DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD – 500 007.

M.Sc. (STATISTICS) I-SEMESTER

SCHEME OF INSTRUCTIONS AND EXAMINATION W.E.F. A.Y. 2023-2024 ONWARDS

Paper #	Paper Code	Paper Title	Credits	Instruction Hours per Week	Semester end Examination duration	Max. Marks in the Semester end Examination	Max. Marks in the Internal Assessments, Assignments, Seminars & Attendance*
THEORY PAPERS							
I	STS-101	Mathematical Analysis	3	3	3	50	40+10
II	STS-102	Linera Algebra & Linear Models	3	3	3	50	40+10
III	STS-103	Probability Theory	3	3	3	50	40+10
IV	STS-104	Distribution Theory	3	3	3	50	40+10
PRACTICAL PAPERS							
V	STS-105	Statistical Methods using Python Programming	2	4	2	40	10
VI	STS-106	Linear Algebra & Linear Models (Conventional & using R)	2	4	2	40	10
VII	STS-107	Distribution Theory ((Conventional & using	2	4	2	40	10
VIII	STS-108	Data Handling (using R)	2	4	2	40	10
Semester Total			20	12+16**	-	440	160

* 75% Attendance is mandatory as per norms; ** For practical with batch of 20 students.

Note: Attendance 10 Marks will be assigned based on proportion of attendance. Four different Internal Assessment tests each with 10 marks. The tests are preferably on multiple choice / fill in the blanks / short answer questions / Quiz / Report writing / Seminar / Assignments.

M.SC. (STATISTICS) I-SEMESTER

STS-101: PAPER-I: MATHEMATICAL ANALYSIS (MA)

Objectives & Course Outcomes:

1. For understanding the Mathematical Statistics the basic concepts of real analysis.
2. To understand the applications of Limits, continuity, Convergences, R-S Integrations etc

UNIT-I

Metric spaces - Compact sets - Perfect sets - Connected sets. Limits of functions - Continuous functions - Continuity and compactness, Continuity and connectedness, Discontinuities - Monotonic functions, Differentiation.

UNIT–II

Riemann-Steiltjes (R-S) Integral and its linear properties. Integration by parts, Euler's summation, Riemann's condition. Integrators of Bounded variations. Statements of necessary and sufficient conditions of Riemann - Steiltjes integral. Differentiation under the integral sign. Interchanging the order of integration.

UNIT- III

Sequences and Series of Functions: Uniform convergence - Uniform convergence and continuity - Uniform convergence and integration - Uniform convergence and differentiation – The Stone-Weierstrass theorem.

REFERENCES

1. Walter Rudin: Principles of Mathematical Analysis, McGraw-Hill International 3rd Edition. (Unit-I: pp 30-46 & pp 83-102) (Unit-II: pp 120-133 & 135-142) (Unit-III: pp 143-154, 159-161, 165-171 & 220-222).
2. H.L. Royden: Real Analysis, PHI 3rd edition
3. Apostol, T.M. (1985): Mathematical Analysis, Narosa, Indian Ed.
4. Malik, S.C. (1984): Mathematical Analysis, Wiley – Eastern.
5. Mathematical Analysis Vol - I by D J H Garling.

M.SC. (STATISTICS) I-SEMESTER

STS-102: PAPER-II: LINEAR ALGEBRA & LINEAR MODELS (LA & LM)

Objectives & Course Outcomes:

1. To find the solution to the given set of equations.
2. To transform the given matrix into another form without changing its characteristics.
3. Knowing the real time applications of the matrix theory.
4. Estimation of parameters using GLM.
5. Computing Best linear unbiased estimator

UNIT – I

Linear Algebra: Vector Spaces with an inner product, Gram –Schmidt orthogonalization process. Ortho-normal basis and orthogonal projection of a vector. Real time applications of orthogonalization in various domains. Solution of matrix equations. Sufficient conditions for the existence of homogeneous and non – homogeneous linear equations. Moore Penrose and generalized inverses and their properties. Real time applications of solving set of equations in various domains.

UNIT–II

Characteristic roots and vectors, Caley–Hamilton theorem algebraic and geometric multiplicity of a characteristic root and spectral decomposition of a real symmetric matrix. Real time applications of characteristic roots and vectors in various domains. Real quadratic forms, reduction and classification of quadratic forms, Index and signature. Simultaneous reduction of two quadratic forms, Extreme of a quadratic form. Matrix Inequalities: Cauchy- Schwartz and Hadamard Inequalities.

UNIT – III

Linear Models: General Linear Model (GLM) and its formulation through examples. Estimability of a linear parametric function. Gauss-Markov linear model, BLUE for linear functions of parameters, relationship between BLUE's and linear Zero-functions. Gauss-Markov theorem, Aitkin's generalized least squares, Concept of Multi-collinearity. Importance and applications of GLM s.

REFERENCES

1. Graybill, F.A. (1983): Matrices with applications in Statistics, 2nd ed., Wards worth.
2. Searle, S.R.(1982) : Matrix Algebra useful for Statistics, John Wiley & Sons.
3. Searles S.R.(1971):Linear statistical Models.
4. Rao, C.R. and Mithra, S.K.(1971) : Generalized inverse of matrices and its applications, John Wiley & Sons.
5. Rao, A.R. and Bhimasankaram, P. (1992): Linear Algebra, Tata McGraw Hill Publishing Co. Ltd.

M.SC. (STATISTICS) I-SEMESTER

STAS-103: PAPER-III: PROBABILITY THEORY (PT)

Objectives & Course Outcomes:

1. To find the probability based on the conditions that are specified.
2. To obtain the distribution function of the random variable based on its probability function and vice-versa.
3. To derive characteristic function from the density and vice-versa and identifying the characteristic function.
4. To obtain the probability bounds or moment bounds for the given random variables.
5. To study convergence properties of the sequence random variables based on its probability laws.

(**Pre-requisite for understanding**: Probability concepts: Classical, statistical and axiomatic definitions, joint, marginal and conditional probabilities, Compound, Addition and Bayes theorems and problems on probability).

UNIT – I

Classes of sets, fields, sigma-fields, minimal sigma-fields, Borel sigma-fields in R, Measure, Properties of a Measure, measurable function, Statements and applications of Caratheodory extension theorem, Monotone and Dominated Convergence theorems and Foatou's lemma. Probability as a measure, Random Variables, distribution function and its properties and their applications. Mathematical Expectation and Expectations of functions of random variables and their applications. Conditional expectation and conditional variance and their applications. Characteristic function and its properties, Uniqueness, Inversion and Continuity theorems and their application problems. Identification functions which are/ not be Characteristic functions.

UNIT – II

Probability and Moment inequalities**:** Chebychev's, Markov, Cauchy-Schwartz, Holder, Minkowsky, Liapunov and Jensen Inequalities. Interrelationships among the inequalities and their applications and simple problems on these inequalities. Sequence of random variables: Borel-Cantelli Lemma; Borel 0-1 law. Statement of Glivenko-Cantelli lemma. Convergence of sequence of random variables: law, probability, almost sure and quadratic mean; their implications, counter implications, Slutzky's theorem. Applications of various modes of convergences and their related problems.

UNIT – III

Weak and Strong Law of Large numbers (WLLN): Bernoulli, Chebychev's and Khintchine's WLLNs. Kolmogorov inequality. Borel's SLLNs. Kolmogorov's SLLNs for independent random variables and i.i.d. random variables, Applications of LLN and their related problems. Central Limit Theorems: Demoviere-Laplace form of CLT, Levy-Lindeberg form of CLT, Liapunov's form of CLT and Lindberg-Feller form of CLT and their application related problems.

REFERENCES

1. Basu A.K. (2012): Measure Theory and Probability, PHI, 2nd edition.
2. Ross, S.M (2004): Introduction to Probability Models, 8th Edition, Academic Press
3. Bhat, B.R. (1985): Modern Probability Theory, Wiley Eastern.
4. Rohatgi, V.K. (1993): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern.

M.SC. (STATISTICS) I-SEMESTER

STAS-104: PAPER-IV: DISTRIBUTION THEORY (DT)

Objectives and Course Outcomes:

1. To derive any property for any distribution that is specified.
2. To obtain the distribution to the given transformed random variables.
3. To obtain sampling distribution to the given statistic.
4. To derive the point estimator for the given parameter in the distribution function.
5. To derive the properties like consistency, unbiasedness, efficiency, sufficiency, MVB, completeness, etc. for the estimator.

(**Pre-requisite**: Basic univariate probability distributions: Discrete Uniform, Bernoulli, Binomial, Poisson, Negative Binomial, Geometric, Hyper-Geometric, continuous uniform, Normal, Exponential, Gamma (one & two parameters), Beta First and second kinds and Cauchy. Univariate and bivariate random variables transformations).

UNIT – I

Definitions and derivations of properties related to Lognormal, Weibull, Pareto, Laplace distributions and their applications and related problems. Truncated distributions (Binomial, Poisson, normal distributions). Mixture Distributions and examples.

UNIT – II

Exponential family of distributions and Power series family of distributions and their means and variances (Binomial, Poisson, Geometric). Compound distributions of Binomial-Poisson, Poisson-Gamma, their means and variances. Functions of random variables and their distributions using Jacobian of transformations and problem on transformations, Distributions of Quadratic forms under normality and its applications.

UNIT – III

Derivations of density functions of Sampling Distributions of central and non-central t, F and χ^2 and their properties (for non-central only statements), distribution of Sample mean and variance, independence of $\overline{X}$ and S^2. Order statistics: Joint and Marginal distributions of order statistics. Distributions of sample range, Problems on computing the distribution of order statistics. Applications of order statistics.

REFERENCES

1. Bhuyan K.C. (2010): Probability distribution theory and Statistical Inference, New Central Book agency (P) Ltd.
2. Parimal Mukhopadhya (2002): Mathematical Statistics, Books & Allied Ltd.
3. Johnson, S. and Kotz (1972): Distribution in Statistics, Vol. I, II and III.
4. Johnson R.A. & Wichern: Applied Multivariate analysis.
5. Kshirasagar, A.M. (1972) : Multivariate Analysis, Marcel Decker

M.SC. (STATISTICS) I-SEMESTER

STS-105: PAPER-V: STATISTICAL METHODS USING PYTHON PROGRAMMING

PRACTICAL-I

Objectives & Outcomes:

1. Use various data types, loop statements, OOPs concepts, Exemptions, string operations etc for a specified problem
2. Design, implement, debug a given problem using Python
3. Execute the programs using derived and user defined data types.
4. Implement programs using modular approach and file I/O
5. Writing Python code for any statistical methods for the given data data set.

(THEORETICAL CONCEPTS)

Weeks 1-5: Introduction to Python Programming, Input, Processing and Output, Displaying Output with the Print Function, Comments, Variables, Reading Input from the Keyboard, Performing Calculations Operators. Type conversions, Expressions, More about Data Output. Decision Structures and Boolean Logic: if, if-else, if-elif-else Statements, Nested Decision Structures, Comparing Strings, Logical Operators, Boolean Variables. Repetition Structures: recursion and non-recursion, while loop, for loop, Calculating a Running Total, Input Validation Loops, Nested Loops. python-syntax, statements, functions, Built-in-functions and Methods, Modules in python, Exception Handling. Functions: Defining and Calling a Void Function, designing a Program to Use Functions, Local Variables, Passing Arguments to Functions, Global Variables and Global Constants, Value-Returning Functions, Generating Random Numbers, Writing Our Own Value-Returning Functions, The math Module, Storing Functions in Modules. File and Exceptions: Introduction to File Input and Output, Using Loops to Process Files, Processing Records, Exceptions. Finding Items in Lists with in-Operator, List Methods and Useful Built-in Functions, Copying Lists, Processing Lists, Two-Dimensional Lists, Tuples. Strings: Basic String Operations, String Slicing, Testing, Searching, and Manipulating Strings.

LIST OF PRACTICALS (Not using Python Packages)

(Programs must be in a position to write with all possibilities like usage of functions, Loops, OOP concepts, methods, built in functions etc. wherever it is possible)

Week-6:	Program to find the sum and product of two matrices. Program to find the Determinant and Inverse of the given matrix
Week-7:	Program to sort the given set of numbers using bubble sort, Quicksort, Merge sort, insertion sort. Program for linear search, binary search.
Week 8:	Program to find the Median, Mode for the given of array of elements. Program for preparation of frequency tables, Computation of mean, median, mode, variance and standard deviation to the given data set.
Week 9:	Program to compute first four Central & Non-central moments, Skewness and Kurtosis to the given data set.
Week-10:	Program to generate random numbers from Uniform, Binomial, Poisson, Normal and Exponential distributions using algorithms.
Week-11:	Program to Fit Binomial, Poisson & Negative Binomial distributions for the given data set and testing their goodness of fit and drawing the curve plots.
Week-12:	Program to Fit Normal, Exponential & Cauchy distributions for given data set.
Week-13:	Program for finding Correlation and regression lines for the given data set.
Week-14:	Program for testing means, variances, correlations.
Week-15:	Program for carryout the analysis of variance for one way and two way.

Note: Practical Exam question paper will have 15 Marks (5Q x 3M) weightage on theory concepts and 25 (2Q out of 4 X 8 M + 9 M Execution) on its programs writing and execution. Practical Record should contain all practical's with their implementation and is Mandatory and carries 5 marks and Assessment test / Viva-Voce- 5 Marks.

M.SC. (STATISTICS) I-SEMESTER

STS-106: PAPER-VI: LINEAR ALGEBRA & LINEAR MODELS
PRACTICAL-II

Objectives & Outcomes:

1. Knowing the manual procedures and also their implementation using R
2. Finding the inverse of a matrix in various methods.
3. Applying any transformations on matrices
4. Applying the matrix operations on the given data sets (Determinant, eigen values, eigen vectors, transformations etc).
5. Summarization of properties of the data sets based on matrix operations.

(LIST OF PRACTICALS: CONVENSIONAL & USING R)

Week-1:	Inverse of a matrix by Partition method
Week-2:	Solutions of linear equations by sweep-out method
Week-3:	Solutions of linear equations by Doolittle Method
Week-4:	Computation of Moore-Penrose inverse by Penrose method
Week-5:	Computation of generalized inverse of a matrix.
Week-6:	Formation of characteristic equation by using traces of successive powers
Week-7:	Spectral decomposition of a square matrix of third order
Week-8:	Simultaneous reduction of a pair of quadratic forms to diagonal and canonical forms.
Week-9:	Finding orthonormal basis by Gram – Schmidt process.
Week-10:	Computation of variance-covariance matrix for data set and study of its characteristics.
Week-11	Fitting of a simple linear regression model, Testing its lack of fit, and computing its R^2, Adj R^2, Pure error and Confidence interval for regression coefficient
Week-12	Fitting of a Multiple Linear regression model, Testing its lack of fit, and computing its R^2, Adj R^2, Pure error and Confidence interval for regression coefficient.
Week-13	Computation of Simple, partial and Multiple correlation Coefficients
Week-14	Testing Multi-Collinearity

Note: Practical Record should contain all practical's with their implementation and is Mandatory and it carries 5 Marks and Assessment test / Viva-Voce- 5 Marks. The Semester end practical exam contains two sections: Section-A: Conventional & Section-B Using R. (Answer ant two out of four choosing at least one from each section)

M.SC. (STATISTICS) I-SEMESTER

STS-107: PAPER-VII: DISTRIBUTION THEORY
PRACTICAL-III (CONVENSIONAL & USING R)

Objectives & Outcomes:

1. Knowing the manual procedures and also their implementation using R
2. Generation of random samples from any distribution
3. Identifying an appropriate probability distribution to the given data.
4. Fitting and testing the probability distribution.
5. Drawing the probability distribution curves and stating its nature of the distributional curve properties for the given data sets.

LIST OF PRACTICALS:

Week-1: Generation of random samples from Uniform distribution.

Week-2: Generation of random samples from the Binomial, Poisson, Geometric, Negative Binomial distributions.

Week-3: Generation of random samples from the Normal, Exponential, Gamma, Beta, Cauchy distributions.

Week-4: Fitting an appropriate discrete distribution to the given data sets

Week-5: Fitting an appropriate continuous distribution to the given data sets (Uniform, Normal, Exponential)

Week-6: Testing its Goodness of fit of Cauchy distribution to the given data set

Week-7: Fitting of Gamma distribution with two parameters to the given data set

Week-8: Fitting of Lognormal Distribution with two parameters to the given data set

Week-9: Fitting of Weibull Distribution with two parameters to the given data set

Week-10: Fitting of Pareto distribution with two parameters to the given data set

Note: Practical Record should contain all practical's with their implementation and is Mandatory and it carries 5 Marks and Assessment test / viva voce is 5 Marks. The Semester end practical exam contains two sections: Section-A: Conventional & Section-B Using R. (Answer ant two out of four choosing at least one from each section).

M.SC. (STATISTICS) I-SEMESTER

STS-108: PAPER-VIII: DATA HANDLING

PRACTICAL-IV

Data sets of Kaggle.com can be used for practice. For example few of the them are: Iris Dataset**;** flights.csv Dataset; Sustainable Development Data; Credit Card Fraud Detection; Employee dataset; Heart Attack Analysis & Prediction Dataset; Dataset for Facial recognition; Covid_w/wo_Pneumonia Chest Xray Dataset; Groceries dataset; Financial Fraud and Non-Fraud News Classification; IBM Transactions for Anti Money Laundering

Data Handling with R:

1. Understanding data with Data types, Measurement of scales, descriptive statistics and data pre-processing steps.
2. Data transformations (Standardize, Normalize, converting data from one scale to other scales).
3. Data Visualization: Drawing One dimensional diagram (Pictogram, Pie Chart, Bar Chart), two-dimensional diagrams (Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot), other diagrammatical / graphical representations like, Gantt Chart, Heat Map, Box-Whisker Plot, Area Chart, Correlation Matrices.
4. Parametric tests (z-, χ^2, t-, F-tests, ANOVA), Correlation & Regression etc.
5. Non-Parametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, Run test).
6. Applying the modelling process, Model evolution, over fitting, under fitting, cross validation concepts, (train/test, K fold and leave out one approaches),
7. Evaluation of Model Performance for classification techniques for qualitative and Quantitative data.
8. Data interpretation and Report writing.

Note: Practical Record should contain all practicals with their implementation and is Mandatory and it carries 5 Marks and Assessment test / viva voce is 5 Marks. The Semester end practical exam contain answer any two out of four questions with their implantations using R.

DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD – 500 007

M.Sc. (STATISTICS) II-SEMESTER

SCHEME OF INSTRUCTIONS AND EXAMINATION W.E.F. A.Y. 2023-24 ONWARDS

Paper #	Paper Code	Paper Title	Credits	Instruction Hours per Week	Semester end Examination duration	Max. Marks in the Semester end Examination	Max. Marks in the Internal Assessments, Assignments, Seminars & Attendance*
THEORY PAPERS							
I	STS-201	Estimation Theory	3	3	3	50	40+10
II	STS-202	Multivariate Analysis	3	3	3	50	40+10
III	STS-203	Design and Analysis of Experiments	3	3	3	50	40+10
IV	STS-204	Sampling Theory	3	3	3	50	40+10
PRACTICAL PAPERS							
V	STS-205	Estimation Theory & Multivariate Analysis	2	4	2	40	10
VI	STS-206	Design and analysis of Experiments and Sampling Theory	2	4	2	40	10
VII	STS-207	Statistical Analysis using SPSS	2	4	2	40	10
VIII	STS-208	Data Science using Python	2	4	2	40	10
Semester Total			20	12+16 **	-	480	120

* 75% Attendance is mandatory as per norms; ** For practical with batch of 20 students

M.SC. (STATISTICS) II-SEMESTER

STS-201: PAPER I: ESTIMATION THEORY (ET)

Objectives and Course Outcomes:

1. To derive any estimation to the parameter of any probability function using ML and moments methods if exists.
2. To examine the properties of the estimator like consistency, unbiasedness, efficiency, sufficiency, completeness, minimum variance bound, CAN, BAN, etc.
3. Estimation of density function based on the sample in nonparametric approach.
4. Estimation of parameters based on the resampling techniques.

