage units, estimation of population mean and its standard error, optimum value of the number of clusters and second stage units in the two stage sampling having equal second stage units. 4.2 Two phase sampling, ratio and regression estimator of population mean under two phase sampling, bias in the estimator and its MSE. 4.3 Sampling and non– sampling errors, response errors, mathematical model for Response errors, Hansen Hurwitz technique, Randomized Response Technique (RRT), Warner's randomized response technique.	

Reference Books:

1. Arnab, R. (2017). Survey Sampling: Theory & Applications, Academic Press
2. Chaudhuri, A. (2014). Modern Survey Sampling, CRC Press
3. Cochran, W.G. (1984). Sampling Techniques, Wiley.
4. Des Raj and Chandhok, P. (1998). Sample Survey Theory, Narosa.
5. Murthy M.N.(1977) Sampling theory and methods (Statistical Publishing Society)
6. S. Sampath (2005) Sampling theory and methods (Narosa)
7. Singh, D. and Chaudhary F.S (1986).Theory and Analysis of Sample Survey Designs, Wiley Eastern Limited.
8. Singh, S. (2003). Advance Sampling Theory and Applications (Volume I and II), Kluwer Academic Publishers.
9. Sukhatme, P.V, Suktatme, B.V., Sukhatme, S. and Asok, C. (1984).Sampling Theory of Surveys with applications, Indian Society for Agricultural Statistics, New Delhi.
10. Thompson, S. K. (2012). Samplig, 3rd Edn., Wiley
11. Parimal Mukhopadhyay: New Central Book Agency: Sampling Theory

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune-5

M.Sc. (Statistics) Semester I

(2023 Course under NEP 2020)

Course Code: 23ScStaP122

Course Name: Survival analysis and statistical quality control

Teaching Scheme: 4 hours/week **Credit: 4**

Examination Scheme: CIA: 50 Marks **End-Sem: 50 Marks**

Prerequisites:

- Basic understanding of parametric and non-parametric inference.

Course Objective:

- To study the different models from Survival Analysis, to understand different types of censoring, learn to estimate and interpret survival characteristics.
- To provide the construction of parametric and non-parametric estimators of survival distributions, and probability density functions based on incomplete data. The models with right-censored, truncated and interval censored data will be considered.

Course Outcomes: After completion of the course students will able to:

- Gain the ability to recognize the difference between parametric and non-parametric survival models.
- Estimate non-parametric survival function of the data.
- To estimate survival function, cumulative hazard rate function using the so-called Kaplan-Meier estimator.
- Use the test of exponentiality against non-parametric classes in real life problems.

Course Contents:

Unit 1		15 lectures
	Concepts of time, order and random censoring (left and right), Life time distribution models - Exponential Gamma, Weibull, Lognormal, Pareto, linear Failure rate, Makeham family, Lehman families (proportional hazard rate family), spacing, normalized spacing and results of an exponential distribution based on normalized spacing. Revision of Parametric inference for complete data: a) Exponential distribution: Point estimation of parameter of exponential distribution and Fisher information, exact and asymptotic Confidence Intervals for λ, obtaining minimal sufficient and consistent estimator of λ, Graphical method for checking exponetiality of data. b) Weibull: Obtaining MLE of scale and shape parameter of Weibull distribution and sample information matrix. c) Gamma: Obtaining MLE of scale and shape parameter of Gamma distribution and sample information matrix. d) Lognormal: Obtaining MLE of of parameter μ and σ, Confidence Interval for μ and σ. Parametric inference for censored data: i) Type I censoring: Exponential distribution ii) Type II censoring: Exponential, gamma, Lognormal	

