Rayat Shikshan Sanstha's

SADGURU GADGE MAHARAJ COLLEGE, KARAD

(An Autonomous College)

SYLLABUS

For

M.Sc. Statistics

Part I (Semester I and II)

As per NEP-2020

To be implemented from academic year 2023-24

M.Sc. Statistics Part I Syllabus: w.e.f. academic year 2023-24

A. Ordinance and Regulations: - (as applicable to degree/ programme)

1. Title of the Programme: **M. Sc. (Statistics)**
2. Faculty: Science and Technology
3. Year of Implementation: M.Sc. part-I: Academic year 2023-24
4. Learning outcomes

   - **Programme Outcomes (POs)**

     Post Graduates of the M.Sc. Statistics programme will be able to:

     1) Have sufficient knowledge of theoretical concepts in Statistics for (i) choosing and applying the most appropriate statistical methods/ techniques for collecting and analyzing data (ii) interpreting the results of analyses in relation to given real life situations.
     2) Have deep understanding and ability to explain the inter-connections between various sub disciplines and apt use of these inter-connections in modelling real life problems.
     3) Have ability to recognize the importance of statistical thinking and training, and to acquire the state-of-the-art developments in Statistics independently from available resources.
     4) Develop expertise in data management and analysis using widely used statistical software.

   - **Programme Specific Outcomes (PSOs)**

     After completion of M.Sc. Statistics programme the student will be able to:

     1) Develop stochastic models for studying, analyzing, interpreting and forecasting real life phenomenon in diverse disciplines.
     2) Effectively use necessary statistical software and computing environment including R, Python and MS-EXCEL among others.
     3) Have the versatility to work effectively in a broad range of establishments (including R&D sectors, analytics, scientific laboratories, government, financial, health, educational) or to continue for higher education, and exhibit ethical and professional behaviour in team work.

5. Duration: 2 Years
6. Fee Structure: As per university rules
7. Eligibility criteria for Admission: Three-year B. Sc. Degree with Statistics as principal / major subject
8. Medium of Instruction: English
9. Exit Option at Level 6: Students can exit after Level 6 with Post Graduate Diploma in Statisticsif he/she completes the courses of minimum 44 credits.

10. Structure of the Programme, Scheme of Teaching and Examination

Sem.	Course type	Course code	No. of credits	Teaching hours per week	Examination Scheme				
					University Assessment			Internal Assessment	
					Max Marks	Min marks for passing	Exam hours	Max Marks	Min marks for passing
I	Major Mandatory	MJ-MST23-101	4	4	80	32	3	20	8
		MJ-MST23-102	4	4	80	32	3	20	8
		MJ-MST23-103	2	2	40	16	1.5	10	4
		MJ-MSP23-106	4	12	--	--	--	100	40
	Major Elective	GE-MST23-104	4	4	80	32	3	20	8
	Research Methodology	RM-MST23-105	4	4	80	32	3	20	8
II	Major Mandatory	MJ-MST23-201	4	4	80	32	3	20	8
		MJ-MST23-202	4	4	80	32	3	20	8
		MJ-MST23-203	2	2	40	16	1.5	10	4
		MJ-MSP23-206	4	12	--	--	--	100	40
	Major Elective	GE-MST23-204	4	4	80	32	3	20	8
	On Job Training/ Field Project	FP-MST23-205	4	4	To be declared				
III	Major Mandatory	MJ-MST23-301	4	4	80	32	3	20	8
		MJ-MST23-302	4	4	80	32	3	20	8
		MJ-MST23-303	2	2	40	16	1.5	10	4
		MJ-MSP23-306	4	12	--	--	--	100	40

	Major Elective	GE-MST23-304	4	4	80	32	3	20	8
	Research Project	RP1-MST23-305	4	4	--	--	--	100	40
IV	Major Mandatory	MJ-MST23-401	4	4	80	32	3	20	8
		MJ-MST23-402	4	4	80	32	3	20	8
		MJ-MSP23-405	4	12	--	--	--	100	40
	Major Elective	GE-MST23-403	4	4	80	32	3	20	8
	Research Project	RP2-MST23-404	6	6	--	--	--	150	60

11. Courses being offered

Sem	Course type	Course code	No. of credits	Course title
I	Major Mandatory	MJ-MST23-101	4	Distribution Theory
		MJ-MST23-102	4	Estimation Theory
		MJ-MST23-103	2	Statistical Computing
		MJ-MSP23-106	4	Practical-I
	Major Elective	GE-MST23-104	4	Statistical Mathematics
		GE-MST23-104a		Real Analysis
		GE-MST23-104b		Linear Algebra
	Research Methodology	RM-MST23-105	4	Research Methodology in Statistics
II	Major Mandatory	MJ-MST23-201	4	Theory of Testing of Hypothesis
		MJ-MST23-202	4	Linear Models and Regression Analysis
		MJ-MST23-203	2	Statistical Programming using Python
		MJ-MSP23-206	4	Practical-II
	Major Elective	GE-MST23-204	4	Probability Theory
		GE-MST23-204a		Bayesian Inference
		GE-MST23-204b		Reliability Theory
		GE-MST23-204c		Circular Data Analysis
	On Job Training/ Field Project	FP-MST23-205	4	--
III		MJ-MST23-301	4	Stochastic Processes

	Major Mandatory	MJ-MST23-302	4	Data Mining
		MJ-MST23-303	2	Multivariate Analysis
		MJ-MSP23-306	4	Practical-III
	Major Elective	GE-MST23-304	4	Design and Analysis of Experiments
		ME32		Econometrics
		ME33		Functional Data Analysis
	Research Project	RP1-MST23-305	4	--
IV	Major Mandatory	MJ-MST23-401	4	Biostatistics
		MJ-MST23-402	4	Time Series Analysis
		MJ-MSP23-405	4	Practical-IV
	Major Elective	GE-MST23-403	4	Optimization Techniques
		ME42		Statistical Quality Control
		ME43		Spatial Data Analysis
		ME44		Actuarial Statistics
	Research Project	RP2-MST23-404	6	--

12. Standard of passing: 40% in each course. Separate passing for internal and semester examinations.
13. Nature of Question paper and Scheme of marking for University examination
    - **Nature of the theory question papers (4 credits):**
        a) There shall be 7 questions each carrying 16 marks.
        b) Question No.1 is compulsory. It consists of 8 questions for 2 marks each.

c) Students have to attempt any 4 questions from question No. 2 to 7.
d) Question No. 2 to 6 shall contain 2 to 4 sub-questions.
e) Question No. 7 shall contain 4 short note type questions, each carrying 4 marks.

- **Nature of the theory question papers (2 credits):**

a) There shall be 4 questions.
b) Question No.1 is compulsory. It consists of 4 questions for 2 marks each.
c) Question No. 2 to 4 shall be of 16 marks each.
d) Students have to attempt any 2 questions from question No. 2 to 4.
e) Question No. 2 to 4 shall contain 2 to 4 sub-questions.

14. Nature of Practical examination: -

Component	**Max marks**
Practical examination: Examination will be of 3 hour duration. There shall be 8 questions each of 12 marks, of which a student has to attempt any 5 questions.	60
Day-to-day practical performance and journal	20
Viva: Viva will be based on all practical’s	20

SEMESTER I

MJ-MST23-101 DISTRIBUTION THEORY

Course outcomes:

Upon successful completion of this course, the students will be able to:

i) Understand and explain the concept of univatiate and multivariate random variables and related entities
ii) Understand and explain the nature of various probability distributions and perform related computations.
iii) Understand probability models for multivariate data and perform related computations
iv) Understand non-central sampling distribution and able to perform their applications, able to perform computations related to order statistics

Unit 1: Random variable, Cumulative distribution function (CDF) and its properties, continuous and discrete distributions, mixtures of probability distributions, decomposition of mixture CDF into discrete and continuous CDFs, computation of probabilities of events using CDF, expectation and variance of mixture distributions. Quantiles of probability distributions. Transformations of univariate random variables, probability integral transformation.