(Pre-requisite for understanding: Concepts of Point estimation, Interval estimation, method of maximum likelihood and moment estimations and related simple problems, Criteria for good estimation, Consistency, unbiasedness, efficiency and sufficiency and its related problems and simple problems on interval estimation using pivot method).

Unit-I

Estimation: Point and Interval estimation, Simple problems related to criterion for good estimator, Minimum Variance Unbiased Estimator, UMVU estimation, Fisher's information, Cramer-Rao inequality, Rao–Blackwell theorem, Completeness, Lehmann – Scheff's theorem and their applications and its related problems, and Bhattacharya bounds. MLE and its properties (statements only). Consistency and asymptotic normality of the consistent solutions of likelihood equations. Definition of CAN and BAN estimators and their properties, related examples.

Unit-II

Estimation of bias and standard deviation of point estimators of Jackknife and Bootstrap methods with examples. Concept of U statistics, Kernal and examples. Statement of Asymptotic distributions of U – statistics. Interval estimation: confidence level CI using pivots and shortest length CI. Confidence intervals for the parameters for Normal, Exponential, Binomial and Poisson Distributions. Confidence Intervals for quintiles. Concept of tolerance limits and examples.

Unit-III

Concepts of loss, risk and decision functions, admissible and optimal decision functions, estimation and testing viewed as decision problems, apriori, aposteriori distributions, conjugate families, Baye's and minimax decision functions with applications to estimation with quadratic loss. Concepts of nonparametric estimation: Density estimates, survey of existing methods. Rosenblatt's naïve density estimator, its bias and variance. Consistency of Kernel density estimators and its MSE.

REFERENCES

1. Goon, Gupta and Das Gupta : Outlines of Statistics, Vol. 2, World Press, Calcutta.
2. Rohatgi, V.K.: An introduction to Probability theory and mathematical statistics, W/E.
3. Rao, C.R.: Linear Statistical Inference and its applications, John Wiley
4. Gray and Schucany : Generalized Jackknife; Marcel Decker
5. Efron B. and Robert J. Tibshirani: An Introduction to the Bootstrap, Chapmen and Hall.
6. Lehman, E.L. (1983): Theory of point estimation, John Wiley

M.SC. (STATISTICS) II-SEMESTER

STS-202: PAPER-II: MULTIVARIATE ANALYSIS (MVA)

Objectives and Course Outcomes:

1. To understand the distribution of Multivariate data.
2. To understand the multivariate statistical tools.
3. To identify and applying of multivariate techniques for data analysis.

UNIT – I

Multinomial distribution Multivariate normal distribution, marginal, conditional distributions. Independence of multivariate vectors. Random sampling from a multivariate normal distribution. Maximum likelihood estimators of parameters. Distribution of sample mean vector. Independence of sample mean vector and variance-covariance matrix. Wishart matrix, Wishart distribution and its properties. Distribution of sample generalized variance. Null distribution of simple correlation coefficients. Null distribution of partial and multiple correlation coefficients. Distribution of sample regression coefficients. Application in testing and interval estimation.

UNIT – III

Null distribution of Hotelling's T^2 statistic. Application in tests on mean vector for one and more multivariate normal populations and also on equality of the components of a mean vector in a multivariate normal population. Mahalanobi's D^2 statistic. Wilk's Λ - criterion and statement of its their properties with simple applications. Linear Discriminant Analysis: Classification and discrimination procedures for discrimination between two multivariate normal populations – sample discriminant function, tests associated with discriminant functions, probabilities of misclassification and their estimation, classification into two multivariate normal populations with equal covariance matrices.

UNIT – III

Principal component analysis and its properties and applications. canonical variables and canonical correlations: definition, use, estimation and computation. Cluster analysis: Definitions, Agglomerative hierarchical clustering methods, Single complete and average linkages, K-means, KNN clustering. Multi-dimensional scaling methods (metric & non metric). Introduction to Factor analysis, orthogonal factor model. Path analysis, Correspondence analysis, conjoint analysis.

REFERENCES

1. Johnson, R.A.Wichern: Applied Multivariate Analysis, PHI
2. Anderson, T.W. (1983) : An Introduction to multivariate statistical analysis, 2nd Edition, Wiley.
3. Kshirasagar, A.M. (1972) : Multivariate Analysis, Marcel Decker.
4. Morrison, D.F. (1976): Multivariate Statistical Methods, 2nd Edition, McGraw Hill

M.SC. (STATISTICS) II-SEMESTER

STS-203: PAPER-III - DESIGN AND ANALYSIS OF EXPERIMENTS (DAE)

Objectives and Course Outcomes:

1. Analysis of the experimental data using full factorials, with partial and total confounding.
2. Analysis of the experimental data using one way and two classifications
3. To estimate the parameters of population and estimating variances.

Pre requisite: Concept of analysis of Variance and ANOVA for one-way and two-way classifications with one observation per cell, expectation of various sums of squares, Statistical analysis, Analysis of Completely randomized, Randomized Block and Latin Square Designs including estimation missing observations and efficiencies)

Unit-I

Analysis of variance for m-observations, n_{ij}-observations per cell. Multiple Comparison tests: Fishers Least Significance Difference (LSD) and Duncan's Multiple Range (DMR) tests. Analysis of Covariance for One-way and Two-way classifications. Factorial experiments: Estimation of Main effects, interaction effects and analysis of 2^k factorial experiments in general and with particular reference to k = 2, 3 and 4 and 3^2 factorial experiment.

Unit-II

Total and Partial Confounding in case of 2^3, 2^4 and 3^2 factorial designs. Concept of balanced partial confounding. Fractional replications of factorial designs: One half replications of 2^3 and 2^4 factorial designs, one-quarter replications of 2^5 and 2^6 factorial designs. Resolution of a design. Split – Plot design. Balanced incomplete block design (BIBD) – parametric relations, intra-block analysis, recovery of inter-block information. Construction of BIBD's. through MOLS.

Unit-III

Partially balanced incomplete block design with two associate classes PBIBD (2) – Parametric relations, intra block analysis. Simple lattice design and Youden-square design. Concept of Response surface methodology (RSM), Response surface designs. Design for fitting first – order and second – order models. Variance of estimated response. Second order rotatable designs (SORD), Central composite designs (CCD), Rotatability of CCD.

REFERENCES

1. Montogomery,D.C.:Design and Analysis of Experiments, John Wiley
2. Das M N and Giri, N.C: Design and Analysis of Experiments
3. Kempthrone: Design and Analysis of Experiments.
4. Cochran and Cox: Experimental Designs.

M.SC. (STATISTICS) II-SEMESTER

STS-204: PAPER-IV: SAMPLING TECHNIQUES (ST)

Objectives and Course Outcomes:

1. Usage of SRS, Stratified, systematic , Cluster and two stage sampling methods.
2. To estimate the parameters of population and estimating variances.

(Pre requisite: basic terminology, Need & Principal steps in sample surveys , census versus sample surveys, sampling and non- sampling errors, sampling methods. SRSWR, SRSWOR, stratified and systematic sampling methods, Estimates of their population mean, variances etc.)

Unit-I

Unequal probability Sampling: ppswr/wor methods (including Lahiri's scheme) and related estimators of a finite population mean. Horowitz – Thompson, Hansen – Horowitz and Yates and Grundy estimators for population mean/total and their variances. Ratio Method Estimation: Concept of ratio estimators, Ratio estimators in SRS, their bias, variance/MSE. Ratio estimator in Stratified random sampling – Separate and combined estimators, their variances/MSE.

Unit-II

Regression method of estimation: Concept, Regression estimators in SRS with pre – assigned value of regression coefficient (Difference Estimator) and estimated value of regression coefficient, their bias, variance/MSE, Regression estimators in Stratified Random sampling – Separate and combined regression estimators, their variance/ MSE. Cluster Sampling: Cluster sampling with clusters of equal sizes, estimator of mean per unit, its variance in terms of intracluster correlation, and determination of optimum sample and cluster sizes for a given cost. Cluster sampling with clusters of unequal sizes, estimator - population mean its variance/MSE.

Unit-III

Sub sampling (Two – Stage only): Equal first stage units – Estimator of population mean, variance/MSE, estimator of variance. Determination of optimal sample size for a given cost. Unequal first stage units – estimator of the population mean and its variance/MSE. Non – Sampling errors: Sources and treatment of non-sampling errors. Non – sampling bias and variance. Randomized Response Techniques (for dichotomous populations only): Warner's model, unrelated question model. Small area estimation : Preliminaries, Concepts of Direct Estimators, Synthetic estimators and Composite estimators.

REFERENCES

1. Parimal Mukhopadhyay (1998) : Theory and methods of Survey sampling, Prentice – Hall of India, New Delhi.
2. Murthy, M.N. (1967): Sampling Theory and methods, Statistical Publishing Society, Calcutta.
3. Des Raj (1976) : Sampling Theory, Tata McGraw Hill, New Delhi.
4. Sukhatme etal (1984) : Sampling Survey methods and its applications, Indian society of Agricultural Statistics.
5. Cochran, W.C. (1977) : Sampling Techniques, Third Edition, Wiley Eastern.

M.SC. (STATISTICS) II-SEMESTER

STS-205: PAPER-V: ESTIMATION THEORY AND SAMPLING THEORY

PRACTICAL-I (CONVENTIONAL)

SECTION-A: LIST OF PRACTICALS ON ESTIMATION THEORY

Week-1: Computation of Jackknife estimates

Week-2: Computation of Boot-strap estimates

Week-3: MLE by Scoring method for Cauchy population

Week-4: Confidence limits for parameters of normal population

Week-5: Large sample confidence limits in case of Binomial, Poisson, Exponential distributions

SECTION-B: LIST OF PRACTICALS ON MULTIVARIATE ANALYSIS

Week-1: MLE of Mean vector and variance covariance Matrix based on the sample drawn from p- Normal population.
Writing the density function based on Mean vector and covariance matrix and identification of parameters from the p-variate normal density.

Week-2: Hotelling's T^2 for test the mean vector based on single sample
Mahalanobi's D^2 for test the mean vector based on single sample

Week-3: Hotelling's T^2 for testing equality of the mean vectors based on two samples
Mahalanobi's D^2 for testing equality of the mean vectors based on two samples.

Week-4: Computation of Principal Components.

Week-5: Classification between two normal populations by discriminant analysis using Maximum likely hood ratio approach and Bayesian mis classification.

Week-6: Cluster analysis using Single, Complete and Average linkages.

Week-7: Computation of Canonical variables and correlation.

Week-8: Computation of Orthogonal Factor Model

Week-9: Computation of Path coefficients and drawing Path diagram

Week-10: Computation of Multidimensional Scaling

Note: Practical Record should contain all practical's with their implementation and is Mandatory and it carries 5 Marks and Assessment test / Viva-Voce is 5 Marks. The Semester end practical exam contains two sections: Section-A: & Section-B.

M.SC. (STATISTICS) II-SEMESTER

STS-206: PAPER VI: DESIGNS & ANALYISIS OF EXPERIMENTS AND MULTIVARIATE ANALYSIS

PRACTICAL-II (CONVENTIONAL)

SECTION-A: LIST OF PRACTICALS ON DESIGN AND ANALYSIS OF EXPERIMENTS

Week-1:	Analysis of Variance for two-way classification m- observations per cell
	Analysis of Variance for two-way classification n_{ij} -observations per cell
Week-2:	Analysis of Covariance for one-way classification
	Analysis of Covariance for two-way classification
Week-3:	Analysis of Variance for 2^3, 2^4 factorial experiments
	Analysis of Variance for 3^2 factorial experiments.
Week-4:	Identification of Confounded terms in 2^3, 2^4 and 3^2 factorial experiments.
	Construction of design with a specified effect is confounded.
Week-5:	Analysis of Variance for Total confounding of $2^{3,}$ 2^4 designs
	Analysis of Variance for Partial confounding of $2^{3,}$ 2^4 designs.
Week-6:	Analysis of Variance for one–half fraction of 2^4 designs and
	Analysis of Variance for one–quarter fraction of 2^5 designs.
Week-7:	Analysis of variance for Split-Plot design.
Week-8:	Analysis of Balanced Incomplete Block Design
Week-9:	Analysis of Youden Square Design
Week 10:	Analysis of Partially Balanced Incomplete Block Design

SECTION-B: LIST OF PRACTICALS ON SAMPLING THEORY

Week-1:	PPS sampling with and without replacements.
Week-2:	Ratio estimators in SRS, comparison with SRS
Week-3:	Separate and combined ratio estimators, Comparison.
Week-4:	Regression estimators in SRS, Comparison with SRS and Ratio estimators
Week-5:	Separate and combined Regression estimators, Comparison.
Week-6:	Cluster sampling with equal cluster sizes.
Week-7:	Sub sampling (Two–stage sampling) with equal first stage units

Note: Practical Record should contain all practical's with their implementation and is Mandatory and it carries 5 Marks and Assessment test / Viva-Voce is 5 Marks. The Semester end practical exam contains two sections: Section-A: & Section-B, both are conventional.

M.SC.(STATISTICS) II-SEMESTER

STS-207: PAPER VII: DATA ANALYSIS USING SPSS
PRACTICAL-III

LIST OF PRACTICALS USING SPSS

1. Basic operations of Data entry, Data import and export, I/O files handling etc.
2. **Data Visualization**: Pie diagram, Bar diagram, Histogram, Line plot, frequency curves & polygons, Scatter Plot, Gantt Chart, Box Plot.
3. **Descriptive Statistics:** Measures of Central Tendencies, Dispersions, Relative measures of Dispersions, Moments, Skewness, Kurtosis.
4. **Parametric Tests**: Testing for Mean(s), Variance(s), Proportion(s), ANOVA for one-way two-way and two way with one and m-observations per cell and with & without interactions,
5. **Non–Parametric tests:** Sign test, Wilxon Sign Rank test, Mann-Whitney U-test, Run test, Kolmogorov Smirnov test, Chi-square test for goodness of fit and Chi-square test independence.
6. **Design & Analysis of Experiments:** Analysis of Variances for Completely randomized, randomized block and latin Square Designs and Factorial experiments (2^2, 2^3 F.E. without confounding).
7. **Regression Analysis:** Analysis of Simple and Multiple Linear Regression models, Selection Best Linear Regression Model (All possible, forward, backward, stepwise and stage wise methods). Binary and multinomial Logistic regression models, Probit analysis.
8. **Multivariate Data Analysis:** Linear Discriminant Analysis, Principal Component analysis, Factor analysis, Multi-dimensional scaling, Cluster analysis.
9. **Statistical Quality Control:** Construction Control charts for variables and attributes.

Note: Practical Record should contain all practical's with their implementation and is Mandatory and it carries 5 Marks and Assessment test / viva-Voce is 5 Marks. The Semester end practical exam contains answer any two with their implementation out of four Questions. (Answer any two out of the four)

M.SC. (STATISTICS) II-SEMESTER

STS-208: PAPER VIII: DATA SCIENCE USING PYTHON
PRACTICAL-IV

Objectives & Outcomes:

1. The main objective of this laboratory is to put into practice the ETL (extract, transform, load) pipeline which will extract raw data, clean the data, perform transformations on data, load data and visualize the data.
2. In this course students are expected to extract, transform and load input data that can be textfiles, CSV files, XML files, JSON, HTML files, SQL databases, NoSQL databases etc.,. For doing this, they should learn the following Python libraries/modules: pandas, numpy, BeautifulSoup, pymysql, pymongo, nltk, matplotlib

Datasets:

For this laboratory, appropriate publicly available datasets, can be studied and used.
MNIST (http://yann.lecun.com/exdb/mnist/),
UCI Machine Learning Repository: (https://archive.ics.uci.edu/ml/datasets.html),
Kaggle: (https://www.kaggle.com/datasets)
Twitter Data

LIST OF PRACTICALS DATA SCIENCE USING PYTHON

1. Write programs to parse text files, CSV, HTML, XML and JSON documents and extract relevant data. After retrieving data check any anomalies in the data, missing values etc.
2. Write programs for reading and writing binary files
3. Write programs for searching, splitting, and replacing strings based on pattern matching using regular expressions.
4. Design a relational database for a small application and populate the database. Using SQL do the CRUD (create, read, update and delete) operations.
5. Create a Python Mongo DB client using the Python module pymongo. Using a collection object practice functions for inserting, searching, removing, updating, replacing, and aggregating documents, as well as for creating indexes
6. Write programs to create Numpy arrays of different shapes and from different sources, reshape and slice arrays, add array indexes, and apply arithmetic, logic, and aggregation functions to some or all array elements.
7. Write programs to use the Pandas data structures: Frames and series as storage containers and for a variety of data-wrangling operations, such as:
    - Single-level and hierarchical indexing
    - Handling missing data
    - Arithmetic and Boolean operations on entire columns and tables
    - Database-type operations (such as merging and aggregation)
    - Plotting individual columns and whole tables
    - Reading data from files and writing data to files

DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD-500 007.

M.Sc. (APPLIED STATISTICS) I-SEMESTER

SCHEME OF INSTRUCTIONS AND EXAMINATION W.E.F. A.Y. 2023-24 ONWARDS

Paper #	Paper Code	Paper Title	Credits	Instruction Hours per Week	Semester end Examination duration	Max. Marks in the Semester end Examination	Max. Marks in the Internal Assessments, Assignments, Seminars & Attendance*
THEORY PAPERS							
I	STAS-101	Mathematical Analysis	3	3	3	50	40+10
II	STAS-102	Linear Algebra & Linear Models	3	3	3	50	40+10
III	STAS-103	Applied Probability Theory	3	3	3	50	40+10
IV	STAS-104	Distribution Theory & Estimation	3	3	3	50	40+10
PRACTICAL PAPERS							
V	STAS-105	Statistical Methods using Python Programming	2	4	2	40	10
VI	STAS-106	Linear Algebra & Linear Models (using R)	2	4	2	40	10
VII	STAS-107	Distributions & Estimation (using R)	2	4	2	40	10
VIII	STAS-108	Data Handling using R	2	4	2	40	10
Semester Total			20	12+16**	-	600	

* 75% Attendance is mandatory as per norms; ** For practical with batch of 20 students.

Note: Attendance 10 Marks will be assigned based on proportion of attendance. Four different Internal Assessment tests each with 10 marks. The tests are preferably on multiple choice / fill in the blanks / short answer questions / Quiz / Report writing / Seminar / Assignments.

.

M.SC. (APPLIED STATISTICS) I-SEMESTER

STAS-101: PAPER-I: MATHEMATICAL ANALYSIS (MA)

Objectives & Course Outcomes:

1. The basic concepts of real analysis required for understanding the Mathematical Statistics.
2. To know the applications of Limits, continuity, Convergences, R-S Integrations etc in Statistics.

UNIT-I

Metric spaces - Compact sets - Perfect sets - Connected sets. Limits of functions - Continuous functions - Continuity and compactness, Continuity and connectedness, Discontinuities - Monotonic functions, Differentiation.

UNIT–II

Riemann-Steiltjes (R-S) Integral and its linear properties. Integration by parts, Euler's summation, Riemann's condition. Integrators of Bounded variations. Statements of necessary and sufficient conditions of Riemann - Steiltjes integral. Differentiation under the integral sign. Interchanging the order of integration.

UNIT- III

Sequences and Series of Functions: Uniform convergence - Uniform convergence and continuity - Uniform convergence and integration - Uniform convergence and differentiation – The Stone- Weierstrass theorem.