	iii) Random censoring: Exponential, Lehman family, Weibull distribution. iv) Non-Parametric estimation of survival Function a) For complete data: Non parametric estimator of distribution function and survival function, distribution of emperical survival function, confidence band for survival function (by Using Kolmogorov – Smirnov statistics) b) For censored data: Actuarial estimator of survival Function, Estimator of variance of actuarial estimator (Greenwoods formula), Kaplan Meier (product limit estimator) and its variance, redistribution to right algorithm.	
Unit 2		15 lectures
	Test for Exponetiality: Estimable function of degree r, Kernel, symmetric Kernel ,U- statistic, variance of U- Statistic, one sample U-Statistic theorem, Hollander and Proschan Test, Test for exponentiality against positive ageing based n sample spacing, Analytical test for exponentiality against NBUE, Deshpande's Test, Two sample U- statistic theorem, Wilcoxon and Mann Whitney test, Gehan's test, Mantel-Haenzel test, Log rank test, concept of covariates Semi-parametric regression for failure rate Cox's proportional hazards model with one and several covariates. Base line model, link function, likelihood function proportional Hazard Rate model	
Unit 3		15 lectures
	Statistical Process control (SPC) Total quality Management(TQM) : meaning and dimensions of quality, Quality improvement, Quality Philosophy, Introduction to TQM, six sigma, DMAIC, and other extension of TQM, quality systems, The ISO 9000 and other Quality systems. Control Chart: Revision of theory of control charts, Concepts of stable industrial processes, Systematic variation, random variation, variation within and between subgroups, estimation of process parameters,. Equivalence between control chart and testing of hypothesis problem. Choice of control limits Operating characteristic (OC curve) of control chart. Probability of false alarm, probability of catching shift in parameter. Concept of Run length, probability distribution of run length, average run length (ARL). Comparison of control chart using ARL,OC curve, criteria of detecting lack of controls (sensitizing rules),patterns on control charts with justification and its effect on Probability of false alarm. Control Charts: (a)X-bar MR chart (b) CUSUM Chart:, Chart statistic (C_i^+ ,C_i^-)and chart parameters(k,h), construction and working of tabular CUSUM chart for mean and variance ,Statement of hypotheses. Estimation of shift in mean of process, Fast initial response or head start feature, Siegmund's approximation for ARL and determination of chart parameters. CUSUM chart for subgroup size n>1, comparison between Shewhart chart and CUSUM chart V mask procedure. (c) EWMA chart: Chart statistic its expectation and variance. Choice	

	of chart parameters (λ, L). Construction and working of EWMA chart for mean and variance. EWMA chart for subgroup size n>1, Comparison of Shewhart control charts with CUSUM charts. Simulation of ARL(δ).	
Unit 4		15 lectures
	Process capability: Different Process capability and performance indicesCp, Cpk, Cpm.Propertis and relation between capability indices.Connection between proportion of defectives(DPPM) and Cp.Interval estimation of mean given Cpm ≥1.Estimation and confidence intervals of estimators of CpandCpk Testing of hypothesis about Cp . Control charts for auto correlated observations: Need ,constructions of control chart for residuals after fitting first order auto correlated model. Hoteling T^2Chart: Testing multivariate normality, (a)Hotelling T^2multivariate control chart for mean vector when (i) dispersion matrix is (i) known (ii) unknown ARL(0), ARL(δ). T^2 control chart when subgroup size n=1 (b)control chart for dispersion matrix when mean vector is (i) known (ii) unknown. Sampling Plans: Description of MIL STD and Dodge Romig sampling plans a) Acceptance Sampling Plan for attributes with Curtailed inspection: Equivalence between sampling plans and testing of hypothesis problem. b)Double, multiple and sequential sampling plans for attributes sampling plan, Operating characteristic functions. AOQL, ATI, ASN. c) Continuous Sampling Plans, d) chain sampling plan.	

Reference Books :

1. Deshpande, J.V, Purohit, S.G.,(2005),Life Time Data :Statistical Models and Methods
2. Cox, D.R. and Oakes, D. (1984) Analysis of Survival Data, Chapman and Hall, New York.
3. Elandt-Johnson, R.E., Johnson N.L. (1980) Survival models and Data Analysis, John Wiley and Sons
4. Gross A.J. and Clark, V. A. (1975) Survival Distributions: Reliability Applications in the Biomedical
   Sciences, John Wiley and Sons.
5. Miller, R.G. (1981) Survival Analysis, John Wiley and Sons.