(12L+3T)

Unit 2: Concepts of location, scale and shape parameters of distributions with examples. Symmetric distributions and their properties. Moment inequalities: Basic, Holder, Markov, Minkowski, Jensen, Chebyshev's inequalities, and their applications. Random vectors, joint distributions, Independence, variance-covariance matrix, joint MGF. Conditional expectation and variances. Transformations of bivariate random variables, Convolutions, compound distributions.

(12L+3T)

Unit 3: Multivariate normal distribution, two definitions and their equivalence, singular and nonsingular normal distribution, characteristic function, moments, marginal and conditional distributions. Maximum likelihood estimators of the parameters of the multivariate normal distribution and their sampling distributions. Wishart matrix and its distribution (statement only), properties of Wishart distribution, distribution of generalized variance(statement only). Marshall-Olkin bivariate exponential distributions.

(12L+3T)

Unit 4: Non-central chi-square, t and F distributions, distributions of linear and quadratic forms involving normal random variables, Fisher Cochran and related theorems (statements only) and their applications. Order Statistics: Distribution of an order statistics, joint distributions of two order statistics, distribution of spacings, normalized spacings with illustration to exponential case, distribution of sample median and sample range.

(12L+3T)

References:

1. Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.
2. DasGupta, A. (2010). *Fundamentals of probability: A first course*. Springer Science & Business Media.
3. Johnson N. L. & Kotz. S. (1996). *Distributions in Statistics Vol-I, II and III*. JohnWiley and Sons New York.
4. Kotz, S., Balakrishnan, N., & Johnson, N. L. (2004). *Continuous multivariate distributions, Volume 1: Models and applications* (Vol. 1). John Wiley & Sons.
5. Rao C. R. (1995). *Linear Statistical Inference and Its Applications.* John Wiley & Sons.
6. Rohatagi V. K. & Saleh A. K. Md. E.(2001). *Introduction to Probability Theory and Mathematical Statistics*. John Wiley and sons Inc.

MJ-MST23-102: ESTIMATION THEORY

Course outcomes:

Upon successful completion of this course, the students will be able to:

i) Explain the principles of data reduction and obtain sufficient, minimal sufficient, and complete statistics for various families of distributions.
ii) Obtain UMVUE of parameters of various distributions and determine Cramer-Rao and Chapman-Robbins-Kiefer lower bounds for the variances of unbiased estimators.
iii) Apply parametric and nonparametric methods to obtain estimators.
iv) Obtain CAN and BAN estimators.

Unit 1: Principles of data reduction: sufficiency principle; sufficient statistics; factorization theorem; minimal sufficient statistic; minimal sufficient statistic for exponential family, power series family, curved exponential family, and Pitman family; completeness; bounded completeness; ancillary statistics, Basu's theorem and it's applications.

(12L + 3T)

Unit 2: Unbiased estimation: unbiased estimator; uniformly minimum variance unbiased estimator (UMVUE); A necessary and sufficient condition for an estimator to be UMVUE; Rao-Blackwell theorem and Lehmann-Scheffe theorem, and their applications in finding UMVUEs; Fisher information function and Fisher information matrix; Cramer-Rao lower bound; Chapman-Robbins-Kiefer lower bound.

(12L + 3T)

Unit 3: Methods of finding estimators: method of moments estimator; maximum likelihood estimator (MLE), properties of MLE, MLE in nonregular families; method of scoring; method of minimum chi-square, EM algorithm. Nonparametric estimation: degree of an estimable parameter, kernel, U-statistic and its properties.

(12L + 3T)

Unit 4: Consistency of an estimator, weak and strong consistency, joint and marginal consistency, invariance property under continuous transformations, methods of constructing consistent estimators, Comparison of consistent estimators, asymptotic relative efficiency, minimum sample size required by the estimator to attain certain level of accuracy. Consistent Asymptotic Normal (CAN) Estimators: Definition of CAN estimator for real and vector valued parameters, invariance of CAN property under non-vanishing differentiable transformation (delta method). Methods of constructing CAN estimators: Method of Moments, method of percentiles, comparison of CAN estimators. BAN estimators, CAN and BAN estimators in one parameter and multi-parameter exponential family of distributions, Crammer family of distributions, Cramer – Huzurbazar theorem (Statement only).

(12L + 3T)

References:

1. Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.
2. Deshmukh S., Kulkarni M. (2022). *Asymptotic Statistical Inference: A Basic Course Using R*. Springer Verlag, Singapor.
3. Dudewicz, E. J., & Mishra, S. (1988). *Modern mathematical statistics*. John Wiley & Sons, Inc.
4. Kale, B. K., & Muralidharan, K. (2015). *Parametric inference: An introduction*. Alpha Science International Limited.
5. Lehmann, E. L. (1983). *Theory of Point Estimation.* John Wiley & sons.
6. Mukhopadhyay, P. (2015). *Mathematical Statistics*, Books and Allied (p) Ltd.
7. Rao C. R. (1995). *Linear Statistical Inference and Its Applications.* John Wiley & Sons.
8. Rohatgi, V. K., & Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons.

MJ-MST23-103: STATISTICAL COMPUTING

Course outcomes:

Upon successful completion of this course, the students will be able to:

i) Perform data organization, data manipulation, statistical and mathematical computations, and data analysis using MSEXCEL.
ii) Perform data organization, data manipulation, statistical and mathematical computations, and data analysis using R.

Unit 1: MSEXCEL: Introduction to MSEXCEL. Cell formatting, conditional formatting, Data manipulation using EXCEL: sort and filter, find and replace, text to columns, remove duplicate, data validation, consolidate, what-if-analysis. Working with Multiple Worksheets and Workbooks. Built-in mathematical and statistical functions for obtaining descriptivestatistic, computing PMF/PDF, CDF and quantiles of the well-known distributions, rand and randbetween function, Logical functions: if, and, or, not. Lookup functions: hlookup, vlookup, Formula Errors, Creating and Working with Charts, Database functions, Text functions, Date and time functions, Excel add-ins: analysis tool pack, Pivot tables and charts.

(12L+3T)

Unit 2: R-software: Introduction to R, data types and objects, operators, data input, data import and export, built in functions for descriptive statistics, random sampling and computation of pdf, cdf and quantiles of well known distribution. Strings and Dates in R. apply family of functions. Saving work in R. Matrix algebra, graphical procedures, frequencies and cross tabulation, built in functions: lm, t.test, prop.test, wilcox.test, ks.test, var.test, chisq.test, aov. Control statements. Programming, user defined functions, R-packages. R-studio.

(12L+3T)

References:

1. Gardener, M. (2012). Beginning R: the statistical programming language. John Wiley & Sons.
2. Held, B., Moriarty, B., & Richardson, T. (2019). Microsoft Excel Functions and Formulas with Excel 2019/Office 365. Mercury Learning and Information.
3. Herkenhoff, L., & Fogli, J. (2013). Applied statistics for business and management using Microsoft Excel. New York: Springer.
4. Purohit, S. G., Gore, S. D., & Deshmukh, S. R. (2015). Statistics using R. Alpha Science International.
5. Thulin, M. (2021). Modern Statistics with R: From wrangling and exploring data to inference and predictive modelling. BoD-Books on Demand.
6. Weblinks:
   https://support.microsoft.com/en-us/excel
   https://cran.r-project.org/manuals.html

GE-MST23-104: STATISTICAL MATHEMATICS

Course outcomes:

Upon successful completion of this course, the students will be able to:

i) Explain the vector Space, its dimension, and linear dependence/independence of vectors, and able to perform relation operations
ii) Understand matrix theory and perform matrix operations
iii) Understand sequences and series of real numbers and their convergence
iv) Understand the concept of real valued function, continuity of functions, convergence of series of functions, integration of functions, and obtain optima of a function.

Unit 1: Vectors, linear dependence and independence of vectors, vector space, subspace, basis, dimension of a vector space, example of vector spaces. Gram-Schmidt orthogonalisation process, Orthonormal basis, orthogonal projection of a vector, Linear transformations, algebra of matrices, types of matrices, row and column spaces of a matrix, elementary operations and elementary matrices, rank and inverse of a matrix, null space and nullity, partitioned matrices.