REFERENCES

1. Walter Rudin: Principles of Mathematical Analysis, McGraw-Hill International 3rd Edition. (Unit-I: pp 30-46 & pp 83-102) (Unit-II: pp 120-133 & 135-142) (Unit-III: pp 143-154, 159-161, 165-171 & 220-222).
2. H.L. Royden: Real Analysis, PHI 3rd edition
3. Apostol, T.M. (1985): Mathematical Analysis, Narosa, Indian Ed.
4. Malik, S.C. (1984): Mathematical Analysis, Wiley – Eastern.
5. Mathematical Analysis Vol - I by D J H Garling.

M.SC. (APPLIED STATISTICS) I-SEMESTER

STAS-102: PAPER-II: LINEAR ALGEBRA & LINEAR MODELS (LA)

Objectives & Course Outcomes:

1. To find the solution to the given set of equations.
2. To transform the given matrix into another form without changing its characteristics.
3. Usage of Matrices for the real time applications.
4. To express the set of equations in a GLM form to find the estimation of parameters.

UNIT – I

Linear Algebra: Vector Spaces with an inner product, Gram –Schmidt orthogonalization process. Ortho-normal basis and orthogonal projection of a vector. Real time applications of orthogonalization in various domains. Solution of matrix equations. Sufficient conditions for the existence of homogeneous and non – homogeneous linear equations. Moore Penrose and generalized inverses and their properties. Real time applications of solving set of equations in various domains.

UNIT–II

Characteristic roots and vectors, Caley–Hamilton theorem algebraic and geometric multiplicity of a characteristic root and spectral decomposition of a real symmetric matrix. Real time applications of characteristic roots and vectors in various domains. Real quadratic forms, reduction and classification of quadratic forms, Index and signature. Simultaneous reduction of two quadratic forms, Extreme of a quadratic form. Matrix Inequalities: Cauchy- Schwartz and Hadamard Inequalities.

UNIT – III

Linear Models: General Linear Model (GLM) and its formulation through examples. Estimability of a linear parametric function. Gauss-Markov linear model, BLUE for linear functions of parameters, relationship between BLUE's and linear Zero-functions. Gauss-Markov theorem, Aitkin's generalized least squares, Concept of Multi-collinearity. Importance and applications of GLM s.

REFERENCES

1. Graybill, F.A. (1983): Matrices with applications in Statistics, 2nd ed., Wards worth.
2. Searle, S.R.(1982) : Matrix Algebra useful for Statistics, John Wiley & Sons.
3. Rao, C.R. and Mithra, S.K. (1971) : Generalized inverse of matrices and its applications, John Wiley & Sons.
4. Rao, A.R. and Bhimasankaram, P. (1992): Linear Algebra, Tata McGraw Hill Publishing Co. Ltd.
5. Searles S.R. (1971): Linear statistical Models.

M.SC. (APPLIED STATISTICS) I-SEMESTER

STAS-103: PAPER-III: APPLIED PROBABILITY THEORY (APT)

Objectives & Course Outcomes:

1. To find the probability based on the conditions that are specified.
2. To obtain distribution function of random variable based on its probability function & vice-versa.
3. To derive characteristic function from the density and vice-versa and identifying the characteristic function.
4. To obtain the probability bounds or moment bounds for the given random variables.
5. To study convergence properties of the sequence random variables based on its probability laws.

(Pre-requisite for understanding: Probability concepts: Classical, statistical and axiomatic definitions, joint, marginal and conditional probabilities, Compound, Addition and Bayes theorems and problems on probability).

UNIT – I

Probability as a measure, Random Variables, distribution function and its properties and their applications. Mathematical Expectation and Expectations of functions of random variables and their applications. Conditional expectation and conditional variances, applications (A list model, random graph, uniform priors, Polyas' urn model and Bose-Einstein distribution, mean time for patterns, the compound Poisson identity, the k-record values of discrete random variables). Characteristic function and its properties, Uniqueness, Inversion and Continuity theorems and their application problems. Identification functions which are/ not be Characteristic functions.

UNIT – II

Probability and Moment inequalities**:** Chebychev's, Markov, Cauchy-Schwartz, Holder, Minkowsky, Liapunov and Jensen Inequalities. Interrelationships among the inequalities and their applications and simple problems on these inequalities. Sequence of random variables: Borel-Cantelli Lemma; Borel 0-1 law. Statement of Glivenko-Cantelli lemma. Convergence of sequence of random variables: law; probability; almost sure, quadratic mean; their implications, counter implications and Slutzky's theorem. Applications of various convergences and their related problems.

UNIT – III

Weak and Strong Law of Large numbers (WLLN): Bernoulli, Chebychev's and Khintchine's WLLNs. Kolmogorov inequality. Borel's SLLNs. Kolmogorov's SLLNs for independent random variables and i.i.d. random variables, Applications of LLN and their related problems. Central Limit Theorems: Demoviere-Laplace form of CLT, Levy-Lindeberg form of CLT, Liapunov's form of CLT and Lindberg-Feller form of CLT and their application related problems.

REFERENCES

1. Ross, S.M (2004): Introduction to Probability Models, 8th Edition, Academic Press
2. Bhat, B.R. (1985): Modern Probability Theory, Wiley Eastern.
3. Bau A.K. (2012): Measure Theory and Probability, PHI, 2nd edition.
4. Rohatgi, V.K. (1993): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern
5. Chandra, T.K. and Chatterji D (2001): A First Course in Probability, Narosa Publishing House.

M.SC. (APPLIED STATISTICS) I-SEMESTER

STAS-104: PAPER-IV: DISTRIBUTION THEORY AND ESTIMATION (DTE)

Objectives and Course Outcomes:

1. To derive any property for any distribution that is specified.
2. To obtain the distribution to the given transformed random variables.
3. To obtain sampling distribution to the given statistic.
4. To derive the point estimator for the given parameter in the distribution function.
5. To derive the properties like consistency, unbiasedness, efficiency, sufficiency, MVB, completeness, etc. for the estimator.

(Pre-requisite: Basic univariate probability distributions: Discrete Uniform, Bernoulli, binomial, Poisson, Negative binomial, Geometric, Hyper geometric, continuous uniform, Normal, Exponential, Gamma (one & two parameters), Beta First and second kinds and Cauchy. Univariate and bivariate random variables transformations).

UNIT – I

Distribution Theory: Definitions and derivations of properties related to Lognormal, Weibull, Pareto, Laplace distributions and their applications and related problems. Compound distribution of Binomial-Poisson, Truncated distributions (Poisson, Exponential and Normal distributions). Exponential family of distribution, Power series distributions, Mixture Distributions. Bivariate Normal distribution. Functions of random variables and their distributions using Jacobian of transformations and problem on transformations, Distributions of Quadratic forms under normality and its applications.

UNIT – II

Multi-nomial and Multivariate Normal distributions and their properties. Derivations of density functions of Sampling Distributions of central and non-central t, F and χ^2 and their properties (for noncentral Statements only), distribution of Sample mean and variance, independence of $\bar{X}$ and S^2. Order statistics: Joint and Marginal distributions of order statistics. Distributions of sample range, Problems on computing the distribution of order statistics. Applications of order statistics.

UNIT – III

Estimation: Point and Interval estimation, Criterion for good point estimator, Minimum Variance Unbiased Estimator, Fisher's information, Cramer-Rao inequality, Rao–Blackwell theorem, Completeness, Lehmann – Scheff's theorem and their applications and its related problems. Estimation of bias and standard deviation of point estimators of Jackknife and Bootstrap methods with examples. MLE and its properties (statements only). Consistency and asymptotic normality of the consistent solutions of likelihood equations. Definition of CAN and BAN estimators and their properties, related examples. Confidence Interval estimation pivotal method for Poisson, Normal and Exponential

REFERENCES

1. Bhuyan K.C. (2010): Probability distribution theory and Statistical Inference, New Central Book agency (P) Ltd.
2. Parimal Mukhopadhya (2002): Mathematical Statistics, Books & Allied Ltd.
3. Johnson, S. and Kotz (1972): Distribution in Statistics, Vol. I, II and III.
4. Rohatgi, V.K. (1984): An Introduction to Probability theory and Mathematical Statistics, Wiley Eastern.
5. Rao, C.R. (1973): Linear Statistical Inference and its applications, Wiley Eastern, 2/e
6. Lehman, E.L. (1983): Theory of Point Estimation, John Wiley and Sons.

M.SC. (APPLIED STATISTICS) I-SEMESTER

STAS-105: PAPER-V: STATISTICAL METHODS USING PYTHON PROGRAMMING

PRACTICAL-I

Objective & Outcomes:

1. Use various data types, loop statements, OOPs concepts, Exemptions, string operations etc for a specified problem
2. Design, implement, debug a given problem using Python
3. Execute the programs using derived and user defined data types.
4. Implement programs using modular approach and file I/O
5. Writing Python code for any statistical methods for the given data data set.

(THEORETICAL CONCEPTS)

Weeks 1-5: Introduction to Python Programming, Input, Processing and Output, Displaying Output with the Print Function, Comments, Variables, Reading Input from the Keyboard, Performing Calculations Operators. Type conversions, Expressions, More about Data Output. Decision Structures and Boolean Logic: if, if-else, if-elif-else Statements, Nested Decision Structures, Comparing Strings, Logical Operators, Boolean Variables. Repetition Structures: recursion and non-recursion, while loop, for loop, Calculating a Running Total, Input Validation Loops, Nested Loops. python-syntax, statements, functions, Built-in-functions and Methods, Modules in python, Exception Handling. Functions: Defining and Calling a Void Function, designing a Program to Use Functions, Local Variables, Passing Arguments to Functions, Global Variables and Global Constants, Value-Returning Functions, Generating Random Numbers, Writing Our Own Value-Returning Functions, The math Module, Storing Functions in Modules. File and Exceptions: Introduction to File Input and Output, Using Loops to Process Files, Processing Records, Exceptions. Finding Items in Lists with in-Operator, List Methods and Useful Built-in Functions, Copying Lists, Processing Lists, Two-Dimensional Lists, Tuples. Strings: Basic String Operations, String Slicing, Testing, Searching, and Manipulating Strings.

LIST OF PRACTICALS (Not using Python Packages)

(Programs must be in a position to write with all possibilities like usage of functions, Loops, OOP concepts, methods, built in functions etc. wherever it is possible)

Week-6:	Program to find the sum and product of two matrices. Program to find the Determinant and Inverse of the given matrix
Week-7:	Program to sort the given set of numbers using bubble sort, Quicksort, Merge sort, insertion sort. Program for linear search, binary search.
Week 8:	Program to find the Median, Mode for the given of array of elements. Program for preparation of frequency tables, Computation of mean, median, mode, variance and standard deviation to the given data set.
Week 9:	Program to compute first four Central & Non-central moments, Skewness and Kurtosis to the given data set.
Week-10:	Program to generate random numbers from Uniform, Binomial, Poisson, Normal and Exponential distributions using algorithms.
Week-11:	Program to Fit Binomial, Poisson & Negative Binomial distributions for the given data set and testing their goodness of fit and drawing the curve plots.
Week-12:	Program to Fit Normal, Exponential & Cauchy distributions for given data set.
Week-13:	Program for finding Correlation and regression lines for the given data set.
Week-14:	Program for testing means, variances, correlations.
Week-15:	Program for carryout the analysis of variance for one way and two way.

Note: Practical Exam question paper will have 15 Marks (5QX3M) weightage on theory concepts and 25 (2Q out of 4 X 8 M + 9 M Execution) on its programs writing and execution. Practical Record should contain all practical's with their implementation and is Mandatory and carries 5 marks and Assessment test / Viva-Voce-5 Marks..

M.SC. (APPLIED STATISTICS) I-SEMESTER

STS-106: PAPER-VI: LINEAR ALGEBRA & LINEAR MODELS
PRACTICAL-II (CONVENSIONAL & USING R)

Objective & Outcomes:

1. Knowing the manual procedures and also their implementation using R
2. Finding the inverse of a matrix in various methods.
3. Applying any transformations on matrices
4. Applying the matrix operations on the given data sets (Determinant, eigen values, eigen vectors, transformations etc).
5. Summarization of properties of the data sets based on matrix operations.

LIST OF PRACTICALS ON LINEAR ALGEBRA & LINEAR MODELS

Week-1:	Inverse of a matrix by Partition method
Week-2:	Solutions of linear equations by sweep-out method
Week-3:	Solutions of linear equations by Doolittle Method
Week-4:	Computation of Moore-Penrose inverse by Penrose method
Week-5:	Computation of generalized inverse of a matrix.
Week-6:	Formation of characteristic equation by using traces of successive powers
Week-7:	Spectral decomposition of a square matrix of third order
Week-8:	Simultaneous reduction of a pair of quadratic forms to diagonal and canonical forms.
Week-9:	Finding orthonormal basis by Gram – Schmidt process.
Week-10:	Computation of variance-covariance matrix for data set and study of its characteristics.
Week-11	Fitting of a simple linear regression model, Testing its lack of fit, and computing its R^2, Adj R^2, Pure error and Confidence interval for regression coefficient
Week-12	Fitting of a Multiple Linear regression model, Testing its lack of fit, and computing its R^2, Adj R^2, Pure error and Confidence interval for regression coefficient.
Week-13	Computation of Simple, partial and Multiple correlation Coefficients
Week-14	Testing Multi-Collinearity

Note: Practical Record should contain all practical's with their implementation and is Mandatory and it carries 5 Marks and Assessment test / Viva-Voce is 5 Marks. The Semester end practical exam contains two sections: Section-A: Conventional & Section-B Using R. (Answer ant two out of four choosing at least one from each section)

M.SC. (APPLIED STATISTICS) I-SEMESTER

STS-107: PAPER-VII: DISTRIBUTION THEORY & ESTIMATION
PRACTICAL-III (CONVENSIONAL & USING R)

Objective & Outcomes:

1. Knowing the manual procedures and also their implementation using R
2. Generation of random samples from any distribution
3. Identifying an appropriate probability distribution to the given data.
4. Fitting and testing the probability distribution.
5. Drawing the probability distribution curves and stating its nature of the distributional curve properties for the given data sets.

LIST OF PRACTICALS ON DISTRIBUTION THEORY & ESTIMATION

Week-1: Generation of random samples from Uniform distribution.
Week-2: Generation of random samples from the Binomial, Poisson, Geometric, Negative Binomial distributions.
Week-3: Generation of random samples from the Normal, Exponential, Gamma, Beta, Cauchy distributions.
Week-4: Fitting an appropriate discrete distribution to the given data sets
Week-5: Fitting an appropriate continuous distribution to the given data sets (Uniform, Normal, Exponential)
Week-6: Testing its Goodness of fit of Cauchy distribution to the given data set
Week-7: Fitting of Gamma distribution with two parameters to the given data set
Week-8: Fitting of Lognormal Distribution with two parameters to the given data set
Week-9: Fitting of Weibull Distribution with two parameters to the given data set
Week-10: Fitting of Pareto distribution with two parameters to the given data set
Week-11 Estimation of parameter in the Cauchy distribution
Week-12 Estimation of Bias, Variance and MSE for mean, median and standard deviation using resampling technique Jackknife
Week-13 Estimation of Bias, Variance and MSE for mean and standard deviation using resampling technique Bootstrap estimator
Week-14 Confidence interval Estimation for the parameters in Poisson, Normal and Exponential distribution using pivot method.

Note: Practical Record should contain all practical's with their implementation and is Mandatory and it carries 5 Marks and Assessment test / viva-voce is 5 Marks. The Semester end practical exam contains two sections: Section-A: Conventional & Section-B Using R. (Answer ant two out of four choosing at least one from each section)

M.SC. (APPLIED STATISTICS) I-SEMESTER

STAS-108: PAPER-VIII: DATA HANDLING
PRACTICAL-IV

Data sets of Kaggle.com can be used for practice. For example few of the them are: Iris Dataset; flights.csv Dataset; Sustainable Development Data; Credit Card Fraud Detection; Employee dataset; Heart Attack Analysis & Prediction Dataset; Dataset for Facial recognition; Covid_w/wo_Pneumonia Chest Xray Dataset; Groceries dataset; Financial Fraud and Non-Fraud News Classification; IBM Transactions for Anti Money Laundering

Data Handling with R:

1. Understanding data with Data types, Measurement of scales, descriptive statistics and data pre-processing steps.
2. Data transformations (Standardize, Normalize, converting data from one scale to other scales).
3. Data Visualization: Drawing One dimensional diagram (Pictogram, Pie Chart, Bar Chart), two-dimensional diagrams (Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot), other diagrammatical / graphical representations like, Gantt Chart, Heat Map, Box-Whisker Plot, Area Chart, Correlation Matrices.
4. Parametric tests (z-, χ^2, t-, F-tests, ANOVA), Correlation & Regression etc.
5. Non-Parametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, Run test).
6. Applying the modelling process, Model evolution, over fitting, under fitting, cross validation concepts, (train/test, K fold and leave out one approaches),
7. Evaluation of Model Performance for classification techniques for qualitative and Quantitative data.
8. Data interpretation and Report writing.

Note: Practical Record should contain all practicals with their implementation and is Mandatory and it carries 5 Marks and Assessment test / viva-voce is 5 Marks. The Semester end practical exam contain answer any two out of four questions with their implantations using R.

DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD – 500 007

M.Sc. (APPLIED STATISTICS) II-SEMESTER

SCHEME OF INSTRUCTIONS AND EXAMINATION W.E.F. A.Y. 2023-24 ONWARDS

Paper #	Paper Code	Paper Title	Credits	Instruction Hours per Week	Semester end Examination duration	Max. Marks in the Semester end Examination	Max. Marks in the Internal Assessments, Assignments, Seminars & Attendance*
THEORY PAPERS							
I	STAS-201	Statistical Inference	3	3	3	50	40+10
II	STAS-202	Sample Theory & Surveys	3	3	3	50	40+10
III	STAS-203	Multivariate Data Analysis	3	3	3	50	40+10
IV	STAS-204	Design and Analysis of Experiments	3	3	3	50	40+10
PRACTICAL PAPERS							
V	STAS-205	Statistical Inference & Sampling Theory	2	4	2	40	10
VI	STAS-206	Multivariate Data Analysis & Design of Experiments	2	4	2	40	10
VII	STAS-207	Statistical Analysis using SPSS	2	4	2	40	10
VIII	STAS-207	Data Science using Python	2	4	2	40	10
		Semester Total	20	12+16*	-	48	120

M.SC. (APPLIED STATISTICS) II-SEMESTER

STAS-201: PAPER-I - STATISTICAL INFERENCE (SI)

(**Pre requisite:** Basic terminologies related to Testing of Hypothesis, Small (t-, F- and χ^2-Tests) and Large sample tests (Z-tests) related to mean, variance, proportions, correlations. NP-lemma and its related problems and concepts of non-parametric tests and basic non parametric tests usage (Sign test, Median test, U-test, run tests).

UNIT – I

Concepts of Most Powerful and Uniformly Most Powerful tests, Neymann – Pearson lemma and its applications to one parameter exponential family of distributions. Concepts of unbiased and consistent tests. Likelihood Ratio Criterion with simple applications (including homogeneity of variances). Statements of asymptotic properties of LR test. The concept of robustness in testing.

UNIT – II

Notions of sequential vs. fixed sample size techniques. Wald's sequential probability Ratio Test (SPRT) for testing Simple null Hypothesis vs. simple alternative. Termination property of SPRT. SPRT procedures for Binomial, Poisson, Normal and exponential distributions and associated OC and ASN functions. Statement of optimality properties of SPRT.