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

PG Part 1(Semester I), M.Sc. (Statistics)

(2023 Course under NEP 2020)

Course Code: 23ScStaP131 **Course Name: Research Methodology**

Section 1: Research Methodology Core Paper

Teaching Scheme: 2 Hours/Week **Credit: 02**

Examination Scheme: CIA: 25 Marks **End-Sem: 25 Marks**

Prerequisite Courses:

- B.A., B.Sc. B.Com, B.Voc., BBA, BBA IB, BBA CA

Course Objectives:

- To make students aware about research and its importance
- To obtain knowledge regarding systematic gathering of data and get advanced knowledge in the selected topic
- To inculcate logical and organized thinking in students
- To investigate some existing situation or problem by creating new system or method
- To help students to design research problem

Course Outcomes:

On completion of the course, student will be able to–

1. Describe importance of research
2. Differentiate between unethical and ethical practices of publication ethics
3. Select research problem appropriately
4. Prepare good hypothesis
5. Design research problem systematically
6. Analyze and organize data correctly
7. Prepare good scientific research report

Course Contents:

Unit 1	Research problem and design	No. of lectures
	Introduction to research :meaning and definition of research, objective of research, importance of research, characteristics of good research, purpose and role of research, classification of research **Research problem :** defining of research problem, Criteria for selecting the research problem, , importance of literature survey in defining research problem. Hypothesis :Defining Hypothesis, types of hypothesis, characteristics of good hypothesis, formulation of	**15**

	hypothesis **Research Design :**Definition and features of research design, Concept of research design, types of research design, preparation of research design, Sampling techniques, characteristics of good sampling designs	
Unit 2	**Data analysis, report writing and publication ethics**	**15**
	Definition of Data, methods of data collection, analysis of data, types of data analysis, Questionnaire, Design of Questionnaire, Testing hypothesis : parametric and non-parametric tests : T-test, Z-test, Chi-square test, ANOVA **Report writing** : importance of interpretation of results, meaning, definition and significance of report /thesis writing, Principals of research report drafting, Types of reports, layout of research report, important parts of reports, precautions of preparation of report/ thesis **Publication ethics :**definition, introduction and importance, best practices/ standard settings initiative and guidelines COPE, WAME, etc, conflict of interest, Publication misconduct :definition, concept problems that lead to unethical behavior, violation of publication ethics, predatory publishers and journals, software tools to identify predatory publications developed by SPPU	

Section 2: Research Methodology (Statistics)

Teaching Scheme: 2 Hours/Week **Credit: 02**

Examination Scheme: CIA: 25 Marks **End-Sem: 25 Marks**

Course Contents:

Unit 1		**No. of lectures**
	• Questionnaire: its reliability and validity. • Types of statistical research: empirical, field experiments, laboratory experiments, and secondary sources of data, exploratory and confirmatory research, planned and ad-hoc methods of data collection, non-response and methods of recovering the missing response. • Missing values and types of missingness, imputations methods for missing values, single and multiple imputations	**15**
Unit 2		

	• Jack-Knife and bootstrap method of estimation, estimate of bias and standard error. • Simulation methods • Bagging and boosting methods with applications.	**15**

References:

- C. R. Kothari (2004) Research Methodology : Methods and Techniques 2nd Edition, New age International (p) Ltd Publications, New Delhi, India

- J.W. Creswell and J.D. Creswell (2017) Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5thEdition, SAGE Publications, USA.

- C. G. Thomas (2021) *Research Methodology and Scientific Writing, 2nd Edition*, Springer Nature, New York.

- M. Kheider lectures from University of Biskra (2017)

  https://univ-biskra.dz/sites/fll/images/houadjli%20Ahmed%20Chaouki.pdf

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester II

(2023 Course under NEP 2020)

Course Code: 23ScStaP211 **Course Name: Multivariate Analysis.**

Teaching Scheme: 4 hours/week **Credits: 4**

Examination Scheme: CIA: 50 Marks **End-Sem: 50 Marks**

Prerequisite Courses:

- Matrix algebra
- Univariate statistical inference

Course Objectives:

- To develop understanding of appropriate and relevant methods of multivariate data analysis.

Course Outcomes:

- A student is able to perform analysis of multivariate data.