(12L+3T)

Unit 2: Generalized inverse, Vector and Matrix differentiation, Spectral decomposition of a real symmetric matrix, singular value decomposition, Choleskey decomposition, real quadratic forms, reduction and classification, index and signature, extrema of a quadratic form, simultaneous reduction of two quadratic forms.

(12 L + 3 T)

Unit 3: Sequences of real numbers, convergence, divergence, monotone, bounded and unbounded sequences, Cauchy sequence, Convergence of bounded monotone sequence. Limit points, Limit inferior and limit superior of the sequences and their properties. Subsequences and properties associated with them. Series of numbers, tests for convergence (without proof) test for absolute convergence, convergence of series of non-negative terms.

(12 L + 3 T)

Unit 4: Real valued functions, continuous functions, Uniform continuity of functions and sequences of functions, Uniform convergence of series of functions with special emphasis on power series, radius of convergence. Riemann, Riemann-Steltjes Integrals and their common properties. integrability of functions, Fundamental theorem on calculus, mean value theorem, their applications in finding functional of probability distributions. Maxima, minima of functions of several variables. Constrained maxima, minima, Lagrange’s method, Taylor’s theorem (without proof), Multiple and Improper integrals, their applications in multivariate probability distributions. Theorem on differentiation under integral sign and Leibnitz rule (statements only) with applications.

(12 L + 3 T)

References:

1. Apostol (1985). *Mathematical Analysis.* Narosa Publishing House, T.M.
2. Bartle, R. G., & Sherbert, D. R. (2000). *Introduction to real analysis* (Vol. 2, p. 2). New York: Wiley.
3. Goldberg, R. R. (1970). *Methods of real analysis*. Oxford and IBH Publishing.
4. Hadely G. (1962). *Linear Algebra.* Narosa Publishing House.
5. Malik, S. C., & Arora, S. (1992). *Mathematical analysis*. New Age International.
6. Narayan, S., Raisinghania M. D. (2013). *Elements of Real Analysis*. S. Chand.
7. Rao C. R. (1995). *Linear Statistical Inference and Its Applications.* John Wiley & Sons
8. Rao, A. R., & Bhimasankaram, P. (2000). *Linear algebra* (Vol. 19). Springer.
9. Royden (1988). *Principles of Real Analysis*. Macmillian
10. Searl S. B.(2006). *Matrix Algebra Useful for Statistics*. Wiley.

GE-MST23-104a: REAL ANALYSIS

Course outcomes:

Upon successful completion of the requirements for this course, students will be able to:

i) Understand set theory, obtain the supremum and infimum of bounded sets, and limit point of a set
ii) Understand sequences and series, the convergence of sequence and series, and compute the limit of sequences.
iii) Examine continuity and uniform continuity of functions, sequences, and series of functions, and evaluate Riemann and Riemann-Steltjes Integrals.
iv) Differentiate vector and matrix valued functions, obtain optima of functions of several variables, determine the Taylor series expansion of functions, evaluate multiple integrals, and use in connection with multivariate distributions

Unit 1: Set of real numbers, countable and uncountable sets, countability of rational numbers,uncountability of the interval (0, 1) and other uncountable sets. Supremum and Infimum of bounded sets, limit point(s) of a set, closur of a set, open, closed, dense and compact sets and their properties. Bolzano-Weierstrass and Heine-Borel Theorems (Statements only). Applications of these theorems.

(12 L + 3 T)

Unit 2: Sequences of real numbers, convergence, divergence, monotone, bounded and unbounded sequences, Cauchy sequence, Convergence of bounded monotone sequence. Limit points, Limit inferior and limit superior of the sequences and their properties. Subsequences and properties associated with them. Series of numbers, tests for convergence (without proof) test for absolute convergence, convergence of series of non-negative terms.

(12 L + 3 T)

Unit 3: Real valued functions, continuous functions, Uniform continuity of functions and sequences of functions, Uniform convergence of series of functions with special emphasis on power series, radius of convergence. Riemann, Riemann-Steltjes Integrals and their common properties. Upper and lower integrals, integrability of functions, Integration by parts, Fundamental theorem on calculus, mean value theorem, their applications in finding functional of probability distributions.

(12 L + 3 T)

Unit 4: Vector and Matrix differentiation, Maxima, minima of functions of several variables. Constrained maxima, minima, Lagrange's method, Taylor's theorem (without proof), implicitfunction theorem and their applications. Multiple integrals, Change of variables, Improper integrals, Applications in multivariate probability distributions. Theorem on differentiation under integral sign and Leibnitz rule (statements only) with applications.

(12 L + 3 T)

References:

1. Apostol (1985). *Mathematical Analysis.* Narosa Publishing House, T.M.
2. Bartle G. R. (1976). *Element of Real Analysis.* Wiley, 2nd edition.
3. Bartle, R. G., & Sherbert, D. R. (2000). *Introduction to real analysis* (Vol. 2, p. 2). New York: Wiley.
4. Goldberg, R. R. (1970). *Methods of real analysis*. Oxford and IBH Publishing.
5. Malik, S. C., & Arora, S. (1992). Mathematical analysis. New Age International.
6. Narayan, S., Raisinghania M. D. (2013). Elements of Real Analysis. Fourteenth Revised Edition, S. Chand.
7. Royden (1988). Principles of Real Analysis. Macmillian

GE-MST23-104b : LINEAR ALGEBRA

Course outcomes:

Upon successful completion of this course, the student will be able to:

i) Understand the vector Space, its dimension, and linear dependence/independence of vectors.
ii) Understand the matrix theory and related computations.
iii) Solve systems of linear equations.
iv) Understand the quadratic form and its classification

Unit 1: Vector space, subspace, linear dependence and independence, basis, dimension of a vector space, example of vector spaces. Gram-Schmidt orthogonalisation process, Orthonormal basis, orthogonal projection of a vector, Linear transformations, algebra of matrices, types of matrices, row and column spaces of a matrix, elementary operations and elementary matrices, rank and inverse of a matrix, null space and nullity, partitioned matrices.

(12L+3T)

Unit 2: Permutation matrix, reducible/irreducible matrix, primitive/ imprimitive matrix, Kronecker product, Generalized inverse, Moore-Penrose generalized inverse, Solution of a system of homogenous and non-homogenous linear equations, theorem related to existence ofsolution and examples.

(12L+3T)

Unit 3: Characteristic roots and vectors of a matrix, algebraic and geometric multiplicities of a characteristic root, right and left characteristic vectors, orthogonal property of characteristic vectors, Cayley-Hamilton Theorem and its applications.

(12L+3T)

Unit 4: Spectral decomposition of a real symmetric matrix, singular value decomposition, Choleskey decomposition, real quadratic forms, reduction and classification, index and signature, extrema of a quadratic form, simultaneous reduction of two quadratic forms.

(12L+3T)

References:

1. Graybill, F. A. (1961). *An introduction to linear statistical models* (No. 04; HA29, G73 V. 1.).
2. Hadely G. (1962). *Linear Algebra.* Narosa Publishing House.
3. Harville D. (1997). *Matrix Algebra From Statistics Perspective*. Springer.
4. Rao C. R. (1995). *Linear Statistical Inference and Its Applications.* John Wiley & Sons
5. Schott, J. R. (2016). *Matrix analysis for statistics*. John Wiley & Sons.
6. Searl S. B.(2006). *Matrix Algebra Useful for Statistics*. Wiley.
7. Rao, A. R., & Bhimasankaram, P. (2000). *Linear algebra* (Vol. 19). Springer.

RM-MST23-105: RESEARCH METHODOLOGY IN STATISTICS

Course outcomes:

Upon successful completion of this course, the students will be able to:

i) Understand the concept of research, research process, and research ethics.
ii) Understand and apply various sampling methods for data collection and estimate the parameters.
iii) Understand the concept of simulation and able to simulate real life processes
iv) Estimate bias and standard error of an estimator using resampling techniques, apply, numerical methods to solve systems of linear equations, to obtain the roots of a nonlinear equation, and to solve definite integrals.