UNIT – III

Non-parametric methods for one-sample problems based on Run test and Kolmogorov – Smirnov test. Wilcoxon Signed rank test for one sample and paired samples. Two sample problems based on Wilcoxon Mann Whitney test. Kolmogorov test (expectation and variances of above test statistics except for Kolmogorov – Smirnov test). Statements about their exact and asymptotic distributions, Wald Wolfowitz Runs test and Normal scores test. Kendall's Tau, Ansari – Bradley test for two-sample dispersion, Kruskal–Wallis test for one – way layout. (k- samples). Friedman test for two-way layout (randomized block).

REFERENCES

1. Lehman, E. L.: Testing of hypothesis, John Wiey
2. Rohatgi, V.K.: An Introduction to Probability Theory and Mathematical Statistics (Wiley Eastern)
3. Gibbons: Non-Parametric Statistical Inference, (Tata Mc Graw Hill)
4. Wald,A. : Sequential Analysis (Dover Publications)
5. Milton and Arnold – Introduction to probability and Statistics (4^{th} Edition)-TMH publication.
6. W.J. Conovar – Practical Non parametric Statistics (John Wiley)

M.SC. (APPLIED STATISTICS) SEMESTER I
STAS-202: PAPER-II: SAMPLING THEORY AND SURVEYS (STS)

Objectives and Course Outcomes:

1. Usage of SRS, Stratified, systematic, Cluster and two stage sampling methods.
2. To estimate the parameters of population and estimating variances.

(Pre requisite: basic terminology, Need & Principal steps in sample surveys , census versus sample surveys, sampling and non- sampling errors, sampling methods. SRSWR, SRSWOR, stratified and systematic sampling methods, Estimates of their population mean, variances etc.)

UNIT – I

Unequal probability Sampling – Probability proportional to size (PPS) sampling with and without replacements (ppswr / wor) methods - drawing samples using Cumulative total and Lahiri's methods. Horwitz -Thompson, Hansen – Horwitz and Yates and Grundy estimators for population mean, total and their variances. Ratio Method of Estimation - Concept of ratio estimators, Ratio estimators in SRS, their bias, variance/MSE. Ratio estimators in Stratified random sampling – Separate and combined estimators, their variances/MSE.

UNIT-II

Regression method of estimation – Concept Regression estimators, Regression estimators in SRS with pre–assigned value of regression coefficient (Difference Estimator) and estimated value of regression coefficient, their bias, variance/MSE, Regression estimators in Stratified Random sampling – Separate and combined regression estimators, their variances/ MSE. Cluster Sampling - Cluster sampling with clusters of equal sizes, estimator of mean per unit, its variance in terms of intracluster correlation coefficient, determination of optimum sample and cluster sizes for a given cost. Cluster sampling with clusters of unequal sizes, estimator of population mean and its variance/MSE.

UNIT – III

Sub sampling (Two–Stage only) - Equal first stage units – Estimator of population mean, variance/MSE, estimator of variance. Determination of optimum sample size for a given cost. Unequal first stage units – estimator of population mean and its variance/MSE. Planning of Sample Surveys - Methods of data collection, problem of sampling frame, choice of sampling design, pilot survey, processing of survey data. Non-sampling errors - Sources and treatment of non-sampling errors. Non – sampling bias and variance.

REFERENCES

1. Parimal Mukhopadhyay (1998) : Theory and methods of Survey sampling, Prentice – Hall of India, New Delhi.
2. Cochran, W.C. (1977) : Sampling Techniques, Third Edition, Wiley Eastern.
3. Daroga Singh and Chowdary (1986) : Theory and Analysis of Sample Survey Designs – Wiley Eastern Ltd.
4. Des Raj (1976) : Sampling Theory, Tata McGraw Hill, New Delhi.
5. Sukhatme et. Al (1984): Sampling Survey methods and its applications, Indian society of Agricultural Statistics.
6. Murthy, M.N. (1967) : Sampling theory, Tata McGraw Hill, New Delhi.

M.SC (APPLIED STATISTICS) II-SEMESTER II

STAS-203: PAPER III: MULTIVARIATE DATA ANALYSIS (MDA)

Objectives and Course Outcomes:

1. To understand the distribution of Multivariate data.
2. To understand the multivariate statistical tools.
3. To identify and applying of multivariate techniques for data analysis.

UNIT – I

Concept of Bivariate and multivariate random variables, concept of random vector, its expectation, and variance-covariance matrix, marginal and joint distributions, stochastic independence of random vectors, conditional distributions. Multinomial Distribution and its properties (Marginal, Conditional, MGF, Ch.F, Correlation), Multivariate normal distributions and its properties (Marginal, Conditional, MGF, Ch.F, Correlation), Distribution of sample mean vector, Independence of sample mean vector and sample variance-covariance matrix, Maximum likelihood estimates of parameters (Mean vector and covariance matrix). Sample dispersion matrix, statement of Wishart distribution and its simple properties; Wilk's λ criterion and its distribution, statements of its properties;

UNIT - III

Hotelling's T^2 statistic, Null distribution of Hotellings' T^2 and Applications of Hotelling T^2 Statistic (single and two sample mean vector cases), Mahalanobis D^2 statistic, Concepts of Discriminant analysis, Computation of linear discriminant function using likelihood ratios based on Multivariate normal populations and Fisher's Linear Disciminent Function, Bayes Mis-classification, Relationship between Discriminant function and Mahalanobis D^2 statistic. Path analysis and computation of path coefficients. Correspondence analysis.

UNIT – IV

Principal component analysis: Introduction, Derivation of Principal components and statements of its properties; Factor analysis: Introduction, simple factor model, Orthogonal factor model construction. Introduction to multidimensional scaling, some related theoretical results, similarities, metric and non-metric multidimensional scaling methods. Canonical variables and canonical correlations, Cluster analysis: Introduction, similarities and dissimilarities, Single, Complete and average linkage methods.

REFERENCES

1. Johnson, R.A, and Dean W. Wichern: Applied Multivariate Statistical Analysis.
2. Morrison, D: An Introduction to Multivariate Analysis.
3. Seber : Multivariate Observations
4. Anderson: An Introduction to Multivariate Analysis.
5. Bishop: Analysis of Categorical data.

M.SC. (APPLIED STATISTICS) II-SEMESTER

STAS-204: PAPER IV: DESIGN OF EXPERIMENTS (DOE)

Objectives and Course Outcomes:

1. Analysis of the experimental data using full factorials, with partial and total confounding.
2. Analysis of the experimental data using one way and two classifications.
3. To estimate the parameters of population and estimating variances.

Pre requisite: Concept of analysis of Variance and ANOVA for one-way and two-way classifications with one observation per cell, expectation of various sums of squares, Statistical analysis, Analysis of Completely randomized, Randomized Block and Latin Square Designs including estimation missing observations and efficiencies)

UNIT-I

Analysis of co-variance: one–way and two–way classifications. Multiple comparisons, Fisher Least Significance Difference (L.S.D) test and Duncan's Multiple range test (DMRT). Estimation of main effects, interactions and analysis of 2^k factorial experiment in general with particular reference to k = 2, 3 and 4 and 3^2 factorial experiments. Total and partial confounding in case of 2^3, 2^4 and 3^2 factorial designs. Concept of Balanced partial confounding.

UNIT-II

Fractional replications of factorial designs – one-half replication of 2^3 & 2^4 design, one-quarter replication of 2^5 and 2^6 designs. Resolution of a design, Split – plot design.

Balanced incomplete block design (BIBD) – parametric relations, intra-block analysis, recovery of inter-block information. Youden Square design and its analysis.

UNIT-III

Partially balanced incomplete block design with two associate classes PBIBD (2), Parametric relations, intra block analysis. Simple lattice design. Concept of Response surface methodology (RSM), the method of steepest ascent. Response surface designs. Design for fitting first – order and second – order models. Variance of estimated response. Second order rotatable designs (SORD), Central composite designs (CCD): Role of CCD as an alternative to 3^k design, Rotatability of CCD.

REFERENCES

1. Das, M.N. and Giri,N.: Design and Analysis of Experiments, Wiley Eastern.
2. Montogomery, D.C. : Design and Analysis of Experiments, John Wiley.
3. Parimal Mukhopadhyay : Applied Statistics, New Central Book Agency.
4. Cochran and Cox : Experimental designs, John Wiley.
5. Kempthrone : Desing and Analysis of Experiments, John Wiley.
6. Kapoor and Gupta : Applied Statistics, Sultan Chand.

M.SC. (APPLIED STATISTICS) II-SEMESTER

STAS-205: PAPER-V: STATISTICAL INFERENCE AND SAMPLING THEORY
PRACTICAL-I

(CONVENTIONAL PRACTICAL)

SECTION-A: LIST OF PRACTICALS ON STATISTICAL INFERENCE

1. Type I and Type II errors
2. Most Powerful tests
3. Uniformly Most Powerful tests
4. Likelihood ratio Tests
5. SPRT procedures for Binomial, Poisson and Normal and their OC, ASN function
6. Wilcoxon Signed rank test
7. Wilcoxon Mann-Whitney test
8. Kolmogorov – Smirnov one sample, two sample tests
9. Ansari – Bradley test for two sample dispersion
10. Krusakal Walli's test for one way layout
11. Friedman test for two-way layout
12. Normal Scores test
13. Kendall's Tau

SECTION-B: LIST OF PRACTICALS ON SAMPLING THEORY

1. PPS sampling with and without replacements.
2. Ratio estimators in SRS, comparison with SRS
3. Separate and combined ratio estimators, Comparison.
4. Regression estimators in SRS, Comparison with SRS and Ratio estimators
5. Separate and combined Regression estimators, Comparison.
6. Cluster sampling with equal cluster sizes.
7. Sub sampling (Two–stage sampling) with equal first stage units.

Note: Practical Record should contain all practical's with their implementation and is Mandatory and it carries 5 Marks and Assessment test / Viva-Voce is 5 Marks. The Semester end practical exam contains two sections: Section-A: & Section-B, both are conventional.

M.SC. (APPLIED STATISTICS) II-SEMESTER

STAS-206: PAPER VI: DESIGNS & ANALYISIS OF EXPERIMENTS AND MULTIVARIATE DATA ANALYSIS
PRACTICAL-I (CONVENTIONAL)

SECTION-A LIST OF PRACTICALS ON DESIGN AND ANALYSIS OF EXPERIMENTS

Week-1: DMR and LSD tests
Week-2: Analysis of Covariance for one-way classification
Analysis of Covariance for two-way classification
Week-3: Analysis of Variance for 2^3, 2^4 factorial experiments
Analysis of Variance for 3^2 factorial experiments.
Week-4: Identification of Confounded terms in 2^3, 2^4 and 3^2 factorial experiments.
Construction of design with a specified effect is confounded.
Week-5: Analysis of Variance for Total confounding of $2^{3,}$ 2^4 designs
Analysis of Variance for Partial confounding of $2^{3,}$ 2^4 designs.
Week-6: Analysis of Variance for one–half fraction of 2^4 designs and
Analysis of Variance for one–quarter fraction of 2^5 designs.
Week-7: Analysis of variance for Split-Plot design.
Week-8: Analysis of Balanced Incomplete Block Design
Week-9: Analysis of Youden Square Design
Week 10: Analysis of Partially Balanced Incomplete Block Design

SECTION-B: LIST OF PRACTICALS ON MULTIVARIATE ANALYSIS

Week-1: MLE of Mean vector and variance covariance Matrix based on the sample drawn from p- Normal population.
Week-2: Hotelling's T^2 for test the mean vector based on single sample
Mahalanobi's D^2 for test the mean vector based on single sample
Hotelling's T^2 for testing equality of the mean vectors based on two samples
Mahalanobi's D^2 for testing equality of the mean vectors based on two samples.
Week-3: Computation of Principal Components.
Week-4: Classification between two normal populations by discriminant analysis using Maximum likely hood ratio approach and Bayesian mis classification.
Week-5: Cluster analysis using Single, Complete and Average linkages.
Week-6: Computation of Canonical variables and correlation.
Week-7: Computation of Orthogonal Factor Model
Week-8: Computation of Path coefficients and drawing Path diagram
Week-9: Computation of Multidimensional Scaling

Note: Practical Record should contain all practical's with their implementation and is Mandatory and it carries 5 Marks and Assessment test is 5 Marks. The Semester end practical exam contains two sections: Section-A: & Section-B, both are conventional.

M.SC. (APPLIED STATISTICS) II-SEMESTER

STAS-207: PAPER VII: DATA ANALYSIS USING SPSS & OR USING TORA
PRACTICAL-III

LIST OF PRACTICAL'S USING SPSS

1. Basic operations of Data entry, Data import and export, I/O files handling etc.
2. **Data Visualization**: Pie diagram, Bar diagram, Histogram, Line plot, frequency curves & polygons, Scatter Plot, Gantt Chart, Box Plot.
3. **Descriptive Statistics:** Measures of Central Tendencies, Dispersions, Relative measures of Dispersions, Moments, Skewness, Kurtosis.
4. **Parametric Tests**: Testing for Mean(s), Variance(s), Proportion(s), ANOVA for one-way two-way and two way with one and m-observations per cell and with & without interactions,
5. **Non–Parametric tests:** Sign test, Wilxon Sign Rank test, Mann-Whitney U-test, Run test, Kolmogorov Smirnov test, Chi-square test for goodness of fit and Chi-square test independence.
6. **Design & Analysis of Experiments:** Analysis of Variances for Completely randomized, randomized block and latin Square Designs and Factorial experiments (2^2, 2^3 F.E. without confounding).
7. **Regression Analysis:** Analysis of Simple and Multiple Linear Regression models, Selection Best Linear Regression Model (All possible, forward, backward, stepwise and stage wise methods). Binary and multinomial Logistic regression models, Probit analysis.
8. **Multivariate Data Analysis:** Linear Discriminant Analysis, Principal Component analysis, Factor analysis, Multi-dimensional scaling, Cluster analysis.
9. **Statistical Quality Control:** Construction Control charts for variables and attributes.

M.SC. (APPLIED STATISTICS) II-SEMESTER

STAS-208: PAPER VIII: DATA SCIENCE USING PYTHON
PRACTICAL-IV

OBJECTIVES & OUTCOMES:

1. The main objective of this laboratory is to put into practice the ETL (extract, transform, load) pipeline which will extract raw data, clean the data, perform transformations on data, load data and visualize the data.
2. In this course students are expected to extract, transform and load input data that can be textfiles, CSV files, XML files, JSON, HTML files, SQL databases, NoSQL databases etc.,. For doing this, they should learn the following Python libraries/modules: pandas, numpy, BeautifulSoup, pymysql, pymongo, nltk, matplotlib

DATASETS:

For this laboratory, appropriate publicly available datasets, can be studied and used.
MNIST (http://yann.lecun.com/exdb/mnist/),
UCI Machine Learning Repository: (https://archive.ics.uci.edu/ml/datasets.html),
Kaggle: (https://www.kaggle.com/datasets)
Twitter Data

LIST OF PRACTICALS

1. Write programs to parse text files, CSV, HTML, XML and JSON documents and extract relevant data. After retrieving data check any anomalies in the data, missing values etc.
2. Write programs for reading and writing binary files
3. Write programs for searching, splitting, and replacing strings based on pattern matching using regular expressions.
4. Design a relational database for a small application and populate the database. Using SQL do the CRUD (create, read, update and delete) operations.
5. Create a Python Mongo DB client using the Python module pymongo. Using a collection object practice functions for inserting, searching, removing, updating, replacing, and aggregating documents, as well as for creating indexes
6. Write programs to create Numpy arrays of different shapes and from different sources, reshape and slice arrays, add array indexes, and apply arithmetic, logic, and aggregation functions to some or all array elements.
7. Write programs to use the Pandas data structures: Frames and series as storage containers and for a variety of data-wrangling operations, such as:
    - Single-level and hierarchical indexing.
    - Handling missing data.
    - Arithmetic and Boolean operations on entire columns and tables.
    - Database-type operations (such as merging and aggregation).
    - Plotting individual columns and whole tables.
    - Reading data from files and writing data to files.

DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD – 500 007

M.Sc. (STATISTICS) III-SEMESTER

SCHEME OF INSTRUCTIONS AND EXAMINATION FOR 2022-2024 BATCH

Paper	Sub. Code	Paper Title	Credits	Instruction Hours per Week	Semester end Exam duration	Max. Marks in Semester end Exam	Max. Marks in the Internal Assessment and Assignments
THEORY PAPERS							
I	STS-301	Non-Parametric Inference (NPI)	3	3	3	70	20+10
II	STS-302	Quality control & Optimization Techniques (QCOT)	3	3	3	70	20+10
III	STS-303 **Elective – I**	A) Applied Regression Models (ARM) B) Econometric Models (EM) C) Advanced Design & Analysis of Experiments (ADAE). D) Statistical Process in Data Science	3	3	3	70	20+10
IV	STS-304 **Elective – II**	A) Data Mining Techniques (DM) B) Bayesian Inference (BI) C) Advanced Machine Learning Techniques (MDLT)	3	3	3	70	20+10
PRACTICALS PAPERS							
V	STS-305	Non-Parametric Inference & Optimization Techniques (NPI & QCOT)	2	4	2	50	-
VI	STS-306	Practicals on E-I & E-II	2	4	2	50	-
VII	STS-307	Statistical Analysis using R & TORA	2	4	2	50	-
VIII	STS-308	Data Analysis Project (Mini project)	2	4	2	**50**	-
Semester Total			**20**	**12+16***	-	**600**	

M. SC. (STATISTICS) III-SEMESTER

STS-301: PAPER I - NON PARAMETRIC INFERENCE (NPI)

Unit–I

Non-parametric density estimation: Density estimates, survey of existing methods. Rosenblatt's naïve density estimator, its bias and variance. Consistency of Kernel density estimators and its MSE.

Unit–II

Nonparametric Tests: one-sample problems based on sign test, Wilcoxon signed Rank test, run test and Kolmogorov – Smirnov test. Two sample problems based on sign test, Wilcoxon signed rank test for paired comparisons, Wilcoxon Mann-Whitney test, (Expectations and variances of above test statistics, Statements about their exact and asymptotic distributions).

Unit–III

Nonparametric Tests: Two sample problems based on Kolmogorov – Smirnov Test, Wald–Wolfowitz Runs test and Normal scores test. Ansari–Bradley test for two sample dispersions. (Expectations and variances of above test statistics, Statements about their exact and asymptotic distributions).

Unit–IV

Nonparametric Tests: Chi–Square test of goodness of fit and independence in contingency tables. Tests for independence based on Spearman's rank correlation and Kendall's Tau. Kruskal–Wallis test for one-way layout (K-samples). Friedman test for two-way layout (randomised block).

Unit–V

Asymptotic Relative Efficiency (ARE) and Pitman's theorem. ARE of one sample, paired sample and two sample locations tests. The concept of Rao's second order efficiency and Hodges–Lehman's deficiency with examples.

REFERENCES

1. Gibbons – Non-parametric Statistical Inference (1978)
2. Myles Hollander and Douglas A. Wolfe: Nonparametric statistical methods (John Wiley)
3. Silverman: Density estimation for statistics and data analyses.
4. W.J. Conover – Practical Non parametric Statistics (John Wiley)
5. Sidney Siegel – Non-parametric Statistics for Behavioural Science, Mc. Graw Hill.
6. Ferguson, T.S. – Mathematical Statistics, A decision theoretic approach (Academic press)

M.SC. (STATISTICS) III-SEMESTER

STS-302: PAPER II: QUALITY CONTROL AND OPTIMIZATION TECHNIQUES

Unit–I

Review of control charts for variable data and attributes : O.C. and A.R.L. functions of control charts for variables and attributes, modified control charts for variables and Acceptance control charts for attributes, control by gauging. Moving Average and exponentially weighted moving average charts, Cu-sum charts using V-Masks and decision intervals.