Course Contents:

Unit 1		15 lectures
	Exploratory multivariate Data Analysis : Sample mean vector, Dispersion Matrix, Correlation Matrix, Linear transformation and its mean and variance, graphical interpretation. Principal component Analysis (by using covariance and correlation method, standardized method), Factor analysis (models, rotation types), Canonical correlation with real life examples.	
Unit 2		16 lectures
	Cluster analysis (Hierarchical and Nonhierarchical, Agglomerative, Single, complete, average , Wald's linkage , K-mean clustering method , qualitative method clustering) Multivariate normal distribution, Singular and non-singular normal distribution, mean, variance of multivariate normal distribution, independence of variables, M.G.F, Characteristic function, moments, Distribution of linear and quadratic form of normal variables, marginal and conditional distribution, multiple and partial correlation coefficient (3 random variable case) with examples on each of the topic.	
Unit 3		16 lectures
	MLES of parameters of multivariate normal distribution and their sampling distribution, Tests and confidence region for the mean when dispersion matrix is known, Wishart distribution (generalized case of chi-square) and its properties, Hotelling T^2statistics and uses of its distribution, Beharen- Fishers problem, confidence region for mean vector of multivariate normal distributions.	

Unit 4		13 lectures
	MANOVA technique , Likelihood ratio test , Test for equality of dispersion matrices Discrimenant analysis (by using prior probabilities, by using cost), Fisher Discriminate analysis, Mahalanobis D^2 Statistics	

Reference Books:

1. Anderson, T. W. (1984). Introduction to Multivariate Analysis, John Wiley
2. Fang ,K., Kotz, S., Ng K. W. (1990). Symmetric Multivariate and Related Distributions, Chapman and Hall
3. Härdle, W. K. &Simar, L. (2012). Applied Multivariate Statistical Analysis, Springer, New York
4. Härdle, W. K., Hlávka, Z. (2007). Multivariate Statistics Exercises and Solutions, Springer, New York
5. Johnson R.A. &Wichern, D.W. (1988). Applied Multivariate Statistical Analysis, Prentice Hall Inc.
6. Kotz, S., Balakrishnan N. and Johnson N. L. (2000). Continuous Multivariate Distributions, Volume 1, Models and Applications, John Wiley & Sons,
7. Kshirsagar, A. M. (1983). Multivariate Analysis, Marcel Dekker
8. Mardia,K. V. and Jupp, P. E. (2000), Directional Statistics, John Wiley & Sons
9. Morrison, D.F. (1990). Multivariate Statistical Methods, McGraw Hill Co.

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester II

(2023 Course under NEP 2020)

Course Code: 23ScStaP212 **Course Name: Regression Analysis**

Teaching Scheme: 4 hours/week **Credit: 4**

Examination Scheme: CIA: 50 Marks **End-Sem: 50 Marks**

Prerequisite Courses:

- Matrix algebra
- Multivariate normal distribution results

Course Objectives:

- A student should be able to handle multivariate data
- A student should be able to predict response variable based on regressors**.**

Course Contents:

Unit 1		15 lectures
	• Multiple regression models • Estimation of the model parameters • Hypothesis testing and confidence interval • Model adequacy checking: residual analysis, detection and treatment of outliers, lack of fit of regression model • Detection and treatment of multicollinearity, ridge regression	
Unit 2		15 lectures
	• Transformations and weighting to correct model inadequacies: variance stabilizing transformations, transformations to linearize the model, analytical methods for selecting a transformation, generalized and weighted least squares • Diagnostics for leverage and influence • Polynomial regression : polynomial models in one variable, polynomial models in two or more variables, non-parametric regression	
Unit 3		15 lectures
	• Variable selection and model building: computational techniques for variable selection (all possible regressions, forward selection method, backward elimination method, stepwise regression methods) • Introduction to nonlinear regression: nonlinear regression model, nonlinear least square, transformation to a linear model, parameter estimation in a nonlinear system, statistical inference in nonlinear regression.	
Unit 4		15 lectures

	• Generalized Linear Models(GLM): link functions and linear predictors, parameter estimation and inference in the GLM, prediction and estimation with the GLM, residual analysis in the GLM, logistic regression models, Poisson regression models • Validation of regression models: validation techniques	

Reference books:

1. Cameron, A. C. and P. K. Trivedi (1998). Regression Analysis of Count Data, Cambridge
2. Draper, N. R. and Smith, H. (1998). Applied Regression Analysis, John Wiley, ThirdEdition.
3. Hosmer, D. W. and Lemeshow, S. (1989). Applied Logistic Regression, Wiley.
4. Kleinbaum, D. G. & Klein, M. (2002). Logistic Regression A Self-Learning Text, Springer
5. McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models, Chapman& Hall.
6. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). Introduction to LinearRegression Analysis, Wiley.
7. Neter, J., W., and Kutner, M. H. (1985). Applied Linear Statistical Models, Wiley.
8. Ratkowsky, D. A. (1983). Nonlinear Regression Modelling, Marcel Dekker, London.
9. Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) Semiparametric Regression, CambridgeUniversity Press.
10. Seber, G. E. F. and Wild, C. J. (1989). Nonlinear Regression, Wiley.
11. Weisberg, S. (2005). Applied Liner Regression, Wiley.

Progressive Education Society's

Modern College of Arts, Science and Commerce,
Shivajinagar, Pune - 5
M.Sc. (Statistics) Semester II
(2023 Course under NEP 2020)

Course Code: 23ScStaP213 **Course Name: Statistical Inference**

Teaching Scheme: 4 hours /week **Credit: 04**

Examination Scheme: CIA: 50 Marks **End-Sem: 50 Marks**

Prerequisite Courses:

- Probability distribution theory

Course Objectives: On completion of this course the student will be able to understand

- Concept of Fisher information.
- Concept of sufficiency and minimal sufficiency.
- Concept of completeness.
- Concept of MVUE
- Testing of hypothesis
- Interval estimation

Course Outcomes:

At the end of the course, a student learns point and interval estimation theory and testing of hypothesis theory.

Course Contents:

Chapter 1		18 lectures
	• Fisher information and information matrix • Concept of Sufficiency, Neymann factorization theorem, • likelihood equivalence, minimal sufficiency, construction of minimal sufficient statistics • Special classes of distributions: one parameter exponential family, multi parameter exponential family, Pitman family, minimal sufficient statistic for special classes of distributions.	
Chapter 2		18 lectures
	• Completeness, complete sufficient statistics, special classes of distributions admitting complete sufficient statistics • Ancillary statistic, Basu's theorem and its application • Estimability of parametric function, Cramer-Rao inequality,minimum variance unbiased estimators (MVUE), necessary and sufficient conditions for existence of MVUE, Minimum variance bound unbiased estimators (MVBUE) • Rao- Blackwell theorem, Lehman- Scheffe theorem.	
Chapter 3		16 lectures

	• Problem of testing of hypothesis, simple and composite hypotheses. randomized and non-randomized tests, most powerful test, Neyman-Pearson Lemma (with proof), powerfunction of a test, existence of UMP tests for one-sided alternatives in one parameter exponential family and Pitmanfamily UMP tests for two sided alternatives, their existence and non-existence, unbiased test, UMP unbiased test • Monotone likelihood ratio property and its applications.	
Chapter 4		8 lectures
	• Confidence interval: Concept of confidence intervals Relationship with testing of hypothesis • shortest expected length confidence intervals, uniformly most accurate confidence bounds.	

Reference Books:

1. Casella, G. and Berger, R. L. (2002). Statistical Inference. Duxbury Advanced Series, Second Edition.
2. Efron, B. and Hastie, T. (2016). Computer Age Statistical Inference Algorithms, Evidence and Data Science. Cambridge University Press
3. Kale, B.K. & Muralidharan, K. (2015) Parametric Inference An Introduction, Alpha Science International Ltd.
4. Lehmann, E.L. and Casella, G. (1998). Theory of Point Estimation. Springer, New York
5. Lehmann, E. L. and Romano, J. (2005). Testing Statistical Hypotheses, Springer
6. Rao, C. R. (1995). Linear Statistical Inference and its Applications, Wiley
7. Rohatgi, V. K. and Saleh, A.K. Md. E. (2001). Introduction to Probability and Statistics, John Wiley & Sons, New York.
8. Shao, J. (2003). Mathematical Statistics, Springer-Verlag, New, New York.

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester II

(2023 Course under NEP 2020)

Course Code: 23ScStaP214

Course Name: Lab Course on 23ScStaP211, 23ScStaP212 and 23ScStaP213

Teaching Scheme: TH: 4 Lectures /Week **Credit: 02**

Examination Scheme: CIA: 25 Marks **End-Sem: 25 Marks**

Prerequisite Courses:

- Regression analysis
- Multivariate analysis

Course Objectives:

- To acquaint students with multivariate data analysis techniques using statistical software.