Unit 1: Meaning of research, objectives of research, motivation in research, types of research, research approaches, significance of research, research methods vs. methodology, research and Scientific method, research process, criteria of good research, defining research problem, research design, Research Ethics, publication of research, Plagiarism, Intellectual property rights, Patents and its filing procedures.

(12L+3T)

Unit 2: Sampling techniques: review of simple random sampling stratified random sampling, systematic random sampling, cluster sampling, two phase sampling, ratio and regression method of estimation. Probability proportional to size sampling: Cumulative total method, Lahiri's method, Hansen-Horwitz estimator and its properties, Horwitz-Thompson estimator, Des Raj estimators for a general sample size. Non-sampling errors, techniques for handling non-response: Hansen–Horwitz and Demings model for the effect of call-backs. Randomised response techniques, dichotomous population, Warners model, MLE in Warners model, unrelated question model.

(12L+3T)

Unit 3: Concept and need of simulation, random number generator, true random number and pseudo random number generators, requisites of a good random number generator. Tests for randomness. Congruential method of generating uniform random numbers. Algorithms for generating random numbers from well-known univariate discrete and continuous distributions, generating random vectors from multinomial, bivariate normal, and bivariate exponential distributions, generating random numbers from mixture of distributions (related results without proofs). Acceptance-Rejection Technique. Use of random numbers to evaluateintegrals, to study the systems involving random variables, to estimate event probabilities and to find expected value of random variables. Use of random numbers for performance evaluation of estimators and statistical tests.

(12L+3T)

Unit 4: Resampling methods: Bootstrap methods, estimation of bias and standard errors, estimation of sampling distribution, confidence intervals. Jackknife method: estimation of bias and standard errors, bias reduction method. Numerical methods for solution to system of linear equations: Jacobi and Gauss-Seidel methods with convergence analysis. Numerical methods for finding roots of nonlinear equation: Newton-Raphson method, bisection method; Newton-Raphson for system of non- linear equations. Numerical integration: quadrature formula, trapezoidal rule and Simpson's rules for single integral.

(12L+3T)

References:

1. Atkinson, K. E. (1989). *An introduction to numerical analysis*, John Wiley and Sons.
2. Chaudhuri, A., & Stenger, H. (2005). *Survey sampling: theory and methods*. CRC Press.
3. Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.
4. Devroye L. (1986). *Non- Uniform Random Variate Generation.* Springer- Verlag New York.

5. Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
6. Kennedy, W. J., & Gentle, J. E. (2021). *Statistical computing*. Routledge.
7. Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.
8. Morgan, B. J. (1984). *Elements of simulation* (Vol. 4). CRC Press.
9. Mukhopadhyay, P. (2008). *Theory and methods of survey sampling*. PHI Learning Pvt. Ltd..
10. Robert, C. P., Casella, G., & Casella, G. (1999). *Monte Carlo statistical methods* (Vol. 2). New York: Springer.
11. Ross, S. M. (2022). *Simulation*. Academic Press.
12. Rubinstein, R. Y., & Melamed, B. (1998). *Modern simulation and modeling* (Vol. 7). New York: Wiley.
13. Singh, D., & Chaudhary, F. S. (1986). *Theory and analysis of sample survey designs*. John Wiley & Sons.
14. Sukhatme P. V., Sukhatme S. & Ashok C (1984). *Sampling Theory of surveys and applications* . Iowa university press and Indian society of agricultural statistics, New Delhi.

MJ-MSP23-106: PRACTICAL –I

Course outcomes:

Upon successful completion of this course, the student will be able to:

1. Sketching of various distribution functions and finding possible probability distribution to observed Data.
2. Compute UMVUE, MME and MLE using various methods.
3. Perform mathematical/statistical computations, statistical data analysis using built-in functions in MS-EXCEL and R and develop programs for various tasks.
4. Perform/Solve various statistical problems though simulation, numerical and re-sampling techniques.

Practical List:

1. Sketching of PDF/PMF and CDF
2. Fitting probability distribution and related inference.
3. Probability plots for various univariate probability distributions and their interpretations.
4. Applications of multivariate normal distribution.
5. UMVUE, MME and MLE
6. Methods of Scoring and method of minimum chi-square estimation
7. Estimation using EM Algorithm
8. Verification of consistency and CAN property of estimators.
9. Computations using MSEXCEL
10. Statistical analysis using MSEXCEL
11. Computations using R
12. R-Programming
13. Formulations of research problem and its design
14. Sampling techniques - I
15. Sampling techniques - II
16. Applications of Simulation techniques
17. Numerical Methods and Resampling Techniques

18-20. Based on elective course

Semester II

MJ-MST23-201: THEORY OF TESTING OF HYPOTHESES

Course outcomes:

Upon successful completion of this course, a student will be able to:

i) Understand the concept of testing of hypothesis, test statistic, critical regions, size and power of a test.
ii) Develop and apply MP test, UMP test, UMPU test, similar tests and a test with Neyman structure.
iii) Obtain and interpret different interval estimates of parameters.
iv) Develop and apply large sample tests.

Unit 1: Problem of testing of hypothesis, null and alternative hypotheses, Simple and composite hypotheses, test function, Randomized and non-randomized tests, power function of a test, Most powerful (MP) test, Neyman-PearsonLemma, Monotone likelihood ratio property, Uniformly Most Powerful (UMP) test and its existence, determination of minimum sample size to achieve the desired strength of a test. Concept of p-value.

(12L+3T)

Unit 2: UMP tests for two sided alternatives, examples of their existence and non-existence. Unbiased test, Generalized Neyman Pearson lemma, UMPU tests and their existence in case of exponential families (Statements of the theorems only). Similar tests, test with Neyman structure. Tests for means of multivariate normal distributions based on Hotelling's T^2 Statistic.

(12L+3T)

Unit 3: Interval estimation: Confidence interval, relation with testing of hypotheses problem, Uniformly Most Accurate(UMA) and Uniformly Most Accurate Unbiased(UMAU) confidence intervals, shortest length confidence intervals, Asymptotic Confidence Intervals based on CAN estimators, Variance stabilizing transformations (VST), confidence interval based on VST, Asymptotic Confidence regions.

(12L+3T)

Unit 4: Likelihood ratio test (LRT) and its asymptotic distribution, Wald test, Rao's Score test, Pearson Chi-square test of goodness of fit, Bartlett's test for homogeneity of variances. Large sample tests based on VST. Consistent test, comparison of tests: asymptotic relative efficiency of tests (Pitman and Bahadur efficiency). Performance evaluation (based on simulation) of asymptotic tests and confidence intervals.

(12L+3T)

References:

1. Kale, B. K., &Muralidharan, K. (2015). *Parametric inference: An introduction*. Alpha Science International Limited.
2. Dudewicz, E. J. and Mishra, S. N. (1988). *Modern Mathematical Statistics*, John Wiley & Sons.
3. Ferguson, T. S. (2014). *Mathematical statistics: A decision theoretic approach*. Academic press.
4. Gibbons, J.D., & Chakraborti, S. (2010). *Nonparametric Statistical Inference* (5th ed.). Chapman and Hall/CRC.
5. Lehman, E. L. (1987). *Theory of testing of hypotheses*. Students Edition.
6. Randles, R. H., & Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics*. John Wiley.
7. Rohatgi, V. K., & Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons.
8. Zacks, S. (1971). *Theory of Statistical Inference*, John Wiley & Sons, New York.

MJ-MST23-202: LINEAR MODELS AND REGRESSION ANALYSIS

Course outcomes:

Upon successful completion of this course, a student will be able to:

i) Understand the concept of general linear model and associated inferential procedures.
ii) Understand and develop multiple linear regression models
iii) Identify the problems in developing multiple linear regression models and apply remedies.
iv) Understand generalized linear models and apply them for analyzing real data.

Unit 1: General linear model: definition, assumptions, concept of estimability, least squares estimation, BLUE, estimation space, error space, Gauss Markov theorem, variances and covariances of BLUEs, Distribution of quadratic forms for normal variables: related theorems(without proof), Tests of hypotheses in general linear models. Description of the ANOVA and linear regression models as the particular cases of the general linear model.