Unit–II

Process Capability Analysis: Capability indices Cp, Cpk and Cpm, estimation, confidence intervals and tests of hypotheses relating to capability indices for normally distributed characteristics. Acceptance sampling plans for attributes, single, double and sequential sampling plans and their properties

Unit-III

Rectifying inspection plans for attributes, AOQ, AOQL, designing of Rectifying Sampling Plans for specified AOQL and LTPD. Sampling Plans for inspection by variables for one–sided and two–sided specifications; Dodges Continuous sampling Plan–l and its properties, modifications over CSP–l.

Unit–III

Review on LPP, Graphical & simplex, Charmers methods, Duality in LPP; Duality and Complementary slackness theorems. Primal and dual relation. Dual simplex Algorithm; Sensitivity Analysis: Discrete changes requirement and cost vectors; parametric programming: Parameterisation of cost and requirement vectors.

Integer Programming Problem: Gomory's cutting plane Algorithm for pure and mixed IPP Branch and bound Technique.

Unit–IV

Basic concepts of Networks constraints; Construction of Network and critical path; PERT and CPM; Network flow problems. Time Cost Analysis.

Inventory: Introduction; ABC analysis and Deterministic Inventory models with and without shortages.

REFERENCES

1. Montgomery, D.C.(1985) : Introduction to Statistical Quality Control, Wiley
2. Wetherill, G.B. (1977): Sampling Inspection and Quality Control, Halsted Press.
3. Cowden, D. J. (1960) : Statistical Methods in Quality Control, Asia Publishing House.
4. Kantiswarup; Gupta P.K. and Singh, M.N.(1985) : Operations Research; Sultan Chand
5. Taha, H.A.(1982): Operations Research : An Introduction; MacMillan
6. Sharma,S.D.: Operations Research.
7. Ott,E.R. (1975) : Process Quality Control, McGraw Hill
8. Phadke, M.S. (1989): Quality Engineering through Robust Design, Prentice Hall.
9. Wetherill, G.B., and Brown, D.W: Statistical Process Control: Theory and Practice, Chapman and Hall.
10. Hillier F.S. and Leiberman,G.J.(1962) : Introduction to Operations Research; Holdon Day

M.SC. (STATISTICS) III-SEMESTER

ELECTIVE-I: STS-303(A): PAPER-III (A): APPLIED REGRESSION MODELS (ARM)

Unit–I

Selection of best linear regression: Introduction to selection of best linear regression, all possible regression, backward, forward, step-wise, stage-wise regressions. Ridge regression.

Unit–II

Non-linear regression: Introduction to non-linear regression model, some commonly used families of non-linear regression functions, statistical assumptions and inferences for non-linear regression, linearizable models, determining the Least squares estimates, The Gauss – Newton method, ML estimation, (D&S).

Unit–III

Logistic regression model: Introduction to simple Logistic model, Fitting the model, testing for the significance of the coefficients, Logistic model for Dichotomous independent variable; Introduction to multiple Logistic regression, fitting the multiple logistic regression model, testing for the significance of the model.

Unit–IV

Probit Analysis: Introduction, Analysis of Biological data, sigmoid curve, fitting a Probit Regression line through least squares method.

Robust Regression: Introduction, Least absolute deviations regression (L_1 Regression), M–estimators, examples, and Least Median of Squares (LMS) regression, Robust Regression with Ranked Residuals.

Unit-V

Generalized Linear Models: Introduction, the exponential family of distributions, fitting GLIM. Concept of Mixed, Random Effects and Fixed Models–Introduction, General description, estimation, estimating variance components from balanced data.

REFERENCES

1. Regression Analysis: Concepts and Applications, Franklin A. Graybill and Hariharan K. Iyer
2. Applied Regression Analysis: Norman R. Draper and Harry Smith
3. Applied Regression Analysis, linear models and related methods: John Fox
4. Non–linear Regression Analysis and its Applications: Douglas M. Bates and Donald G. Watts
5. Applied Logistic Regression: David W. Hosme and Stanley Lemeshow.
6. Linear Models for unbalanced Data: Shayler Searle
7. Residuals and Influence in Regression: R. Dennis Cook and Sanford Weisberg
8. Log–linear models and Logistic Regression: Ronald Christensen.

M.SC. (STATISTICS) III-SEMESTER

ELECTIVE-I: STS-303(B): PAPER-III (B): ECONOMETRIC MODELS (EM)

Unit–I

Meaning and scope of econometrics. Concepts of dummy variables and proxy variable. Problems and methods of estimation in single equation regression Models

Multicollinearity: Consequences of multicollinearity, tests to detect its presence and solutions to the problem of multicollinearity.

Unit–II

Generalised Least Squares: Estimates of regression parameters – Properties of these estimates.

Heteroscedasticity: Consequences of heteroscedastic disturbances – test to detect its presence and solutions to the problem of heteroscedasticity.

Unit–III

Auto Correlation: Consequences of autocorrelated disturbances, Durbin – Watson test – Estimation of autocorrelation coefficient (for a first order autoregressive scheme).

Distributed lag models: study of simple finite lag distribution models – Estimation of the coefficients of Kayak geometric lag model.

Instrumental Variable: Definition – derivation of instrument variable estimates and their properties.

Unit–IV

Errors in variables: Problem of errors in variables simple solutions using instrumental variables technique. Simulation equation models and methods of estimation: distinction between structure and Model–Exogenous and Endogenous variables – Reduced form of a model.

Unit–V

Problem of identification – Rank and order conditions and their application.

Methods of estimation: Indirect least squares. Two stages least squares, three stages least squares. A study of merits and demerits of these methods.

REFERENCES

1) Johnston – Econometrics Methods (2^{nd} Edition) :
2) G. S. Maddala – Econometrics
3) A. Koutsoyiennis – Theory of econometrics

M.SC. (STATISTICS) III-SEMESTER

ELECTIVE-I: STS-303(C): PAPER-III(C): ADVANCED DESIGN OF EXPERIMENTS (ADE)

Unit–I

General incomplete block designs and its information matrix. Balanced Incomplete block design (BIBD) – Parametric relations, intra–block analysis, recovery of inter–block information. Concepts of Symmetric, Resolvable and Affine resolvable BIBDS. Construction of BIBDS using MOLS.

Unit–II

Partially balanced incomplete block design with two–associate classes PBIBD(2)–Parametric relations, intra–block analysis, Four different association schemes.

UNIT-III

Youden Square design and its analysis. Lattice designs, Balanced Lattice Design, Simple Lattice Design and their analysis. Construction of Youden square, balanced lattice designs

Unit–IV

Concept of Response surface methodology (RSM), the method of Steepest ascent. Response surface designs–designs for fitting first–order and second– order models, Variance of estimated response. Second order rotatable designs (SORD), central composite designs (CCD)–role of CCD as alternative to 3^k designs, rotatability of CCD.

Unit–V

Experiments with mixtures–Simplex Lattice designs, first-order and second-order mixture models and analysis. Optimum designs–various optimality criteria and their interpretations. Repeated measurements designs. Cross–over designs and Row–Column designs.

REFERENCES

1. Montgomery, D.C.: Design and Analysis of Experiments
2. Parimal Mukhopadhyay : Applied Statistics
3. Das, M.N., and Giri, N.: Design and Analysis of Experiments
4. Myers, R.H. : Response Surface Methodology
5. Aloke Dey : Theory of Block Designs
6. Cornell, M : Mixture Experiments

M.SC.(STATISTICS) III-SEMESTER

ELECTIVE-I : STS-303 (D): PAPER-III (D) : STATISTICAL PROCESS IN DATA SCIENCE

Unit – I

Data Visualization: Data types, Measurement of scales, understanding data with descriptive statistics. Data visualization techniques: Pictogram, Pie Chart, Bar Chart, Histogram, Line plot, frequency curves& polygons, ogive curves, Scatter Plot, Gantt Chart, Heat Map, Box and Whisker Plot, Waterfall Chart, Area Chart, Stacked Bar Charts - Sub Plots – Matplotlib, Seaborn Styles, Box plot - Density Plot - Tree map - Graph Networks. Visual Perception and Cognition, Applications of Principles of Information Visualization, Dashboard Design.

Unit-II

Data Pre-processing: Understanding data with Descriptive statistics. Data pre-processing steps, Data transformations (Standardize, Normalize, converting data from one scale to other scales). Identification suitable basic statistical tools / tests Parametric tests (z-, χ^2, t-, F-tests), Nonparametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, K-S, Wald-Wolfowitz run test) for the data sets. Feature selection methods

Unit-III

Introduction to Data Modelling: Review of the modelling process, Concepts of Classification & Clustering, Supervised and Un-supervised Modelling, Concepts of Model evolution, Cross validation concepts, (train/test, K fold and Leave out one approaches), Model Performance evaluation for Qualitative and Quantitative data, Model improvement and saving models for future use (classification matrix, Precision and Recall, F1 score, Sensitivity, Specificity, ROC curve) and Model performance concepts for regression (MSE, RMSE, R^2, adj R^2, MAPE),

Unit-IV

Concepts of Model improvement (Tuning parameters using manual search, Manual grid search, random search) and saving models for future use. Simple linear regression and its analysis (model fitting, regression ANOVA, testing lack of fit, MSE, RMSE, R^2, adj R^2, testing regression coefficients and confidence limits).

Unit-V

Basic concepts on Multivariate data; Simple, Partial & Multiple correlations; Multi collinearity; Multiple linear regression and its analysis; Selection of best linear regression (over fitting & under fitting) & its methods in outline (all possible, forward, backward, step-wise and stagewise). Simple and Multiple Logistic models fitting and its analysis.

REFERENCES

1) Foster Provost & Tom Fawcett, Data science for Business, O'REILLY Publications
2) Henrik Brink, Joseph W. Richards. Mark Fetherolf, Real World Machine Learning, Manning Publications.
3) Foster Provost & Tom Fawcett, Data science for Business, O'REILLY Publications
4) Henrik Brink, Joseph W. Richards. Mark Fetherolf, Real World Machine Learning, Manning Publications
5) Brett Lantz, Machine Learning with R, Packt Publications.

M.SC.(STATISTICS) III-SEMESTER

ELECTIVE-II : STS-304(A): PAPER-IV(A) : DATA MINING (DM)

Unit-I

Introduction: Challenges, Origins of Data Mining, Data Mining Tasks; **Data:** Types of Data, Data Quality, Data Preprocessing, Measures of Similarity and Dissimilarity; **Exploring Data:** Visualization, OLAP and Multidimensional Data Analysis

Unit-II

Classification: Preliminaries, General approach to solving a classification problem, Decision tree induction, Model Over-fitting, – Evaluating the performance of a classifier – Methods of comparing classifiers; Rule-based classifier, Nearest-Neighbor classifiers, Bayesian classifiers

Unit-III

Classification: Artificial Neutral Networks, Perceptron classifier, Support vector machine, Ensemble methods, Class imbalance problem – Multiclass problem

Unit-IV

Cluster Analysis: Agglomerative hierarchical clustering, K-means, DBSCAN, C4.5, CART Cluster evaluation.

Unit-V

Association Analysis: Problem definition, Frequent item set generation, Rule generation, Compact representation of frequent item sets, Alternative methods for generating frequent item sets, FP-Growth Algorithm, Evaluation of Association patterns, Effect of Skewed support distribution; Handling categorical attributes. Handling continuous attributes, Handling a concept hierarchy.

REFERENCES

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2008): “Introduction to Data Mining”, Pearson Education.
2. Arun K Pujari, Data Mining Techniques, University Press, 2nd Edn, 2009.
3. K.P. Soman, Shyam Diwakar, V.Ajay, Insight into Data Mining Theory and Practice, PHI, 2010.
4. Vikram pudi P. Radha Krishna, Data Mining, Oxford University Press, 1st Edition 2009
5. Galit S, Nitin RP, Peter C Bruce. Data Mining for Business Intelligence. Wiley India Edition,2007.

M.SC. (STATISTICS) III-SEMESTER

ELECTIVE-II : STS-304(B): PAPER-IV(B): BAYESIAN INFERENCE (BI)

Unit-I

Bayes theorem, Bayesian Concept to priors and posteriors, computation of the posterior distribution.) subjective prior distribution, Conjugate family of priors of a parameter. Hyper parameters of a prior from conjugate family. Conjugate families for (i) exponential family models, (ii) models admitting sufficient statistics of fixed dimension.

Unit-II

Subjective prior distribution of a parameter. Non informative, improper and invariant priors. Jeffrey's invariant prior. Bayesian point estimation as a prediction problem from posterior distribution. Bayes estimators for (i) absolute error loss (ii) squared error loss (iii) 0 - 1 loss. Generalization to convex loss functions.

Unit-III

Evaluation of the estimate in terms of the posterior risk. Bayesian interval estimation: Credible intervals. Highest posterior density regions. Interpretation of the confidence coefficient of an interval and its comparison with the interpretation of the confidence coefficient for a classical confidence interval.

Unit-IV

Bayesian testing of Hypothesis: Specification of the appropriate form of the prior Distribution for a Bayesian testing of hypothesis problem. Prior odds, Posterior odds, Bayes factor for various types of testing hypothesis problems depending upon whether the null hypothesis and the alternative hypothesis are simple or composite. Bayesian prediction problem. Large sample approximations for the posterior distribution.

UNIT-V

Estimation of parameters using Markov Chain Monte Carlo methods: Gibbs Sampler and Metropolis-Hasting Method and other computer simulation methods. Bayesian calculations for non-conjugate priors: (i) Importance sampling, (ii) Obtaining a large sample of parameter values from the posterior distribution using Acceptance - Rejection methods.

REFERENCES

1. Berger, J. O. Statistical Decision Theory and Bayesian Analysis, Springer Verlag.
2. Robert C. P. and Casella, G. Monte Carlo Statistical Methods, Springer Verlag.
3. Leonard T. and Hsu, J. S. J. Bayesian Methods. Cambridge University Press.
4. Box, G. P. and Tiao, G. C. Bayesian Inference in Statistical Analysis, Addison - Wesley.

M.SC.(STATISTICS) III-SEMESTER

ELECTIVE-II : STS-304(C): PAPER-IV(C) : ADVANCED MACHINE LEARNING TECHNIQUES

Unit – I

Basic Concepts to Statistical Pattern Recognition, Pattern Recognition System, Fundamental problems in Pattern Recognition. Linear classifies, Multiple Linear regression, Logistic regression, Linear Discriminant Function (for binary outputs) with minimum squared error, Naïve Bayes classifier, Support Vector Machines, KNN algorithm

Unit – II

Decision Tree algorithms, Random Forest algorithm, Bagging, Gradient boosting, Ada-Boosting and XG-Boosting algorithm, Market-Basket Analysis.

Unit – III

Cluster Analysis: Introduction, similarities and dissimilarities, Hierarchical clustering, Single linkage method, k-means and k-Nearest Neighbourhood (KNN) clustering,

Unit – IV

Introduction to Artificial Neuron Networks and its characteristics; Algorithms of Perceptron Learning; Multi-layer Perceptron Learning, Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning. Back-Propagation and their applications.

UNIT – V

Reinforcement learning, Markov Decision Process, Hidden Markov Model, Convolutional Neural Networks, Recurrent Neural Networks, Long-Short Term Memory Networks.

REFERENCES

1. Shai Shalev-Shwartz, Shai Ben-David Understanding Machine Learning: From Theory to Algorithms, Cambridge University press.
2. Marc Peter Deisenroth, A Aldo faisal, Cheng soon Ong: “Mathematics for Machine Learning”, Cambridge University Press, First Edition.
3. Hayes: Artificial Neural networks

M.SC.(STATISTICS) III-SEMESTER

STS-305: PAPER-V : NON-PARAMETRIC INFERENCE, QUALITY CONTROL & OPTIMIZATION TECHNIQUES (NPI & QCOT)

PRACTICAL -I (CONVENTIONAL)

SECTION-A: LIST OF PRACTICALS ON NON–PARAMETRIC INFERENCE

1. Sign test and Wilcoxon signed rank test (including paired comparison)
2. Run test for randomness
3. Two Samples:
    a) Wilcoxon Mann-Whitney test
    b) Kolmogorov – Smirnov test
    c) Wald Wolfowitz test
4. Goodness of fit: Chi–square and Kolmogorov – Smironov test
5. Normal Scores test
6. Kruskal–Wallis for one–way layout
7. Friedman test for two–way layout
8. Tests for independence in contingency tables: Spearman's rank correlation, Kendall's Tau
9. Ansari-Bradley test for two sample dispersions.

SECTION-B: LIST OF PRACTICALS ON QUALITY CONTROL, OPTIMIZATION TECHNIQUES

1. Construction of OC and ASN curves for $\overline{X}$ and R - charts
2. Construction of OC and ASN curves for Control charts for attributes (p,d,c; for a fixed n)
3. Construction of simple and exponentially weighted moving average control chart and simple moving range control chart.
4. Construction of CUSUM charts V – Mark and Tabular methods
5. Construction of AOQ and AFI curves for CSP–I
6. Dual Simplex Method
7. Sensitivity Analysis
8. Parametric Programming Problem
9. Integer Programming Problem
10. Evaluation of project time through CPM and PERT
11. Evaluation of Time cost analysis through CPM and PERT

M.SC. (STATISTICS) III-SEMESTER

STS-306: PAPER-V: SECTION-A (ELECTIVE-I) & SECTION-B (ELECTIVE-II)

PRACTICAL-II (CONVENTIONAL & USING SOFTWARE)

ELECTIVE-I (A): LIST OF PRACTICALS APPLIED REGRESSION MODELS

1. Problems on All possible Regression using R^2.
2. Problems on Stage wise Regression.
3. Computation of odds ratio (Dichotomous).
4. Computation of Multiple Logistic regression.
5. Fitting a probit regression line through least squares method.
6. Computation of variance components.
7. Computation of mean and variance for exponential family of distributions.

ELECTIVE-I (B): LIST OF PRACTICALS ON ECONOMETRIC MODELS

1. Use of dummy variables (dummy variable trap) and seasonal adjustment
2. GLS estimation and predictors
3. Tests for heteroscedasticity.
4. Tests for Autocorrelations
5. Instruments variable estimation
6. Estimation with lagged dependent variable
7. Identification problems – Checking rank and order condition
8. Two SLS estimation

ELECTIVE-I (C): LIST OF PRACTICALS ON ADVANCED DESIGNS AND ANALYSIS OF EXPERIMENTS

1. Intra-block analysis of BIBD
2. Analysis of Youden Square Design
3. Intra-block analysis of PBIBD (2)
4. Analysis of Balanced Lattice design
5. Analysis of Simple Lattice design
6. Fitting of Response surface design model for first and second order
7. Construction of rotatable design
8. Analysis of Mixture Experiments.

ELECTIVE-I(D): STATISTICAL PROCESS IN DATA SCIENCE USING PYTHON

Data sets of Kaggle.com can be used for practice. For example, few of the them are: Iris Dataset; flights.csv Dataset; Sustainable Development Data; Credit Card Fraud Detection; Employee dataset; Heart Attack Analysis & Prediction Dataset; Dataset for Facial recognition; Covid_w/wo_Pneumonia Chest Xray Dataset; Groceries dataset; Financial Fraud and Non-Fraud News Classification; IBM Transactions for Anti Money Laundering

List of Practicals in Data Handling with Python (USING Packages):

1. Understanding data with Data types, Measurement of scales, descriptive statistics and data pre-processing steps.
2. Data transformations (Standardize, Normalize, converting data from one scale to other scales).
3. Parametric tests (z-, χ^2, t-, F-tests, ANOVA), Correlation & Regression etc.
4. Non-Parametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, Run test).
5. Applying the modelling process, Model evolution, over fitting, under fitting, cross validation concepts, (train/test, K fold and Leave out one approaches),
6. Evaluation of Model Performance for classification techniques for qualitative and Quantitative data.
7. Drawing One dimensional diagrams (Pictogram, Pie Chart, Bar Chart,).
8. Drawing two-dimensional (Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot)
9. Drawing Gantt Chart, Heat Map, Box - Whisker Plot, Correlation Matrices.