Course Outcomes:

- After learning this course student will be able to apply regression analysis and multivariate analysis techniques to real life data.

Course Contents:

Title and Content
1. Exploratory Multivariate data Analysis. (Sample mean, variance and covariance matrix, Correlation Matrix) 2. Principal component Analysis (covariance & Correlation technique and their interpretation) 3. Factor analysis (PCA., MLE, all Rotations and their interpretation) 4. Cluster analysis (Single, Complete, Average, Wards , k- mean linkage method) 5. Canonical correlation Multivariate Analysis. (Multivariate normality, Marginal, Conditional, Q-Q plot, contour plot) 6. Model Sampling from Multivariate Distribution, and computation of MLE's of Parameters. 7. Application of Hotelling T^2 Statistic 8. MANOVA technique 9. Likelihood Ratio Test(Equality of means, Equality of variance, R=0) 10. Discriminant Analysis (Fishers linear discriminant function) 11. Selection of variables in Multiple regression and lack of fit 12. Transformation and weighting to correct model inadequacies. 13. Polynomial regression model (one and two regressors), Multicolinarity and ridge regression. 14. Spline Regression 15. Generalized linear models : Logistic Regression & Poisson Regression

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester II

(2023 Course under NEP 2020)

Course Code: 23ScStaP221 **Course Name: Probability Theory**

Teaching Scheme: 4 hours/week **Credit: 4**

Examination Scheme: CIA: 50 Marks **End-Sem: 50 Marks**

Pre requisite Courses:

- Basic concepts of probability.
- Basics of Real analysis and Calculus.

Course Objectives:

- To understand the Basics of Probability theory.

Course Outcomes:

- Student is able to solve the problems of probability theory and Convergence of random variables.

Course Contents:

Chapter 1		15 lectures
	1.1 Review of algebra of sets, sequence of sets, limsup, liminf and limit of a sequence of sets, classes of sets, field, sigma field, minimal sigma field, Borel fields, measurable space, monotone classes, measurable function **1.2** Real and Vector valued random variables, simple random variable, random variable as a limit of sequence of simple random variables. 1.3 Probability measure on a measurable space, probability space, properties of probability measure: continuity, mixture of probability measures, Lebesgue and Lebesgue-Steltjes measures.	
Chapter 2		15 lectures
	2.1 Distribution function, decomposition of a distribution function, discrete and continuous type random variable, Correspondence theorem, 2.2 Expectation of simple random variable, non-negative random variable, arbitrary random variable, properties of expectation, moments, moment inequalities.	
Chapter 3		15 lectures
	3.1 Convergence of a sequence of random variables, convergence in probability, convergence in distribution, convergence in rth mean, almost sure convergence, their inter-relations, Slutkey's Theorem, convergence theorem for expectations**.**	

Chapter 4		15 lectures
	4.1 Independence of events, class of independent events, independence of classes, independence of random variables, expectation of the product of independent random variables, equivalent definitions of independence, 4.2 Kolmogorov 0-1 Law, Borel 0-1 criterion, Khintchin's WLLN, Strong Law of Large Numbers (SLLN) (Statement only), Central Limit Theorem (CLT), Levy continuity theorem, CLT for i.i.d. random variables, Liaponove's form, Lindeberg Feller form and their applications.	

Reference Books:

1. Athreya, K. B. and Lahiri S. (2006). Probability Theory, Hindustan Book Agency,
2. Bhat, B. R. (2007). Modern Probability Theory: An Introductory Text Book, New Age Internationals.
3. Billingsley, P. (1995). Probability and Measure, 3rdEdition, John Wiley, New York
4. Chung, K. L. (2001). A Course in Probability Theory, Third Edition, Academic Press, London
5. Gut, Allan (2005), Probability: A Graduate Course. Springer, New

Progressive Education Society's

Modern College of Arts, Science and Commerce,

Shivajinagar, Pune - 5

M.Sc. (Statistics) Semester II

(2023 Course under NEP 2020)

Course Code: 23ScStaP222

Course Name: Statistical methods in epidemiology

Teaching Scheme: 4 hours/week **Credit : 4**

Examination Scheme: CIA: 50 Marks **End-Sem: 50 Marks**

Prerequisites:

- Basic understanding of probability distributions.