(12L+3T)

Unit 2: Multiple linear regression model, Least squares estimates (LSE) of parameters, Properties of LSE, Hypothesis testing, confidence and prediction intervals, General linear hypothesis testing, Model adequacy checking, Dummy variables and their use in regression analysis. Transformations to correct model inadequacies: VST and Box-Cox power transformation.

(12L+3T)

Unit 3: Multicollinearity: Consequences, detection and remedies, ridge regression. Autocorrelation: sources, consequences, detection (Durbin-Watson test) and remedies. Parameter estimation using Cochrane-Orcutt method. Variable Selection Procedures: R- square, adjusted R-square, Mallows’ Cp, forward, backward and stepwise selection methods, AIC, BIC. Robust Regression: need for robust regression, M-estimators, properties of robust estimators: breakdown and efficiency. Asymptotic distribution of M-estimator (Statement only).

(12L+3T)

Unit 4: Generalized linear models: concept of generalized linear model, Link function, ML estimation, large sample tests about parameters, goodness of fit, analysis of deviance. Residual analysis, types of residuals: raw, Pearson, deviance, Anscombe, quantile; residual plots. Variable selection: AIC and BIC. Logistic regression: logit, probit and cloglog models for dichotomous data, ML estimation, Odds ratio and its interpretation, hypothesis tests about model parameters. Hosmer-Lemeshow test, multilevel logistic regression, Logistic regression for Nominal response. Poisson regression.

(12L+3T)

References:

1. Birkes, D., & Dodge, Y. (2011). *Alternative methods of regression*. John Wiley & Sons.
2. Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.
3. Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. John Wiley & Sons.
4. Huber,P.J.andRonchetti,E.M (2011)*Robust Statistics*,Wiley,2ndEdition.
5. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Wasserman, W. (2004). *Applied linear regression models*. New York: McGraw-Hill/Irwin.
6. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. 5th Ed. John Wiley & Sons.
7. Seber,G.A.,Wild,C.J.(2003).*Non linear Regression*,Wiley.
8. Weisberg, S. (1985). *Applied Linear Regression*, John Wiley & Sons. New York.

MJ-MST23-203: STATISTICAL PROGRAMMING USING PYTHON

Course outcomes:

Upon successful completion of this course, a student will be able to:

i) Develop programs in Python.
ii) Perform data organization, data manipulation, statistical and mathematical computations, and data analysis using Python.

Unit 1: Introduction, installation, keywords, identifiers: variables, constants, literals; comments, Operators, statements and expressions, data types with methods: numbers, string, lists, tuple, dictionary, set; indexing and slicing of each data type, data type conversion, built-in functions, control statements and loops, list comprehensions, user defined functions, anonymous/lambda function, local and global variables, modules: math, stat, random;creating own modules.

(12L+3T)

Unit 2: Concept of library and its working, Data storage, manipulation, visualization and analysis using the libraries: *Numpy, Pandas, Scipy, statsmodels, Matplotlip, Seaborn, Regular Expressions (RegEx), Ski-kit learn.*

(12L+3T)

Reference:

1. Gowrishankar, S., &Veena, A. (2018). *Introduction to Python programming*. CRC Press.
2. Guttag, J. V. (2021). *Introduction to Computation and Programming Using Python: With Application to Computational Modeling and Understanding Data*. Mit Press.
3. Haslwanter, T. (2016). An Introduction to Statistics with Python. *With Applications in the Life Sciences. Switzerland: Springer International Publishing*.
4. Nelli, F. (2018). Python data analytics with Pandas, NumPy, and Matplotlib.
5. Unpingco, J. (2016). *Python for probability, statistics, and machine learning* (Vol. 1). Springer International Publishing.
6. VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. " O'Reilly Media, Inc.".
7. URLs:
    - *https://scikit-learn.org/stable/*
    - *https://numpy.org/*
    - *https://scipy.org/*
    - *https://www.statsmodels.org/stable/index.html*
    - *https://matplotlib.org/*
    - *https://pandas.pydata.org/*

GE-MST23-204: PROBABILITY THEORY

Course outcomes:

Upon successful completion of this course, the student will be able to:

i) Understand the concept of sets and events, field, measurable functions, measure space, random variable.
ii) Understand the concept of probability space, probability measure, distribution function, moments and expectation, convergence of random variables
iii) Understand the concept of independence of events and 0-1 laws.
iv) Understand and apply characteristic function, law of large numbers and central limit theorems.

Unit 1: Sets and classes of events: the event, algebra of sets, classes of events, limsup, liminf and limit of sequence of sets, Monotone Class, fields, sigma fields and properties. Minimal σ-field generated by a class of sets, Borel σ-field. Tail events and tail Sigma field. Measurable Functions: functions and inverse functions, measure, measurable space and measure space, finite measure, sigma finite measure. Random variable: Definition of random variable, Indicator function and its properties, Simple and elementary random variable, limits of random variables and properties.

(12L+3T)

Unit 2: Probability space: definition of probability, its properties, discrete probability space, general probability space, induced probability space, conditional probability measure, Dirac Measure, Counting measure, Lebesgue and Lebesgue - Steltjes measures, linear combinations of probability measures. Distribution Function and its properties. Expectations and moments: Definition and properties of expectation, conditional expectation, moments, moment generating functions. Convergence of random variables: Types of convergence, Cauchy convergence, Convergences in probability, Criterion for Convergences in probability, Slutsky's theorems, Almost sure convergence, Convergence in distribution, Convergence in r^{th} mean, Interrelations

(12L+3T)

Unit 3: Convergence Theorem for expectation: Monotone convergence theorem, Fatous Lemma, Dominated Convergence theorem. Independence: Independence of events, class of independent events, independence of classes, independence of random variables, expectation of product of independent random variables, equivalent definitions of independence, Zero-One laws: Kolmogorov 0-1 law, Borel 0-1 criterion, Borel-Cantelli Lemma

(12L+3T)

Unit 4: Characteristic function: definition and properties of characteristic function, inversion formula (without proof), characteristic function and moments. Laws of large numbers: Convergence of distribution functions, convergence of series of independent random variables, Kolmogorov inequalities and almost sure convergence, weak law of large numbers (without Proof) for iid and non-iid random variables, Strong law of large numbers (without Proof). Central limit Theorem(CLT) (without proof): Lindeberg-Levy, Liaponove's and Lindeberg-Feller forms and applications.

(12L+3T)

References:

1. Alan Karr. (1993): *Probability Theory*- Springer Verlag.
2. Athreya, K. B., &Lahiri, S. N. (2006). *Probability theory*. Hindustan Book Agency.
3. Bhat, B. R. (2007). *Modern probability theory*. New Age International.
4. Billingsley, P. (1986). *Probability and measure*.,(John Wiley and Sons: New York). Billingsley
5. Loeve, M. (1978). *Probability Theory* (Springer Verlag). Fourth edition.
6. Rohatgi, V. K., & Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons.

GE-MST23-204a: BAYESIAN INFERENCE

Course outcomes:

Upon successful completion of this course, the student will be able to:

i) Describe the role of the prior and posterior distribution in Bayesian inference about a parameter and provide various prior distributions
ii) Carry out the statistical inference (point estimate, testing of hypothesis and interval estimate) of the parameter using Bayesian approach
iii) Perform Bayesian analysis with various prior distribution
iv) Generate random samples from distributions for which direct sampling is difficult and use iterative methods for estimation

Unit 1: Basic elements of Statistical Decision Problem. Expected loss, decision rules (non-randomized and randomized). Overview of Classical and Bayesian Estimation. Advantage of Bayesian inference, Prior distribution, Posterior distribution, Subjective probability and its uses for determination of prior distribution. Importance of non-informative priors, improper priors, invariant priors. Conjugate priors, construction of conjugate families using sufficient statistics, hierarchical priors. Admissible and minimax rules and Bayes rules.