ELECTIVE-II (A): LIST OF PRACTICALS IN DATA MINING

1. Nearest-Neighbor classifiers
2. Bayesian classifiers
3. Support vector machine K-means
4. DBSCAN
5. Compact representation of frequent item sets
6. FP-Growth Algorithm

ELECTIVE-II (B): LIST OF PRACTICALS IN BAYESIAN INFERENCE.

1. Data Simulation for Uniform, Normal, Exponential, Cauchy and Poisson Distributions.
2. Bayesian estimation of parameters for p in Binomial(n,p) with their conjugate paired distributions using Metropolis Hasting and Gibbs Sampler).
3. Bayesian Estimation of parameters and μ in Normal (μ, σ^2) distribution with their conjugate paired distributions (using R) with Metropolis Hasting / Gibbs sampler.

ELECTIVE-II(C): LIST OF PRACTICALS ON ADVANCED MACHINE LEARNING TECHNIQUES

(Implementation using Python)

1. Multiple linear regression and Multiple Logistic regression
2. KNN & K-means
3. Naïve Bayes classier
4. Support vector machines
5. Random forest
6. Bagging and Boosting
7. Implementation of Perceptron Learning Algorithm.
8. Implementation of Multi-layer Perceptron Learning
9. Implementation of Back-Propagation Algorithms.
10. Implementation of Hidden Markov Model,

M.SC. (STATISTICS) III- SEMESTER

STS-307: PAPER VII: STATISTICAL ANALYSIS USING R AND TORA
PRACTICAL-III

Practical with R Package for the following topics.

1. **Data Visualization**: Pie diagram, Bar diagram, Histogram, Line plot, frequency curves & polygons, Scatter Plot, Gantt Chart, Box Plot.
2. **Descriptive Statistics:** Measures of Central Tendencies, Dispersions, Relative measures of Dispersions, Moments, Skewness, Kurtosis.
3. **Parametric Tests**: Testing for Mean(s), Variance(s), Proportion(s), ANOVA for one-way two-way and two way with one and m-observations per cell and with & without interactions,
4. **Non–Parametric tests:** Sign test, Wilxon Sign Rank test, Mann-Whitney U-test, Run test, Kolmogorov Smirnov test, Chi-square test for goodness of fit and Chi-square test independence.
5. **Design & Analysis of Experiments:** Analysis of Variances for Completely randomized, randomized block and latin Square Designs and Factorial experiments (2^2, 2^3 F.E. without confounding).
6. **Regression Analysis:** Analysis of Simple and Multiple Linear Regression models, Selection Best Linear Regression Model (All possible, forward, backward, stepwise and stage wise methods). Binary and multinomial Logistic regression models, Probit analysis.
7. **Multivariate Data Analysis:** Linear Discriminant Analysis, Principal Component analysis, Factor analysis, Multi-dimensional scaling, Cluster analysis.
8. **Statistical Quality Control:** Construction Control charts for variables and attributes.

TORA

Operations Research (TORA Package):

Solving a Linear Programming Problems: Graphical method, simplex method, Big-M method, two Phase method, Duality, Dual simplex, transportation problem, Assignment Problem, sensitivity analysis.

M.SC.(STATISTICS) III- SEMESTER

STS-308: PAPER VII: DATA ANALYSIS MINI PROJECT
PRACTICAL-IV

Objectives and Outcomes:

1. To familiarize tools and techniques and content for presentation
2. To enhance practical presentation, effective communication and professional skills
3. To expose the students to answer the queries raised on the topic of presentation.
4. To encourage students to work with innovative and entrepreneurial ideas
5. Demonstrate the ability to synthesize and apply the knowledge and skills acquired in the academic program to real-world problems
6. Evaluate different solutions based on economic and technical feasibility
7. Effectively plan a project and confidently perform all aspects of project management
8. Demonstrate effective written and oral communication skills

PROJECT GUIDELINES:

1. The Head of Department will appoint Internal supervisor to Guide the students in each group.
2. Each group should consist of Five students.
3. Each student in the group must actively participate and report to the internal supervisor.
4. Each group has to search for the internship from any industry/ institution, if not found they have to choose a project with the help of supervisor allotted such that, the aim of project work is to develop solutions to realistic problems applying the knowledge and skills obtained on the courses studied with specializations, new technologies and current industry practices.
5. Each student has to give minimum two seminars, one in the second week ("Project Design Seminar") another on 8^{th} week (project progress seminar).
6. Submit Title of the project and one page abstract /synopsis about the project in the first week to the Head, forwarded by the internal supervisor.
7. Each project should give a 30 minutes presentation using power point presentation and followed by 10 minutes of discussion.
8. Project seminar presentations should contain, source of the data, Sample data, data description, literature survey on the similar studies, objectives of the study, Methodology, statistical techniques, work plan etc. and details of progress of the work, individual roles and their work distribution and their plan etc.
9. Each group Project Report should follow the Ph.D. thesis norms with Plagiarism report and each group has to submit two copies duly signed by the Students, Supervisor, industry certificate (if exists) and Head of the Department on before the last instruct date of the semester.
10. Project Marks will be awarded based on all stages of the project and the topic chosen, seminar presentation, communication skills, role/ contribution of the student in the project etc and viva-voce conducted by the internal & External examiners.

M.SC. (STATISTICS) SEMESTER-IV

Paper	Sub. Code	Paper Title	Credits	Instruction Hours per Week	Semester end Exam duration	Max. Marks in Semester end Exam	Max. Marks in Internal Assessment and Assignments
THEORY PAPERS							
I	STS-401	Stochastic Processes (SP)	3	3	3	70	20+10
II	STS-402	Time Series Analysis (TSA)	3	3	3	70	20+10
III	STS-403	A) Advanced Operations Research (AOR) B) Text Analytics (TA) C) Demography (DGY)	3	3	3	70	20+10
IV	STS-404	A) Artificial Neural Networks (ANN) B) Design & Analysis of Algorithms (DAA) C) Clinical Trails (CT)	3	3	3	70	20+10
PRACTICAL PAPERS							
V	STS-405	Stochastic Processes & Time Series Analysis	2	4	2	50	-
VI	STS-406	Elective – I & II	2	4	2	50	-
VII	STS-407	Major Project	4	8	2	100	-
Total			20	12+16*	-	600	

M.SC.(STATISTICS) IV-SEMESTER
STS-401: PAPER-I: STOCHASTIC PROCESSES (SP)

UNIT – I

Introduction to stochastic processes; classification of stochastic process according to state-space and time-domain. Finite and countable state Markov chains; time-homogeneity; Chapman-Kolmogorov equations; marginal distribution and finite – dimensional distribution;.

UNIT – II

classification of states of a Markov chain – recurrent, positive recurrent, null - recurrent and transient states. Period of a state. Canonical form of transition probability matrix of a Markov chain. Fundamental matrix; probabilities of absorption from transient states into recurrent classes, in a finite Markov Chain; mean time for absorption. Ergodic state and ergodic chain.

Unit-III

Stationary distribution of a Markov chain. Existence and evaluation of stationary distribution. Random walk and gambler's ruin problem. Weiner process as limit of random walk. First passage time of the process.

UNIT – IV

Discrete state-space, continuous time Markov Processes – Kolmogorov difference - differential equations. Poisson process and its properties. Birth and Death Process, application in queuing. Pure Birth and pure Death processes.

UNIT – V

Renewal process, elementary renewal theorem and its applications. Statement and uses of Key – renewal theorem. Residual life time. Branching process – Galton-Watson branching process, mean and variance of size of n^{th} generation; probability of ultimate extinction of a branching process – fundamental theorem of Branching process – Examples.

REFERENCES

1. Medhi,J. (1982) : Stochastic Processes – Wiley Eastern
2. Karlin, S. and Taylor, H.M. (1975): A First Course in Stochastic Processes, Vol. I, Academic Press.
3. Bhat, B.R. (2000): Stochastic Models: Analysis and applications – New Age International India.
4. Basu, A.K. (2003): Introduction to Stochastic Process, Narosa Publishing House.

M.SC. (STATISTICS) IV-SEMESTER

STS-402: PAPER-II: TIME SERIES ANALYSIS (TSA)

Unit–I

Stationery stochastic processes. The autocovariance and Auto correlation functions and their estimation. Standard errors of autocorrelation estimates. Bartlett's approximation (without proof). The periodogram, the power spectrum and spectral density functions. Link between the sample spectrum and autocorrelation function.

Unit–II

Linear Stationary Models: Two equivalent forms for the general linear process. Autocovariance generating function and spectrum, stationarity and invertibility conditions for a linear process. Autoregressive and moving average processes, autocorrelation function (ACF), partial autocorrelation function (PACF).

Unit-III

Spectrum for AR processes up to 2. Moving average process, stationarity and Invertibility conditions. ACF and PACF for M.A. (q), spectrum for M.A. processes up to order 2. Duality between autoregressive and moving average processes, Mixed AR and MA(ARMA) process. Stationarity and invertibility properties. ACF and spectrum of mixed processes. The ARMA(1.1) process and its properties. Linear Non-Stationary Models – Autoregressive integrated and moving average (ARIMA) processes. The three explicit forms the ARIMA models (viz) Difference equation, random shock and inverted forms.

Unit–IV

Model Identification–Stages in the identification procedures. Use of autocorrelation and partial auto–correlation, functions in identification. Standard errors for estimated autocorrelation and partial autocorrelations. Initial estimates MA, AR and ARMA processes and residual variance.

Model Estimation: Least squares and Maximum likelihood estimation and interval estimation of parameters.

Unit–V

Model Diagnostic checking – checking the stochastic model diagnostic checks applied to residuals. Forecasting: Minimum mean square error forecasts and their properties, derivation of the minimum mean square error forecasts, calculating and updating forecasts at any lead time.

REFERENCES

1. Box and Jenkins: Time Series Analysis
2. Anderson, T.W. : Time Series Analysis.
3. Brockwell,P.J., and Davis,R.A.: Time Series : Theory and Methods (Second Edition). Springer–Verlag.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-I: STS-403(A): PAPER-III(A): ADVANCED OPERATIONS RESEARCH (AOR)

Unit–I

Non-linear Programming problem – Formulation, Generalised Lagrange multiplier technique, Kuhn-Tucker necessary and sufficient conditions for optimality of an NLPP, Wolfe's and Beale's Algorithms for solving QPP. Separable Programming Problem; Piecewise linear Approximation method. Linear Fractional Programming Problem and its applications.

Unit–II

Dynamic Programming, Principle of optimality, solution of LPP by Dynamic Programming technique, Knapsack problem by Dynamic Programming Technique. General goal Programming model and formulation of its objective function. Solutions to linear goal programming and linear integer goal programming.

Unit–III

Game Theory: Two person zero sum game, pure strategies with saddle point, mixed strategies with saddle point, principles of dominance and games without saddle point, 2xm, mx2, mx n games Decision Analysis: Introduction, Steps in Decision theory approach, Types of Decision making environments, Decision making under uncertainty – criterion of optimism, pessimism, equally likely decision criterion, criterion of realism, criterion of regret. Decision tree analysis, Decision making with utilities.

Unit–IV

S-S policy for inventory and its derivation in the case of exponential demand; Models with variable supply and models for perishable Items. Replacement Problems; Introduction, block and age replacement policies, replacement of items with long life. Machine interference problems.

Unit-V

Introduction to simulation, generation of random numbers for Uniform, Normal, Exponential, Cauchy and Poisson Distributions. Estimating the reliability of the random numbers, Simulation to Queuing and Inventory problem.

REFERENCES

1. Taha, H.A.(1982): Operations Research : An Introduction; McMillan
2. Kantiswarup;Gupta P.K. and Singh,M.N.(1985) : Operations Research; Sultan Chand.
3. Sharma,S.D.: Operations Research.
4. Sharma J.K : Operation Research
5. Hillier F.S. and Leiberman,G.J.(1962) : Introduction to Operations Research; Holdon Day.
6. Philips,D.T.,Ravindran,A.and Solberg,J.(2000): Operations Research principles and practice.
7. Taha, H.A.(1982): Operations Research : An Introduction; McMillan
8. Kantiswarup;Gupta P.K. and Singh,M.N.(1985) : Operations Research; Sultan Chand.
9. Sharma,S.D.: Operations Research.

M.SC.(STATISTICS) IV-SEMESTER

ELECTIVE-I: STS-303(B): PAPER-III (B): TEXT ANALYTICS (TA)

Unit - I

Introduction to Natural Language Processing Basic, Language Syntax and Structure (Words, Phrases, Clauses, & Grammar), Language Semantics Processing, (Lexical Semantic Relations, Homonyms, Homographs, and Homophones, Capitonyms, Hyponyms and Hypernyms),Text Corpora (Corpora Annotation and Utilities),Accessing Text Corpora (Brown Corpus, WordNet Corpus) and NLP Applications (Machine Translation, Text Summarization and Text categorization).

Unit – II

Concept of the Tokenization, Sentence Tokenization, Word Tokenization, Concept of the Text Normalization, (Cleaning Text, Removing Special characters, Removing stop words,..etc) correcting words using stemming and Lemmatization and Understanding text syntax and structure.(POS tagging and Parsing)

Unit – III

Concepts of feature extraction, Methods of Feature extraction (Bag of words Model, TF-IDF Models, Advanced word Factorization Models likes Word2vec), Strengths and weakness of models and Word cloud … etc, Concepts of Document term matrix, Term Document Matrix.

Unit – IV

Concepts of Topic Modelling, Algorithms of Topic Modelling (Latent Semantic Indexing (LSI) , Latent Dirichlet Allocation (LDA), Non Negative Matrix Factorization (NMF) and Similarity based text clustering models).

Unit-V

Text Classification using supervised methods (Like Multinomial Naïve Bayes, Support vector machines, Random Forest …), concept of Sentiment Analysis and its applications.

REFERENCES

1) Chapman & Hall : Handbook of Natural Language Processing, Second Edition.
2) CRC: Machine Learning & Pattern Recognition, 2nd Edition.
3) Christopher Manning and Hinrich Schuetze: Foundations of Statistical Natural Language Processing.
4) Dipanjan Sarkar: Text Analytics with Python, A press Publication.
5) Julia Silge: Text Mining with R: A Tidy Approach, 1st Edition.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-I: STS-403(C): PAPER III(C): DEMOGRAPHY (DGY)

Unit–I

Coverage and content errors in demographic data, use of balancing equations and Chandrasekharan, Deming formula to check completeness of registration data.

Unit-II

Adjustment of age data - use of Whipple, Myer and UN indices. Population composition, dependency ratio.

Unit–III

Measures of fertility; stochastic models for reproduction, distributions of time to first birth, inter-live birth intervals and of number of births (for both homogeneous and nonhomogeneous groups of women), estimation of parameters; estimation of parity progression ratios from open birth interval data.

Unit–IV

Measures of Mortality; construction of abridged life tables. Distributions of life table functions and their estimation. Stable and quasi-stable populations, intrinsic growth rate. Models for population growth and their fitting to population data. Stochastic models for population growth.

Unit–V

Stochastic models for migration and for social and occupational mobility based on Markov chains. Estimation of measures of mobility. Methods for population projection. Use of Leslie matrix.

REFERENCES

1. Bartholomew, D. J. (1982). Stochastic Models for Social Processes, John Wiley.
2. Benjamin, B. (1969). Demographic Analysis, George, Allen and Unwin.
3. Chiang, C. L. (1968). Introduction to Stochastic Processes in Biostatistics; John Wiley.
4. Cox, P. R. (1970). Demography, Cambridge University Press.
5. Keyfitz, N. (1977). Applied Mathematical Demography; Springer Verlag.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-II:STS-404(A): PAPER IV(A): ARTIFICIAL NEURAL NETWORKS (ANN)

Unit – I

Basics of Artificial Neural Networks (ANN), Human vs Computers, Organization of the Brain, Biological Activations of Neuron; Artificial Neuron Models: McCulloch-Pitts, Perceptron, Adaline, Hebbian Models; Historical Developments of ANN, Characteristics of ANN, Types of Neuron Activation Function, Signal functions and their properties, monotonicity. ANN Architecture, Classification Taxonomy of ANN, Un-supervised and Reinforcement learning; Learning tasks, Memory, Adaptation, Statistical nature of the learning process. Statistical learning theory**.** Gathering and partitioning of data for ANN and its pre and post processing.

Unit – II

Perceptron Learning Algorithm, Derivation, Perceptron convergence theorem (statement); Multi-layer Perceptron Learning rule, limitations. Applications of the Perceptron learning. Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning. Feed-forward and Fedd-back Back-Propagation Algorithms and derivation, learning rate, Momentum, Difficulties and Improvements. Bias and Variance. Under- Fitting and Over-Fitting

Unit-III

Radial Basis Function Networks: Introduction, Regularization theory, Regularization Networks, Generalized Radial Basis Function Networks, Approximation properties of Radial Basis Function Networks, Comparison with Multi-layer Perceptron, Applications.

Unit-IV

Hebbian Learning, Competitive learning**.** Self Organizing Maps**:** Two basic feature mapping models, Self-Organizing Map, SOM algorithm, properties of feature map, computer simulations, Vector quantization, Learning vector quantization, Hierarchical Vector Quantization,

Unit-V

Boltzman Machine and its learning rule, Hopfield model and its learning. Sigmoid belief network learning procedure, Stochastic machines. Applications of ANN in Classification, Clustering, Regression, Time series forecasting.

REFERENCES

1. Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. New York: Macmillan Publishing. A comprehensive book and contains a great deal of background theory
2. Yagnanarayana, B. (1999): “Artificial Neural Networks” PHI
3. Bart Kosko(1997): Neural Networks and Fuzzy systems, PHI
4. Jacek M. *Zurada(1992): Artificial Neural Systems,* West Publishing Company.
5. Carling, A. (1992). Introducing Neural Networks. Wilmslow, UK: Sigma Press.
6. Fausett, L. (1994). Fundamentals of Neural Networks. New York: Prentice Hall.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-II: STS-404(B): PAPER IV(B): DESIGN AND ANALYSIS OF ALGORITHMS

UNIT I

Introduction to Algorithms: Algorithm, Time & space complexity, Asymptotic Notations. Writing pseudocode, Design Techniques.

Divide and Conquer: Control Abstraction, Binary Search, Finding the Maximum and Minimum, Merge Sort; Quick Sort, Selection sort, Strassen's Matrix Multiplication, Convex Hull.

UNIT-II

Greedy Method: Control Abstraction, Knapsack Problem, Job Sequencing with Deadlines, Minimum-Cost Spanning Trees (Kruskal's & Prim's), Single Source Shortest Paths (Dijkstra's).

Dynamic Programming: Control Abstraction, Multistage Graphs, All-Pairs Shortest Paths, Single-Source Shortest Paths, Optimal Binary Search Trees, 0/1 Knapsack, Traveling Salesperson Problem.

UNIT-III

Basic Traversal and Search Techniques: Techniques for Binary Trees, Techniques for Graphs, Connected Components and Spanning Trees, Biconnected Components and DFS.

Back Tracking: Control Abstraction, , 8-Queens Problem, Sum of Subsets, Graph Colouring, Hamiltonian Cycles, Knapsack Problem.

Branch-Bound: Control Abstraction, 0/1 Knapsack Problem, Traveling Sales Person problem.

UNIT -IV

NP-Hard and NP-Complete Problems: Basic Concepts, Cook's Theorem, NP-Hard. Graph Problems, NP-Hard Scheduling Problems, NP-Hard Code Generation, Some Simplified NP-Hard Problems.