Course Objective:

- The Course aims to provide students with the key statistical knowledge and skills needed to analyze and interpret data associated with epidemics/pandemics.

Course Outcomes: Students who successfully complete this course should be able to:

- Utilize the basic terminology and definitions of epidemiology, describe key features and applications of descriptive and analytic epidemiology.
- Use statistical techniques in the analysis, predictions and presentation of epidemiological data.
- Calculate and interpret ratios, proportions, incidence rates, mortality rates, prevalence rate etc.
- To model the spread and related aspects of a given epidemics so as to gain insight into its management, to use statistical methods for analyzing the shut down strategy, testing strategy, vaccination strategy etc.

Course Contents:

Unit 1		15 lectures
	Epidemiologic terms and parameters: Infection period, incubation period, latent period, number of asymptomatic carriers, disease frequency, disease frequency association, concept of prevalence, measures of risk, reproduction numbers, preventive reproduction numbers, infection rate, fatality rate, transmission intensity, doubling time, flattening of the curve, prevention strategies	
Unit 2		7 lectures
	Concepts of disease occurrence, chains of infections, disease occurrence patterns, SIR epidemic models, Reed-Frost chain binomial epidemic models, SEIR models, random networks for epidemics, models for spatiotemporal spread, incorporating the effects of interventions, predicting the course of the spread	
Unit 3		15 lectures
	Epidemiological study designs, cohort studies, case-control studies, randomized control studies, Intervention, Statistical Inference for	

	the epidemiological parameters, Bayesian inference for latent (unobserved) variables (MCMC, adaptive MCMC); Inference for outbreaks when virus sequence data is available, Methods based on evolution of the virus by comparing virus sequences of diagnosed cases.	
Unit 4		8 lectures
	Testing, Sensitivity, Specificity and ROC Curve related to diagnostic testing, Sample size determination, Pooled testing, Composite sampling,	
Unit 5		15 Lectures
	COVID 19 related topics, Mathematical models for COVID 19 spread, Applications of SIR and SEIR models, Assessment of Lock down effect, Introduction to spatial epidemiology such as spatial exploration of epidemiological data, quantification of spatial patterns and clusters, spatial exposure assessment, methods for assessing risk with examples/models from COVID 19	

Reference Books:

1. Armitage, P., Berry, G. and Matthews, J. N. S. (2002). Statistical Methods in Medical Research, Wiley.
2. Becker, N. G. (2015). Modeling to Inform Infectious Disease Control, CRC Press.
3. Elston, R. C. and Johnson W. D. (2008). Basic Biostatistics for Geneticists and Epidemiologists: A Practical Approach, Wiley
4. HardeoSahai and Khushid, A. (2009). Statistics in Epidemiology: Methods, Techniques and Applications
5. Krämer, A. Kretzschmar, M. and Krickeberg, K. (Editors) (2010). Modern Infectious Disease Epidemiology: Concepts, Methods, Mathematical Models, and Public Health, Springer
6. Lawson, A. B. (2006). Statistical Methods for Spatial Epidemiology, Wiley
7. Lawson, A. B. (2018). Bayesian disease mapping: Hierarchical Modeling in Spatial Epidemiology, CRC Press
8. Marschner, I. C. (2014). Inference Principles for Biostatisticians, CRC Press
9. Merril, R. M. (2015). Statistical Methods in Epidemiologic Research
10. Merrill, R. M. (2012). Fundamentals of Epidemiology and Biostatistics
11. Nigel, B., Daniel, P. and Debbi, S. (2018). Quantitative Methods for Health Research : A Practical Interactive Guide to Epidemiology and Statistics, Wiley
12. Pagano, M. and Gauvreau, K. (2018). Principles of Biostatistics, CRC Press
13. Selvin, S. (1991). Statistical Analysis of Epidemiological Data
14. Stewart, A. (2016). Basic Statistics and Epidemiology: A Practical Guide, Fourth Edition, CRC Press
15. Sullivan, L. M. (2018). Essentials of Biostatistics in Public Health, 3rd Edition, Jones & Bartlett Learning.