(12L + 3T)

Unit 2: Point estimation, Concept of Loss functions, Bayes estimation under symmetric loss functions, Bayes credible intervals, highest posterior density intervals, testing of hypotheses. Comparison with classical procedures. Predictive inference. One- and two-sample predictive problems.

(12L + 3 T)

Unit 3: Bayesian analysis with subjective prior robustness and sensitivity, classes of priors, conjugate class different methods of construction of objective priors: Jeffrey's prior, probability matching prior, conjugate priors and mixtures, posterior robustness: measures and techniques. Bayes factors large sample methods: Limit of posterior distribution, consistency of posterior distribution, asymptotic normality of posterior distribution.

(12L + 3 T)

Unit 4: Bayesian Computations: Analytic approximation, E-M Algorithm, Monte Carlo sampling, Markov Chain Monte Carlo Methods, Metropolis-Hastings Algorithm, Gibbs sampling, examples, convergence issues.

(12L + 3T)

References:

1. Bolstad, W. M. (2007). Introduction to Bayesian Statistics Second Edition. Amerika: A John Wiley & Sons.
2. Christensen, R., Johnson, W., Branscum, A., & Hanson, T. E. (2011). *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. CRC press.
3. Congdon, P. (2007). *Bayesian statistical modelling*. John Wiley & Sons.
4. Ghosh, J. K., Delampady, M., & Samanta, T. (2006). *An introduction to Bayesian analysis: theory and methods* (Vol. 725). New York: Springer.
5. Albert, J. (Ed.). (2009). *Bayesian computation with R*. New York, NY: Springer New York.
6. Rao. C.R. and Day. D. (2006). *Bayesian Thinking, Modeling & Computation, Handbook of Statistics*, Vol. 25. Elsevier

MJ-MSP23-206: PRACTICAL II

Course Outcomes:

Upon successful completion of this course, the student will be able to:

i) Test hypothesis about the parameters and provide interval estimates involved in random experiments based on random sample.

ii) Perform statistical analysis, such as estimation, hypothesis testing, and analysis of variance, under generalized linear models

iii) Fit linear regression model or Generalized Linear Regression Models to the data, perform diagnostic analysis and apply rectifying measures to overcome the problem of Multicollinearity, auto-correlation, outliers and non-linearity.

iv) Perform mathematical/statistical computations, statistical data analysis using built-in functions in Python and develop programs for various tasks.

Practical List:

1. MP, UMP, and UMPU Tests
2. Tests based on Hotelling's T^2 statistic
3. Confidence Intervals
4. Likelihood ratio tests
5. Performance evaluations of confidence intervals and tests through simulation.
6. Linear Estimation: Estimation and Hypothesis testing
7. Multiple linear regression
8. Variable selection in regression
9. Dealing with multicollinearity, autocorrelation and outliers
10. Logistic Regression
11. Poison regression
12. Monte Carlo Simulation of Regression Models
13. Python programming-I
14. Python programming-II
15. Python Programming-III
16. 16-20. Based on elective paper

Rayat Shikshan Sanstha's

SADGURU GADGE MAHARAJ COLLEGE, KARAD

(An Autonomous College)

SYLLABUS

For

M.Sc. Statistics

Part II (Semester III and IV)

As per NEP-2020

To be implemented from academic year 2024-25

SEMESTER III

MJ-MST23-301 STOCHASTIC PROCESSES

Course Outcomes: Students should to understand

1. To acquire more detailed knowledge about Markov processes with a discrete state space , including Markov chains, Poisson processes and birth & death processes.
2. To know about queuing systems, in addition to mastering the fundamental principals of simulation of stochastic processes.
3. To formulate simple stochastic process models in the time domain and provide qualitative & quantitative analysis of such models.

Unit 1: Definition of stochastic process, classification of stochastic processes according to state space and time domain, finite dimensional distributions. Examples of various stochastic processes. Definition of Markov chain. Examples of Markov chains, Formulation of Markov chain models, initial distribution, transition probability matrix, Chapman-Kolmogorov equations,calculation of n-step transition probabilities. Simulation of Markov Chain.
(12L +3T)

Unit 2: Classification of states, irreducible Markov chain, period of the state, random walk and gambler's ruin problem, first entrance theorem, first passage time distribution. Long-Run proportions and limiting probabilities, relation with mean recurrence time, stationary distribution.

(12L +3T)

Unit 3: Discrete state space continuous time Markov chain, Poisson process and related results. Birth and death processes and associated cases. Renewal and delayed renewal processes, related theorems, key renewal theorem (Without proof) and its application. Simulation of Poissonprocess and discrete state space Markov processes

(12L +3T)

Unit 4: Galton-Watson Binary Branching process. Generating functions and its properties,moments. Probability of ultimate extinction. Distribution of population size and association results. Simulation of branching process. Basic elements of Queuing model. Steady state probabilities and various average characteristics for the models: M/M/1, M/M/1 with balking, M/M/c and M/G/1. (12 L+ 3T)

References:

1. Bhat B. R. (2000). Stochastic Models: Analysis and Applications, (New Age International)
2. Cinlar E. (2013): Introduction to Stochastic Process. (Courier Corporation)
3. Feller W.(2008): An Introduction to Probability Theory and Its Applications. (Wiley)
4. Hoel P. G., Port S. C. and Stone C. J. (1987): Introduction to Stochastic Processes. (Waveland Press)
5. Karlin S. and Taylor H. M. (1968): A First Course in Stochastic Process. (Academic Press)
6. Medhi J. (2009): Stochastic Process, (New Age International Publications)
7. Ross S. (1996): Stochastic Processes. (Wiley)
8. Ross S. (2014): Introduction to Probability Models. (Academic Press)
9. Taylor H. M. and Karlin S. (2014): An Introduction to Stochastic Modeling (Academic Press)

MJ-MST23-302 DATA MINING

Course Outcomes: Students should to understand

1. To understand process of data understanding, cleaning.
2. To familiarize with various data mining functionalities and how it can be applied to various real-world problems.
3. To familiarize with various machine learning algorithms used in data mining.

Unit 1: Data understanding and data cleaning, concept of supervised and unsupervised learning. Problem of classification, classification techniques: k-nearest neighbor, decision tree, Naïve Bayesian, classification based on logistic regression, Bayesian belief Network.

(12L+3T)

Unit 2: Model evaluation and selection: Metrics for Evaluating Classifier Performance, Holdout Method and Random Subsampling, Cross-Validation, Bootstrap, Model Selection Using Statistical Tests of Significance, Comparing Classifiers Based on Cost–Benefit and ROC Curves. Techniques to Improve Classification Accuracy: Introduction to Ensemble Methods, Bagging, Boosting and AdaBoost, Random Forests, Improving Classification Accuracy of Class- Imbalanced Data. (12L+3T)

Unit 3: ANN and SVM: Artificial Neutral Network (ANN): Introduction to ANN, types of activation function, McCulloch-Pitts AN model, single layer network, multilayer feed forward network model, training methods, ANN & regression models. Support vector machine: Introduction to support vector machine, loss functions, soft margin, optimization hyperplane, support vector classification, support vector regression, linear programming support vector machine for classification and regression.

(12L+3T)

Unit 4: Unsupervised learning: Clustering: k-mediods, CLARA, DENCLUE, DBSCAN, Probabilistic model based clustering. Market Basket Analysis: Association rules and prediction, Apriori Algorithm, data attributes, applications to electronic commerce.

(12L+3T)

References:

1. Berson and Smith S.J. (1997) : Data warehousing, Data Mining, and OLAP, McGraw-Hill.
2. Breiman J.H Friedman, R.A. Olshen and stone C.J. (1984) : Classification and Regression Trees, Wadsworth and Brooks / Cole.
3. Han, J. and Kamber, M. and Pei, J. (2012) : Data Mining: Concepts and Techniques. MorganGaufmann.3rd Edition.
4. Mitchell T.M. (1997) : Machine Learning , McGraw-Hill.
5. Ripley B.D. (1996) : Pattern Recognition and Neural Networks. Cambridge University Press.
6. Vapnik V.N. The nature of Statistical learning theory, Springer.
7. Cristianini N. and Shawe-Taylor J. An Introduction to support vectormachines.
8. Data set source: http://www.ICS.uci.edu/~mlearn/MLRepository.html
9. Mehrika, K., Mohan, C., and Ranka (1997) Elements of Artificial neural networks. Penram international.
10. Hastie T, Tibshirani R, Friedmant J, (2009): The elements of statistical Learning, Springer.
11. Chattamvelli, R. (2015). Data mining methods. Alpha Science International.