REFERENCE BOOKS

1. E Horowitz, S Sahni, S Rajasekaran, (2007): Fundamentals of Computer Algorithms, 2/e, Universities Press.
2. T.H. Cormen, CE Leiserson, R.L Rivert, C Stein, (2010): Introduction to Algorithms, 3/e, PHI.
3. R. Pannerselvam (2007): Design and Analysis of Algorithms, PHI.
4. Hari Mohan Pandey, (2009): Design, Analysis and Algorithm, University Science Press.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-II: STS-404(C): PAPER IV(C): CLINICAL TRIALS (CT)

Unit–I

Introduction to clinical trials : The need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I-IV trials, multi-center trials. Data management: data definitions, case report forms, database design, data collection systems for good clinical practice.

Unit-II

Determination of sample size: for two independent samples of Dichotomous Response variables, for two independent samples of Continuous Response variables and for repeated variables.

Unit–III

Design of clinical trials : parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, review of 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, design of bioequivalence trials.

Unit–IV

Reporting and analysis: analysis of categorical outcomes from Phase I - III trials, analysis of survival data from clinical trials.

Unit–V

Surrogate endpoints: selection and design of trials with surrogate endpoints, analysis of surrogate endpoint data. (2L) Meta-analysis of clinical trials.

REFERENCES

1. S. Piantadosi (1997). Clinical Trials : A Methodological Perspective. Wiley and Sons.
2. C. Jennison and B. W. Turnbull (1999). Group Sequential Methods with Applications to Clinical Trials, CRC Press.
3. L. M. Friedman, C. Furburg, D. L. Demets (1998). Fundamentals of Clinical Trials, Springer Verlag.
4. J. L. Fleiss (1989). The Design and Analysis of Clinical Experiments. Wiley and Sons.
5. E. Marubeni and M. G. Valsecchi (1994). Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley and Sons.

M.SC.(STATISTICS) SEMESTER IV

STS-405: PAPER-V: STOCHASTIC PROCESSES & TIME SERIES ANALYSIS
PRACTICAL-I (CONVENTIONAL)

SECTION-A: LIST OF PRACTICALS ON STOCHASTIC PROCESSES

1. Formulation of problems as Markov chain models
2. Computation of finite dimensional and marginal distributions; higher dimensional transition probabilities.
3. Classification of states, identification of recurrent classes and reduction to canonical form of t.p.m.
4. Probabilities of absorption into recurrent classes (from transient states)
5. Computation of stationary distribution (unique case)
6. Computation of stationary distribution (non-unique case)
7. M|M|1 queue – operating characteristics
8. Mean and variance of n^{th} generation size and probability of extinction of Branching processes.

LIST OF PRACTICALS ON TIME SERIES ANALYSIS

1. Generation of Time series by means of simple time series models
2. Sample and theoretical correlograms
3. Periodogram analysis
4. Writing the models in B notation and stationarity and invertibility of the models
5. Classification of ARIMA models and computation of weights
6. Identification AR, MA, ARMA models
7. Estimation of parameters in AR, MA and ARMA models
8. Computation of forecasts, updating and probability limits for forecasts

M.SC.(STATISTICS) SEMESTER IV

STS-406: PAPER VI: ELECTIVE-I & ELECTIVE-II

PRACTICAL-II (CONVENTIONAL & WITH SOFTWARE)

ELECTIVE-I	ELECTIVE-II
A) Advanced Operations Research (AOR) B) Text Analytics (TA) C) Demography (DGY)	A) Artificial Neural Networks (ANN) B) Design & Analysis of Algorithms (DAA) C) Clinical Trails (CT)

ELECTIVE-I (A): ADVANCED OPERATIONS RESEARCH

1. Wolfe and Beale's methods for QPP
2. Separable Programming problem
3. Dynamic Programming Problem
4. Goal Programming Problem
5. Problems on Decision under uncertainty
6. Replacement Problem

ELECTIVE-I(B): TEXT ANALYTICS (TA)

1. Perform data collection by web scrapping with python and Perform following tasks (i) Find the URL that you want to scrape (ii) Inspecting the Page (iii) Find the data you want to extract (iv) Write the code (v) Run the code and extract the data (vi) Store the data in the required format.
2. Perform following Data Pre-processing tasks in Python using Scikit-learn. standardization, normalization, encoding, discretization, imputation of missing values. Use your own dataset to perform all pre-processing tasks as suggested in given reference.
   (i) https://www.analyticsvidhya.com/blog/2016/07/practical-guide-datapreprocessing-python-scikit-learn/
   (ii) https://scikit-learn.org/stable/modules/preprocessing.html
3. Answer the following question in your blog (As per dataset taken by you): Dataset Description: Task to be performed: How to decide variance threshold in data reduction? Code Snapshot, Output Snapshot, Task-2, Code Snapshot, Output Snapshot.
   Perform following Data Pre-processing tasks using python
   Data reduction using variance threshold, univariate feature selection,
   recursive feature elimination, PCA, correlation
   Reference:
   1. https://medium.com/analytics-vidhya/feature-selection-using-scikit-learn5b4362e0c19b
   2. https://machinelearningmastery.com/rfe-feature-selection-in-python/
   3. https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60
   4. https://towardsdatascience.com/feature-selection-using-python-for-classificationproblem-b5f00a1c7028
   5. https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/
   Answer the following question in your blog (As per dataset taken by you):
   Dataset Description; Task to be performed; Why feature selection is important?? Its advantages/disadvantages. Code Snapshot; Output Snapshot;What is the impact on accuracy, with or without data reduction? Code Snapshot; Output Snapshot.
   Amongst all methods, which method avoids overfitting and improves model performance

ELECTIVE– I(C): DEMOGRAPHY

1. Construction of Abridged life tables
2. Fitting of population growth models
3. Estimation of population projection
4. Estimation of Life table functions

Elective-II (A): Artificial Neural Networks (ANN)

1. Implementation of Perceptron Learning Algorithm.
2. Implementation of Multi-layer Perceptron Learning
3. Implementation of Gradient Descent Learning,
4. Implementation of Least Mean Square learning,
5. Implementation of Widrow-Hoff Learning.
6. Implementation of Back-Propagation Algorithms.
7. Implementation of Markov Decision Process,
8. Implementation of Hidden Markov Model,

Elective-II(B): DESIGN & ANALYSIS OF ALGORITHMS (Using Python)

1. Write a program for sorting the given list using: Merge Sort, Quick Sort, Heap Sort.
2. Write a program to find the given number in a list using Binary Search.
3. Write a program to find the minimal spanning tree using Kruskal's and Prims Algorithms.
4. Write a program to find the shortest path using Dijkstra's Algorithm.
5. Write a program to solve using dynamic programming technique for Travelling sales man problem. Multistage Graph problem, Optimal Binary Search Trees.
6. Write a program to solve Knapsack problem using Back tracking

ELECTVE-II (C): CLINICAL TRIALS

1. Determination of Sample size
2. Multiple Logistic Regression with two or Three variables
3. Analysis of Clinical trial data using Cross over design
4. Analysis of Clinical trial data using Parallel design
5. Meta-analysis of Clinical trials
6. Analysis of Clinical data using Factorial Experiments

M.SC. (STATISTICS) SEMESTER-IV

STS-407: PAPER VII: MAJOR PROJECT

Note: Follow the guidelines of the project specified in STS-308: Mini project.

****

DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD – 500 007

M.Sc. APPLIED STATISTICS III-SEMESTER

Paper	Sub. Code	Paper Title	Credits	Instruction Hours per Week	Semester end Exam duration	Max. Marks in Semester end Exam	Max. Marks in the Internal Assessment and Assignments
THEORY PAPERS							
I	STAS-30I	Operations Research (OR)	3	3	3	70	20+10
II	STAS-302	Forecasting Models (FM)	3	3	3	70	20+10
III	STAS-303 Elective – I	A. Reliability Theory (RT) B. Actuarial Science (ASC) C. Econometric Models (EM) D. Statistical Process for Data Science (SPDS)	3	3	3	70	20+10
IV	STAS-304 Elective – I	A. Data Mining Techniques (DM) B. Bayesian Inference (BI) C. Statistical Pattern Recognition **(SPR)** D. Advanced Machine Learning Techniques (AMLT)	3	3	3	70	20+10
PRACTICAL PAPERS							
V	STAS-305	Operations Research & Forecasting Models (OR + FM)	4	2	2	50	-
VI	STAS-306	Elective-I & II	4	2	2	50	-
VII	STAS-307	Statistical Analysis using R & TORA	4	2	2	50	-
VIII	STAS-308	Data Analysis Project (Mini project)	4	2	2	50	-
Semester Total			34	***	***	480	120

M.SC. (APPLIED STATISTICS) III-SEMESTER

STAS-301: PAPER-I : OPERATIONS RESEARCH (OR)

Unit–I

General Linear programming problem, Graphical method, Simplex method, Big-M and two phase methods, Concept of Duality, Primal-Dual relation; Dual Simplex algorithm;

Unit–II

Sensitivity Analysis: Introduction, definition of sensitivity analysis; discrete changes in requirement and cost vectors. Parametric Programming: Introduction, parameterization of cost and requirement vectors. Integer Programming Problem: Gomory's cutting plane algorithm for pure and mixed IPP; Branch and bound Technique. Fractional and O-1 Knapsack problems, Sequencing Problems: 2 machine n-job and 3 machine n-job problems;

Unit-III

Queuing Theory: Introduction, essential features of Queuing system, Operating characteristics of Queuing system (transient and steady states).Queue length, General relationships among characteristics. Probability distribution in queuing systems, distribution of Arrival and interarrival. Distribution of death (departure) process, service time. Classification of Queuing models and solution of Queuing models; M/M/1:∞/FIFO and M/M/1:N/FIFO

Unit–IV

Inventory: Analytical structure of inventory problems; ABC analysis; EOQ problem with and without shortages with (a) production is instantaneous (b) Finite constant rate (c) shortages permitted random models where the demand follows uniform distribution. Multi-item inventory subject to constraints.

Unit–V

Networks: Basic concepts constraints in networks, construction of networks. Time calculation in Networks. PERT, CPM, Network problems.

REFERENCES

1. Kantiswarup; Gupta P.K. and Singh,M.N.(1985):Operations Research; Sultan Chand
2. Sharma,S.D.: Operations Research
3. Taha, H.A.(1982): Operations Research: An Introduction; MacMillan
4. Gillet.: Introduction to O. R.

M.SC. (APPLIED STATISTICS) III-SEMESTER

STAS-302: PAPER-II: FORECASTING MODELS (FM)

Unit–I

Forecasting: The role of forecasting in decision-making, forecasting techniques. Smoothing Techniques: Simple Moving Averages, exponential smoothing and Winter's linear and seasonal exponential smoothing. Stationary stochastic processes, Autocovariance and Autocorrelation functions and their estimation. Standard error of autocorrelation estimates. Bartlett's approximation (without proof).

Unit–II

Periodgram, power spectrum and spectral density functions. Simple examples of autocorrelation and spectral density functions. Link between sample spectrum and auto-correlation function.

Linear Stationary Models: Two equivalent forms for the general linear process. Autocovariance generating function and spectrum. Stationarity and invertibility conditions for a linear process.

Unit-III

Autoregressive and moving average processes, autocorrelation function (ACF), partial autocorrelation function (PACF). Spectrum for AR processes up to 2. Moving average process, stationarity and invertibility conditions. ACF and PACF for M.A.(q) spectrum for M.A. processes up to order 2, Duality between autoregressive and moving average processes. Mixed AR and MA (ARMA) process. Stationarity and invertibility properties, ACF and spectrum of mixed processes. The ARMA(1,1) process and its properties.

Unit–IV

Linear Non-Stationary Models–Autoregressive integrated and moving average (ARIMA) processes. The three explicit forms for the ARIMA models viz., difference equation, random shock and inverted forms.

Model Identification: Stages in the identification procedures, use of autocorrelation and partial auto–correlation functions in identification. Standard errors for estimated auto correlations and partial autocorrelations. Initial estimates of parameters of MA, AR and ARMA processes and residual variance.

Model estimation: Least squares and Maximum likelihood estimation and interval estimation of parameters.

Unit–V

Model diagnostic checking–Checking the stochastic model. Diagnostic checks applied to residuals.

Forecasting-minimum: Mean square error forecasts and their properties, derivation of the minimum mean square error forecasts, calculating and updating forecasts, probability limits of the forecasts at any lead time.

REFERENCES

1. Weel Wright, S.C. and Makridakis,S. (1973): Forecasting methods for Management, John–Wiley & sons, New York.
2. Box, G.E.P. and Jankins,G.M.(1970) : Time series Analysis (Forecasting and control), Holden day publication.
3. Anderson, T.W.(1971) : The statistical analysis of Time series, John Wiley, New York.
4. Brockwell,P.J. and Davis, R.A.: Time Series : Theory and methods(Second Edition), Springer-Verlag.

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE-I(A) : STAS-303(A): PAPER III (A): RELIABILITY THEORY (RT)

Unit–I

Coherent Systems: Reliability concepts – Systems of components. Series and parallel systems – Coherent structures and their representation in terms of paths and cuts, Modular decomposition.

Unit–II

Reliability of coherent systems – Reliability of Independent components, association of random variables, bounds on systems reliability and improved bounds on system reliability under modular decomposition.

Unit–III

Life Distribution: Survival function – Notion of aging IFR, DFR, DFRA, NBU and NBUE classes, Exponential distributions and its no-ageing property, ageing properties of other common life distribution, closures under formation of coherent structures, convolutions and mixtures of these cases.

Unit–IV

Maintenance and replacement policies, relevant renewal theory, availability theory, maintenance through spares and repair.

Unit-V

Reliability estimation: Estimation of two and three parameter Gamma, Weibull and log normal distributions.

REFERENCES

1. Barlow, R.E. and Proschen, F. (1975): Statistical Theory of Reliability and life testing. Halt, Reinhart and Winston Inc.
2. Barlow and Proschen (1965): Mathematical Theory of Reliability, John Wiley
3. Balaguru Swamy – Reliability Engineering
4. L.J. Bain: Statistical analysis of Reliability and like testing Marcel Decker.
5. Sinha, S.K., and Kale, S.K., (1980): Life testing and Reliability estimation, Wiley Eastern.

M.SC. (APPLIED STATISTICS) III- SEMESTER IV

ELECTIVE I(B) : STAS-303(B): PAPER III (B): ACTUARIAL SCIENCE (ASC)

Unit–I

Economics of Insurance - Utility theory, insurance and utility theory, models for individual claims and their sums, survival function, curate future lifetime, force of mortality.

Unit–II

Life table and its relation with survival function examples, assumptions of fractional ages, some analytical laws of mortality, select and ultimate tables**.** Types of Life insurance products – Term insurance, Whole-life insurance, Endowment insurance and Annuities. Measurement of risk in life insurance and fundamental principles underlying rate-making.

Unit–III

Elements of compound interest: Nominal and effective rates of interest, discount, accumulation factor and continuous compounding. Multiple life functions, joint life and last survivor status, insurance and annuity benefits through multiple life functions, evaluation for special mortality laws.

Unit-IV

Multiple decrement models, deterministic and random survivorship groups, associated single decrement tables, central rates of multiple decrement, net single premiums and their numerical evaluations. Distribution of aggregate claims, compound Poisson distribution and its applications.

Unit–V

Net premiums: Continuous and discrete premiums, true monthly payment premiums, apportionable premiums, commutation functions, and accumulation type benefits.

Net premium reserves: continuous and discrete net premium reserve, reserves on a semi continuous basis, reserves based on true monthly premiums, reserves on an apportionable or discounted continuous basis reserves at fractional durations.

REFERENCES

1. N. L. Bowers, H. U. Gerber, J. C. Hickman, D. A. Jones and C. J. Nesbitt (1986): Actuarial Mathematics, Society of Actuaries, Ithaca, Illinois, USA .
2. S. S. Huebner and J. R. Kenneth Black (1976) : Life Insurance, Ninth Ed., PHI Pvt. Ltd.
3. S. P. Dixit, C. S. Modi and R. V. Joshi (2000): Mathematical Basis of Life Insurance, Indian Institute of India.
4. Neill, A.(1977): Life contingencies, Heinemann.
5. Spurgeon E.T.(1972): Life contingencies, Cambridge University Press
6. Benjamin, B and Pollard, J. H. (1980): Analysis of Mortality and other Actuarial Statistics.

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE I(C) : STAS-303(C): PAPER-III (C): ECONOMETRIC MODELS (EM)

Unit–I

Meaning and scope of econometrics. Concepts of dummy variables and proxy variable. Problems and methods of estimation in single equation regression Models

Multicollinearity: Consequences of multicollinearity, tests to detect its presence and solutions to the problem of multicollinearity.

Unit–II

Generalised Least Squares: Estimates of regression parameters – Properties of these estimates.

Heteroscedasticity: Consequences of heteroscedastic disturbances – test to detect its presence and solutions to the problem of heteroscedasticity.

Unit–III

Auto Correlation: Consequences of autocorrelated disturbances, Durbin – Watson test – Estimation of autocorrelation coefficient (for a first order autoregressive scheme).

Distributed lag models: study of simple finite lag distribution models – Estimation of the coefficients of Kayak geometric lag model.

Instrumental Variable: Definition – derivation of instrument variable estimates and their properties.

Unit–IV

Errors in variables: Problem of errors in variables simple solutions using instrumental variables technique. Simulation equation models and methods of estimation: distinction between structure and Model–Exogenous and Endogenous variables – Reduced form of a model.

Unit–V

Problem of identification – Rank and order conditions and their application.

Methods of estimation: Indirect least squares. Two stages least squares, three stages least squares. A study of merits and demerits of these methods.

REFERENCES

1) Johnston – Econometrics Methods (2nd Edition) :
2) G. S. Maddala – Econometrics
3) A. Koutsoyiennis – Theory of econometrics

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE I(D): STAS-303 (D): PAPER-III (D) : STATISTICAL PROCESS IN DATA SCIENCE (SPDS)

Unit – I

Data visualization: Data types, Measurement of scales, understanding data with descriptive statistics. Data visualization techniques: Pictogram, Pie Chart, Bar Chart, Histogram, Line plot, frequency curves& polygons, ogive curves, Scatter Plot, Gantt Chart, Heat Map, Box and Whisker Plot, Waterfall Chart, Area Chart, Stacked Bar Charts - Sub Plots – Matplotlib, Seaborn Styles, Box plot - Density Plot - Tree map - Graph Networks. Visual Perception and Cognition, Applications of Principles of Information Visualization, Dashboard Design.

Unit-II

Data Pre-processing: Understanding data with Descriptive statistics. Data pre-processing steps, Data transformations (Standardize, Normalize, converting data from one scale to other scales). Identification suitable basic statistical tools / tests Parametric tests (z-, χ^2, t-, F-tests), Nonparametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, K-S, Wald-Wolfowitz run test) for the data sets. Feature selection methods

Unit-III

Introduction to Data Modelling: Review of the modelling process, Concepts of Classification & Clustering, Supervised and Un-supervised Modelling, Concepts of Model evolution, Cross validation concepts, (train/test, K fold and Leave out one approaches), Model Performance evaluation for Qualitative and Quantitative data, Model improvement and saving models for future use (classification matrix, Precision and Recall, F1 score, Sensitivity, Specificity, ROC curve) and Model performance concepts for regression (MSE, RMSE, R^2, adj R^2, MAPE),

Unit-IV

Concepts of Model improvement (Tuning parameters using manual search, Manual grid search, random search) and saving models for future use. Simple linear regression and its analysis (model fitting, regression ANOVA, testing lack of fit, MSE, RMSE, R^2, adj R^2, testing regression coefficients and confidence limits).