MJ-MST23-303 MULTIVARIATE ANALYSIS

Course Outcomes: Students should to understand

1. To understand the conceptual framework for the different multivariate analysis.
2. To apply and interpret methods of dimensions reduction including principal component analysis.

Unit 1: Exploratory multivariate data analysis, sample mean vector, sample dispersion matrix, correlation matrix, graphical representation, means, variances, covariances, Partial and multiple correlation coefficients. Correlations of linear transformation. Hotelling's T^2 Statistic and its null distribution. Applications of T^2 statistics and its relationship with Mahalanobis D^2 statistic. Confidence region for the mean vector.

(12L+3T)

Unit 2: Discrimination and classification. Fisher's discriminant function and likelihood ratio procedure, minimum ECM rule, Rao's U statistics and its use in tests associated with discriminant function, classification with three populations. Cluster analysis, Hierarchical methods: Single, Complete, average linkage method and non-hierarchical clustering method-k- means clustering. Canonical correlation analysis, Introduction to principal component analysis and related results, Introduction to factor analysis and estimation.

(12L+3T)

References:

1. Kshirsagar A. M.(1972) : Multivariate Analysis. Marcel-Dekker.
2. Johnson, R.A. and Wichern . D.W (2002) : Applied multivariate Analysis. 5thEd.Prentice – Hall.
3. Anderson T. W. (1984) : An introduction to Multivariate statistical Analysis2nd Ed. John Wiely.
4. Morrison D.F. (1976) : Multivariate Statistical Methods McGraw-Hill.

GE-MST23-304 DESIGN AND ANALYSIS OF EXPERIMENT

Course Outcomes: Students should to understand

1. To understand General linear model, Guass Markov theorem, variances and covariance's of BLUEs.
2. To recognize one way classification, two way classifications without interaction and with interaction.
3. To apply this theory to the analysis of specific models in designing statistical experiments.

1. The concept of General linear model
2. The concept of factorial designs, confounding factorial design.
3. The statistical aspects of experimental design as a whole within the structure provided by general linear models.

Unit 1: Concept of design of experiments (DOE), applications of DOE; Basic principles of DOE; Analysis of completely randomized design using the fixed effect model and estimation of the model parameters; Contrasts, orthogonal contrasts, Scheffé's method for comparing contrasts; Comparing pairs of treatment means: controlling false discovery rate, Tukey's test, Fisher least significant difference method; Comparing treatment means with a control; Analyses of randomized complete block design, Latin square design, balanced incomplete block design using fixed effect models and estimation of the model parameters.

(12L + 3T)

Unit 2: Concepts of factorial designs, main effects, and interaction effects; The two-factor factorial design and its analysis using fixed effect model; The general factorial design; Analysis of replicated and unreplicated 2^k full factorial designs; Blocking and confounding in a 2^k factorial design; Construction and analysis of 2^{k-p} fractional factorial designs and their alias structures; Design resolution, resolution III and resolution IV designs; fold over designs; saturated designs.

(12L + 3T)

Unit 3: The 3^k full factorial design and its analysis using fixed effect model; Confounding in 3^k factorial designs; Construction and analysis of 3^{k-p} fractional factorial designs and their alias structures; Factorials with mixed levels: factors at two and three levels, factors at two and four levels; Design optimality criteria; Concept of random effects and mixed effects models, analysis of 2^k factorial designs using the random effect model, analysis of 2^k factorial designs using the mixed effect model, rules for expected mean squares, approximate F-tests.

(12L + 3T)

Unit 4: Response surface methodology: the method of steepest ascent, analysis of the response surface using first and second order model, characterizing the response surface, ridge systems, multiple responses, designs for fitting response surfaces: simplex design, central composite design (CCD), spherical CCD, Box–Behnken design; Robust parameter design: crossed array designs and their analyses, combined array designs and the response model approach; The concepts of nested and split-plot designs.

(12L + 3T)

References

1. Montgomery D.C. (2017): *Design and Analysis of Experiments*, 9th edition, John Wiley& Sons, Inc.
2. Phadke, M. S.(1989). *Quality Engineering using Robust Design*, Prentice-Hall.
3. Voss, D., Dean, A., and Dean, A.(1999). *Design and Analysis of Experiments*, Springerverlag Gmbh.
4. Wu, C. F., Hamada M. S.(2000). *Experiments: Planning, Analysis and ParameterDesign Optimization, 2nd edition*, John Wiley & Sons.

MJ-MSP23-306 PRACTICAL-III

Course outcomes:

Upon successful completion of this course, the student will be able to:

1. To explore increasingly large databases and to improve market segmentation.
2. To carry out derivations involving conditional probability distributions and conditional expectations.
3. To study performance and evaluation of various confidence intervals.
4. To apply multivariate methods in the framework of the multivariate analysis.

1. Realization of stochastic process.
2. Classification of t.p.m. and computation of n- step probability matrix.
3. Classification of states: Computations of absorption probabilities.
4. Stationary distribution and recurrence time.
5. Classification
6. Market Basket Analysis..
7. Artificial Neural Network.
8. Support Vector Machine
9. Application of Multivariate Normal Distribution.
10. Cluster Analysis
11. Principal Component Analysis and Factor Analysis

12 -15 Based on elective paper

MJ-MST23-401 Biostatistics

Course Outcomes: Students should to understand

1. To understand the protocol or study design and the concept of clinical trails.
2. To study the process of drug development.
3. To study the Ethics & regulatory perspective on clinical research trials activities.
4. To identify and classify different types of trail designs.

Unit 1: Introduction to clinical trials: Aim, need and ethics of clinical trials, Role of ethics in clinical trials, conduct of clinical trials, preclinical research, phase I-IV trials, multi-center trials, bias and random error in clinical studies, randomization; concept of blinding/masking in clinical trials, Data management: data definitions, case report forms, database design, data collection systems for good clinical practice.

(12L+3T)

Unit 2: Design of clinical trials: parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, factorial designs, objectives and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials, design, and monitoring of Phase III trials with sequential stopping.

(12L+3T)

Unit 3: Bioavailability, pharmacokinetics, and pharmaco-dynamics, two compartment model. Design of bio-equivalence trials, Decision rules for bioequivalence, Inference for 2x2 crossover design: Classical methods of interval hypothesis testing for bioequivalence, Bayesian methods, nonparametric methods. Reporting and analysis: analysis of categorical outcomes from Phase I - III trials, analysis of survival data from clinical trials.

(12L+3T)

Unit 4: Epidemiological studies: aims, case-control and cohort designs. Measures of disease occurrence and association, variation and bias, identifying non-causal association and confounding, communicating results of epidemiological studies, ethical issues in epidemiology. Causal Inference.

(12L+3T)

References:

1. C. Jennison and B. W. Turnbull (1999): Group Sequential Methods with Applications toClinical Trials, CRC Press.
2. Chow S.C. and Liu J.P. (2004). Design and Analysis of Clinical Trials. 2nd Ed. MarcelDekkar.
3. Chow S.C. and Liu J.P.(2009). Design and Analysis of Bioavailability and bioequivalence.3rd Ed. CRC Press.
4. Clayton, D. and Hills, M. (2013). Statistical methods in Epidemiology, OUP.
5. Daniel, W. W. and Cross, C. L. (2012). Biostatistics: A Foundation for Analysis in theHealth Sciences, 10th Edition, Wiley.
6. J. L. Fleiss (1989). The Design and Analysis of Clinical Experiments.Wiley and Sons.
7. L. M. Friedman, C. Furburg, D. L. Demets (1998). Fundamentals of Clinical Trials,Springer Verlag.
8. Marubeni .E. and Valsecchi M. G. (1994). Analyzing Survival Data from Clinical Trials andObservational Studies, Wiley.
9. S. Piantadosi (1997). Clinical Trials: A Methodologic Perspective, Wiley and Sons.

MJ-MST23-402 Time Series Analysis

Course Outcomes: Students should to understand

1. To understand how time series works and what factors affect a certain variables at different points in time.
2. To forecast the trend pattern exhibited by the given data by using various methods.
3. To run interprete time series models and regression models in time series.

Unit 1: Exploratory time series analysis, Exponential, Double exponential and Holt-Winter smoothing and forecasting, auto-covariance, auto-correlation functions and their properties and characterization (without proof), partial auto covariance function, auto-covariance generating function. First and second order stationary time series, white noise process, Linear Process, estimates of mean, auto-covariance, auto-correlation and partial auto-covariance functions.

(12 L + 3 T)

Unit 2: Wold representation of linear stationary processes, linear time series models: autoregressive(AR), moving average(MA), autoregressive moving average models (ARMA). causality and invertibility of ARMA processes, computation of π -weights and ψ- weights, computation of ACVF, ACF and PACF for AR(1), AR(2), MA(1), MA(2), ARMA(1,1) processes and general procedure for ARMA(p,q) process. The need for differencing a time series, autoregressive integrated moving average models(ARIMA).

(12 L + 3 T)

Unit 3: Estimation of ARMA models: Yule-Walker estimation for AR Processes, Maximum likelihood and least squares estimation for ARMA Processes, Residual analysis and diagnostic checking. Minimum mean squared error forecasting for ARMA and ARIMA models, updating forecasts. Introduction to SARIMA models, Spectral Representation of the ACVF, Spectral density of an ARMA process, its computation for simple models.

(12 L + 3 T)

Unit 4: Introduction to ARCH and GARCH models. Properties and estimation under ARCH(1) and GARCH(1,1) model. Vector time-series models: Covariance and Correlation Matrix functions, MA and AR representation of vector processes, Covariance matrix function of the vector AR(1) and MA(1) models.

(12L + 3T)

Reference:

1. W. S. Wei (2005) Time Series Analysis: Univariate and Multivariate Methods
2. Box, G.E.P and Jenkins G.M. (1970) Time Series Analysis, Forecasting & Control, Holden-Day.
3. Brockwell, P.J and Davis R.A. (1987) Time Series: Theory and Methods, Springer-
4. TsayR. S. Analysis of Financial Time Series, 3rd Ed. (Wil. Ser. in Prob. and Statistics)
5. Kendall, M.G. (1978) Time Series, Charler Graffin
6. Chatfield, C. (2004) The Analysis of Time Series - An Introduction, Sixth edition, Chapmanand Hall.

GE-MST23-403 a. Optimization Techniques

Course Outcomes: Students are able to

1. Understand basics and formulation of linear programming problems and appreciate their limitations; solve linear programming problems using graphical method.
2. Apply simplex method to solve real life problems.
3. Solve artificial variable technique, duality theory, revised simplex method, sensitivity analysis.
4. Understand the concept of Game theory and dynamic programming to solve their problems and understand their real life applications.

Unit 1: Convex Sets and Functions: Convex sets, supporting and separating hyperplanes, convex polyhedra and polytope, extreme points, convex functions. Linear programming problem (LPP): Definition and applications, methods of solving LPP: Graphical method, Simplex method, theorems related to the development of simplex algorithm, theorems related to a basic feasible solution, reduction of a feasible solution to a basic feasible solution, improvement of a basic feasible solution, existence of unbounded solution, optimality conditions and other related theorems (statements only), Examples. Artificial variable technique: Two phase method, Big M method, degeneracy.

(12L+3T)

Unit 2: Concept of Duality, related theorems, complementary slackness property and development of dual simplex algorithm. Sensitivity Analysis: Changes in the cost vector, requirement vector and non-basic activity vector; addition of new variables and addition of new constraints.

(12L+3T)

Unit 3: Integer Linear Programming Problem (ILPP): The concept of cutting plane, cutting plane method for all ILPP and mixed ILLP, Branch and Bound method. Quadratic programming: Kuhn-Tucker conditions, methods due to Beale, Wolfe.

(12L+3T)

Unit 4: Theory of games: two person zero sum games, minimax and maximin principles, Saddle point, mixed strategies; rules of dominance, solution of 2 x 2 game by algebraic method, Graphical method, Reduction of the game problem as LPP. Dynamic Programming: The Recursion Equation Approach, Computational Procedure, Characteristics of Dynamic Programming, Solution of L.P.P. by Dynamic Programming.

(12L+3T)

References:

i. Hadley G.(1969): Linear Programming, Addison Wesley
ii. Taha H. A. (1971): Operation Research: An Introduction, Macmillan N.Y.
iii. KantiSwaroop& Gupta M. M.(1985): Operations Research, Sultan Chand & Co. ltd.
iv. P.Gupta&D.S.Hira(2010): Operation Research, Sultan Chand & Co. ltd.
J. K. Sharma. (2003): Operation Research: Theory and Applications. Macmillan

GE-MST23-403 b. ACTUARIAL STATISTICS

Course Outcomes: Students are able to

1. To estimate outcomes in uncertain situations concerning return on investment, inflation rates and market volatility.
2. To demonstrate accurate and proficient use of probability theory techniques.
3. To asses risk in insurance and finance using statistical methods.
4. To calculate quantities such as premiums, reserves using actuarial techniques.

Unit 1: Introduction to Insurance Business, Concept of risk, types of risk, characteristics of insurable risk ,Risk models for Insurance: Individual and aggregate Risk models for short term, Distribution of aggregate claims, compound Poisson distribution and its applications. Survival function and Life tables: Survival function, Distribution function, Density functions and Force of mortality. Time-until death random variable and Curtate-future lifetime random variable. (8)

Unit 2: Life tables, Select and ultimate life tables. Assumptions for fractional ages and some analytical laws of mortality. Life Insurance: Principles of compound interest: Nominal and effective rates of interest and force of interest and discount, compound interest, Insurance payable at the moment of death and at the end of the year of death ,Whole life insurance, endowment insurance, term insurance, deferred insurance and varying benefit insurance..

(14)

Unit 3: Annuities: annuity certain, discrete annuity, monthly annuity, continuous annuity, deferred annuity, present values and accumulated values of these annuities, Continuous life annuity, discrete life annuity, such as whole life annuity, temporary life annuity, n-year certain and life annuity, life annuities. with mthly payments, Present value random variables for these annuity payments, their means andvariances, Actuarial present value of the annuity. (14)

Unit 4: Loss at issue random variable, various principles to decide net premiums for insurance products and annuity schemes defined in unit II and III, fully continuous premiums and fully discrete premiums, True monthly payment premiums. Extended equivalence principle to decide gross premiums, Concept of reserve, Fully continuous reserve, Fully discrete reserve. (12)

References:

1.Deshmukh S. R.,An Introduction to Actuarial Statistics, University Press, 2009
2.Robin Cunningham, Thomas N. Herzog, Richard L. Models for Quantifying Risk, 4th Edition,ACTEX Publications, 2011.
3. Dickson, David C. M., Hardy, Mary R. and Waters, Howard R., Actuarial Mathematics forlife contingent risks, International series on actuarial science, Cambridge 2009.
4. Narang, Uma, Insurance Industry in India: Features, Reforms and Outlook, New Century Publications

MJ-MSP23-405: PRACTICAL IV

Course Outcomes:

Upon successful completion of this course, the student will be able to:

i) To Apply simplex method to solve real life problems.

ii) To Understand the concept of Game theory and dynamic programming to solve their problems and understand their real life applications.

iii) To forecast the trend pattern exhibited by the given data by using various methods.

Practical List:

1. Pharmacokinetic Parameter
2. Confidence Interval
3. Non Parametric Test Based On Clinical Trial Dynamic Programming Problem
4. Analysis Of Categorical Outcome
5. Auto covariance And Autocorrelation
6. Casual And Invertible
7. Smoothing The Series
8. .Forecasting

9-12 Based on elective paper