Unit-V

Basic concepts on Multivariate data; Simple, Partial & Multiple correlations; Multi collinearity; Multiple linear regression and its analysis; Selection of best linear regression (over fitting & under fitting) & its methods in outline (all possible, forward, backward, step-wise and stagewise). Simple and Multiple Logistic models fitting and its analysis.

REFERENCES

1) Foster Provost & Tom Fawcett, Data science for Business, O'REILLY Publications
2) Henrik Brink, Joseph W. Richards. Mark Fetherolf, Real World Machine Learning, Manning Publications.
3) Foster Provost & Tom Fawcett, Data science for Business, O'REILLY Publications
4) Henrik Brink, Joseph W. Richards. Mark Fetherolf, Real World Machine Learning, Manning Publications
5) Brett Lantz , Machine Learning with R, Packt Publications

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE-II (A): STAS-304(A): PAPER-IV(A) : DATA MINING (DM)

Unit-I

Introduction: Challenges, Origins of Data Mining, Data Mining Tasks; **Data:** Types of Data, Data Quality, Data Preprocessing, Measures of Similarity and Dissimilarity; **Exploring Data:** Visualization, OLAP and Multidimensional Data Analysis

Unit-II

Classification: Preliminaries, General approach to solving a classification problem, Decision tree induction, Model Over-fitting, – Evaluating the performance of a classifier – Methods of comparing classifiers; Rule-based classifier, Nearest-Neighbor classifiers, Bayesian classifiers

Unit-III

Classification: Artificial Neutral Networks, Perceptron classifier, Support vector machine, Ensemble methods, Class imbalance problem – Multiclass problem

Unit-IV

Cluster Analysis: Agglomerative hierarchical clustering, K-means, DBSCAN, C4.5, CART Cluster evaluation.

Unit-IV

Association Analysis: Problem definition, Frequent item set generation, Rule generation, Compact representation of frequent item sets, Alternative methods for generating frequent item sets, FP-Growth Algorithm.

Unit-V

Evaluation of Association patterns, Effect of Skewed support distribution; Handling categorical attributes. Handling continuous attributes, Handling a concept hierarchy.

REFERENCES

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2008): “Introduction to Data Mining”, Pearson Education.
2. Arun K Pujari, Data Mining Techniques, University Press, 2nd Edn, 2009.
3. K.P. Soman, Shyam Diwakar, V.Ajay, Insight into Data Mining Theory and Practice, PHI, 2010.
4. Vikram pudi P. Radha Krishna, Data Mining, Oxford University Press, 1st Edition 2009
5. Galit S, Nitin RP, Peter C Bruce. Data Mining for Business Intelligence. Wiley India Edition,2007.

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE-II (B): STAS-304(B): PAPER IV(B): BAYESIAN INFERENCE (BI)

Unit-I

Bayes theorem, Bayesian Concept to priors and posteriors, computation of the posterior distribution.) subjective prior distribution, Conjugate family of priors of a parameter. Hyper parameters of a prior from conjugate family. Conjugate families for (i) exponential family models, (ii) models admitting sufficient statistics of fixed dimension.

Unit-II

Subjective prior distribution of a parameter. Non informative, improper and invariant priors. Jeffrey's invariant prior. Bayesian point estimation as a prediction problem from posterior distribution. Bayes estimators for (i) absolute error loss (ii) squared error loss (iii) 0 - 1 loss. Generalization to convex loss functions.

Unit-III

Evaluation of the estimate in terms of the posterior risk. Bayesian interval estimation: Credible intervals. Highest posterior density regions. Interpretation of the confidence coefficient of an interval and its comparison with the interpretation of the confidence coefficient for a classical confidence interval.

Unit-IV

Bayesian testing of Hypothesis: Specification of the appropriate form of the prior Distribution for a Bayesian testing of hypothesis problem. Prior odds, Posterior odds, Bayes factor for various types of testing hypothesis problems depending upon whether the null hypothesis and the alternative hypothesis are simple or composite. Bayesian prediction problem. Large sample approximations for the posterior distribution.

UNIT-V

Estimation of parameters using Markov Chain Monte Carlo methods: Gibbs Sampler and Metropolis-Hasting Method and other computer simulation methods. Bayesian calculations for non-conjugate priors: (i) Importance sampling, (ii) Obtaining a large sample of parameter values from the posterior distribution using Acceptance - Rejection methods.

REFERENCES

1. Berger, J. O. Statistical Decision Theory and Bayesian Analysis, Springer Verlag.
2. Robert C. P. and Casella, G. Monte Carlo Statistical Methods, Springer Verlag.
3. Leonard T. and Hsu, J. S. J. Bayesian Methods. Cambridge University Press.
4. Box, G. P. and Tiao, G. C. Bayesian Inference in Statistical Analysis, Addison - Wesley.

M.SC.(APPLIED STATISTICS) III-SEMESTER

ELECTIVE-II (C): STAS-304 (C): PAPER-IV (C) : STATISTICAL PATTERN RECOGNITION (SPR)

Unit - I

Basic Concepts to Statistical Pattern Recognition, Pattern Recognition System, Fundamental problems in Pattern Recognition. Linear classifiers: Linear Discriminant Function (for binary outputs) with minimum squared error; Linear Discriminant function (for the normal density), Error bounds for Normal density. Statistical Decision Theory: Introduction, Bayes theorem, Bayes Decision Theory (continuous and discrete features), Bayes Classifier. Simple problems.

Unit – II

Probability of errors: Two classes, Normal distribution, equal covariance matrix assumptions, Chernoff bounds and Bhattacharya distance. Nearest Neighbour Decision rules: Nearest Neighbor Algorithm for classification, K-Nearest Neighbor Estimation. Variants of the Nearest Neighbor Algorithm, description convergence, finite sample considerations. Estimation of probability of error in case Nearest Neighbour and Bayes classifiers. Minimum Error Rate Classifier, Estimation of Probabilities. Comparison of Nearest Neighbour with the Bayes Classifier. Simple problems

Unit – III

Hidden Markov Model and its use for pattern recognition. Branch and Bound Technique for the use of classification. Neural Networks: Perception linear classifier. Support Vector Machines: construction of Support Vectors, Support Vector Machines algorithm for Classification. Simple problems. Combination of Classifiers: Introduction, Methods for Constructing Ensembles of Classifiers, Methods for Combining Classifiers.

Unit – IV

Feature selection and extraction: Feature extraction and Feature selection techniques Inter and intra class distance measures, Probabilistic distance measures, Principal Components Analysis for variable selection and dimensionality reduction.

Unit-V

An Application-Hand Written Digit Recognition: Description of the Digit Data, Preprocessing of Data, Classification Algorithms, Selection of Representative Patterns, Results.

REFERENCES

1. R.O. Duda & H.E. Hart(1978): Pattern Recognition and scene analysis, Wiley
2. Earl Gose, Richard Johnson Baugh and Steve Jost (2009): Pattern Recognition and Image Analysis, PHI.
3. Pattern Recognition: An Algorithmic Approach: Murty, M. Narasimha, Devi, V. Susheela, Spinger Pub,1st Ed.
4. Duda, Hast & Strok: Pattern Recognition.
5. Pattern Recognition: An Algorithmic Approach: Murty, M. Narasimha, Devi, V. Susheela, Spinger Pub,1st Ed.

M.SC.(APPLIED STATISTICS) III-SEMESTER

ELECTIVE-II(D) : STAS-304(D): PAPER-IV(D) : ADVANCED MACHINE LEARNING TECHNIQUES (AMLT)

Unit – I

Basic Concepts to Statistical Pattern Recognition, Pattern Recognition System, Fundamental problems in Pattern Recognition. Linear classifies, Multiple Linear regression, Logistic regression, Linear Discriminant Function (for binary outputs) with minimum squared error, Naïve Bayes classifier, Support Vector Machines, KNN algorithm

Unit – II

Decision Tree algorithms, Random Forest algorithm, Bagging Gradient boosting, Ada-Boosting and XG-Boosting algorithm, Market-Basket Analysis.

Unit – III

Cluster Analysis: Introduction, similarities and dissimilarities, Hierarchical clustering, Single linkage method, k-means and k-Nearest Neighbourhood (KNN) clustering,

Unit – IV

Introduction to Artificial Neuron Networks and its characteristics; Algorithms of Perceptron Learning; Multi-layer Perceptron Learning, Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning. Back-Propagation and their applications.

UNIT – V

Reinforcement learning, Markov Decision Process, Hidden Markov Model, Convolutional Neural Networks, Recurrent Neural Networks, Long-Short Term Memory Networks.

REFERENCES

1. Shai Shalev-Shwartz, Shai Ben-David Understanding Machine Learning: From Theory to Algorithms, Cambridge University press.
2. Marc Peter Deisenroth, A Aldo faisal, Cheng soon Ong: “Mathematics for Machine Learning”, Cambridge University Press, First Edition.
3. Hayes: Artificial Neural networks
4. Henrik Brink, Joseph W. Richards. Mark Fetherolf, Real World Machine Learning, Manning Publications
5. Trevor Hastie & Robert Tibshirani, An introduction to statistical learning with R, Springer Publications
6. Brett Lantz , Machine Learning with R, Packt Publications

M.SC. (APPLIED STATISTICS) III-SEMESTER

STAS-305 : PAPER V – OPERATIONS RESEARCH AND FORECASTING METHODS

PRACTICAL-I

SECTION-A: OPERATIONS RESEARCH

1. Solving an LPP by Simplex Method
2. Solving an LPP by Big M Method
3. Solving an LPP by Two Phase Method
4. Solving an LPP by Revised Simplex
5. Solving an LPP by using its Duality.
6. Solving an LPP by Dual Simplex Method
7. Sensitivity Analysis for cost and requirement vectors.
8. Parametric Programming for cost and requirement vectors.
9. Sequencing problem with 2 machines n- jobs and 3 machines n- jobs
10. Integer Programming Problem- Gomery's cutting plane method.
11. Evaluation of project time through CPM and PERT
12. Time cost Analysis for CPM and PERT

SECTION-B: FORECASTING MODELS

1. Moving Averages and exponential smoothing.
2. Generation of Time series by means of simple time series models.
3. Sample and theoretical correlograms.
4. Periodogram analysis.
5. Writing the models in B notation and stationarity and invertability of the models.
6. Classification of ARIMA models and computation of weights.
7. Identification AR, MA and ARMA models.
8. Estimation of parameters in AR, MA and ARMA models.
9. Computation of forecasts, updating and probability limits for forecasts.

M.SC. (APPLIED STATISTICS) SEMESTER III

STAS-306 : PAPER-VI: SECTION-A (ELECTIVE-I) & SECTION-B (ELECTIVE-II)

PRACTICAL-II

ELECTIVE-I(A): RELIABILITY THEORY

1. Finding Minimal path sets and Minimal cut sets and their representations.
2. Computation of System reliability – parallel, Series and k out of n system.
3. Computations of reliability of Structures when components are independent.
4. Computation of estimated reliability and hazard rates.
5. Computation of bounds on systems reliability.
6. Graphing the reliability function of the systems when the life times of components are exponentially distributed.

ELECTIVE-I(B): ACTUARIAL SCIENCE

1. Computation of values of utility function.
2. Computation of various components of life tables.
3. Computation of compound interest (nominal and effective rate of interests).
4. Annuities and annuity dues.
5. Computation of premium for Term insurance and Whole life insurance.
6. Computation of premium for Endowment insurance.
7. Construction of multiple decrement table for deterministic survival group.
8. Determination of distribution function, survival function and force of mortality.
9. Construction of multiple decrement table for random survivorship group.
10. Construction of select, ultimate and aggregate mortality.
11. Calculation of p.d.f. and distribution function of aggregate claims.
12. Computation of discrete and continuous net premiums.
13. Assurances payable at the moment of death.

ELECTIVE – I(C): ECONOMETRIC MODELS

1. Use of dummy variables (dummy variable trap) and seasonal adjustment
2. GLS estimation and predictors
3. Tests for heteroscedasticity.
4. Tests for Autocorrelations
5. Instruments variable estimation
6. Estimation with lagged dependent variable
7. Identification problems – Checking rank and order condition
8. Two SLS estimation.

ELECTIVE-I(D): STATISTICAL PROCESS FOR DATA SCIENCE

Data sets of Kaggle.com can be used for practice. For example, few of the them are: Iris Dataset; flights.csv Dataset; Sustainable Development Data; Credit Card Fraud Detection; Employee dataset; Heart Attack Analysis & Prediction Dataset; Dataset for Facial recognition; Covid_w/wo_Pneumonia Chest Xray Dataset; Groceries dataset; Financial Fraud and Non-Fraud News Classification; IBM Transactions for Anti Money Laundering
List of Practicals in Data Handling with Python:

1. Understanding data with Data types, Measurement of scales, descriptive statistics and data pre-processing steps.
2. Data transformations (Standardize, Normalize, converting data from one scale to other scales).
3. Parametric tests (z-, χ^2, t-, F-tests, ANOVA), Correlation & Regression etc.
4. Non-Parametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, Run test).
5. Applying the modelling process, Model evolution, over fitting, under fitting, cross validation concepts, (train/test, K fold and Leave out one approaches),
6. Evaluation of Model Performance for classification techniques for qualitative and Quantitative data.
7. Drawing One dimensional diagrams (Pictogram, Pie Chart, Bar Chart,).
8. Drawing two-dimensional (Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot)
9. Drawing Gantt Chart, Heat Map, Box - Whisker Plot, Correlation Matrices.

ELECTIVE–II(A): DATA MINING

1. Nearest-Neighbor classifiers
2. Bayesian classifiers
3. Support vector machine K-means.
4. DBSCAN
5. Compact representation of frequent item sets.
6. FP-Growth Algorithm.

ELECTIVE–II(B): BAYESIAN INFERENCE

1. Simulation of the data based on the sample drawn from Normal, Binomial, Poisson, Beta, Exponential, Gamma.
2. Estimation of parameters in Bayesian approach based on the sample using Gibbs sampler under various priors and posterirs (Normal, Binomial, Poisson, Beta with conjugate priors)
3. Estimation of parameters in Bayesian approach based on the sample using Metro Pollis hasting under various priors and posterirs (Normal, Binomial, Poisson, Beta)

ELECTIVE-II(C): STATISTICAL PATTERN RECOGNITION

1. Computation of Linear discriminant classifier function for two-multivariate normal classes.
2. Computation of Linear discriminant classifier function using Minimum Squared function for Binary data.

3. Bayes classifier and the computation of its Error rate (Probability of Error).
4. Nearest Neighbor Classifier and computation of its error rate.
5. Classification using Hidden Markov Model.
6. Feature selection using Principal Component Analysis.

ELECTIVE-II(D): ADVANCED MACHINE LEARNING TECHNIQUES

1. Implement and demonstrate the use of set of training data samples. Read the training data from a .CSV file.
2. Write a program to demonstrate the working of the decision tree-based ID-3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
3. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
4. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
5. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.
6. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.
7. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.
8. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.
9. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Note: Consider the data sets in https://www.kaggle.com/datasets

M.SC. (APPLIED STATISTICS) III-SEMESTER

STAS-307: PAPER-VII: DATA ANALYSIS USING R-SOFTWARE PRACTICAL-III

Practical with R-software using proper data sets available in https://www.kaggle.com/datasets

1. **Data visualization**: Pie Chart, Bar Chart, Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot, Gantt Chart, Heat Map, Box and Whisker Plot, Waterfall Chart, Area Chart, Box plot, Density Plot, Tree map, Graph Networks.
2. **Descriptive Statistics:** Measures of Central Tendencies, Dispersions, Relative measures of Dispersions, Percentile, Deciles, Moments, Skewness, Kurtosis. (Grouped and ungrouped data sets)
3. **Design & Analysis of Experiments:** Analysis of variances for CRD, RBD, LSD, Factorial experiments (2^2, 2^3, 2^4, 2^5 FE with and without confounding), BIBD, PBIBD.
4. **Multivariate Analysis:** estimation of Mean vector and covariance matrix, Hotelling T^2, Mahanobis D^2, linear Discriminant analysis, Principal Component analysis, Factor analysis, multi-dimensional scaling, Cluster analysis (Agglomerative and k-means),
5. **Time Series Analysis:** Computation of auto covariance, auto correlations, Smoothening, model fitting, AR, MA, ARMA and ARIMA models and their analysis
6. **Parametric tests**: Testing means, variances, proportions for single, two and more than two sample cases; Testing significance of correlation coefficient and regression coefficients,
7. **Non–Parametric tests:** Sign test, Wilxon Sign Rank test, Mann-Whitney U-test, Run test, Kolmogorov Smirnov test, chisquare test for goodness of fit and independence, Kendal tau,
8. **Regression Analysis:** Fitting simple and multiple linear regression models, Analysis, model Lack of fit, R^2, Adj R^2, selection best linear regression using all possible, forward, backward, stepwise and stage wise methods. Simple and multiple logistic regression models, Probit analysis.

M.SC. (APPLIED STATISTICS) SEMESTER III

STAS-308: PAPER-VIII: DATA ANALYSIS MINI PROJECT

Objectives and Outcomes:

1. To familiarize tools and techniques and content for presentation
2. To enhance practical presentation, effective communication and professional skills
3. To expose the students to answer the queries raised on the topic of presentation.
4. To encourage students to work with innovative and entrepreneurial ideas
5. Demonstrate the ability to synthesize and apply the knowledge and skills acquired in the academic program to real-world problems
6. Evaluate different solutions based on economic and technical feasibility
7. Effectively plan a project and confidently perform all aspects of project management
8. Demonstrate effective written and oral communication skills

PROJECT GUIDELINES:

1. The Head of Department will appoint Internal supervisor to Guide the students in each group.
2. Each group should consist of Five students.
3. Each student in the group must actively participate and report to the internal supervisor.
4. Each group has to search for the internship from any industry/ institution, if not found they have to choose a project with the help of supervisor allotted such that, the aim of project work is to develop solutions to realistic problems applying the knowledge and skills obtained on the courses studied with specializations, new technologies and current industry practices.
5. Each student has to give minimum two seminars, one in the second week ("Project Design Seminar") another on 8^{th} week (project progress seminar).
6. Submit Title of the project and one page abstract /synopsis about the project in the first week to the Head, forwarded by the internal supervisor.
7. Each project should give a 30 minutes presentation using power point presentation and followed by 10 minutes of discussion.
8. Project seminar presentations should contain, source of the data, Sample data, data description, literature survey on the similar studies, objectives of the study, Methodology, statistical techniques, work plan etc. and details of progress of the work, individual roles and their work distribution and their plan etc.
9. Each group Project Report should follow the Ph.D. thesis norms with Plagiarism report and each group has to submit two copies duly signed by the Students, Supervisor, industry certificate (if exists) and Head of the Department on before the last instruct date of the semester.
10. Project Marks will be awarded based on all stages of the project and the topic chosen, seminar presentation, communication skills, role/ contribution of the student in the project etc and viva-voce conducted by the internal & External examiners.

DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD – 500 007.

M.Sc. APPLIED STATISTICS IV- SEMESTER

Paper	Sub. Code	Paper Title	Credits	Instruction Hours per Week	Semester end Exam duration	Max. Marks in Semester end Exam	Max. Marks in the Internal Assessment and Assignments
THEORY							
I	STAS4-I	Statistical Process and Quality Control (SPQC)	3	3	3	70	20+10
II	STAS4-II	Applied Stochastic Processes (ASP)	3	3	3	70	20+10
III	STAS4-III Elective - I	A) Operations Research-II (OR-II) B) Text Analytics (TA**)** C) Demography (DGY)	3	3	3	70	20+10
IV	STAS4-IV Elective - II	**A)** Artificial Neural Networks (ANN) B) Design & Analysis of Algorithms (DAA) C) Clinical Trails (CT)	3	3	3	70	20+10
PRACTICALS							
V	STAS-405	Statistical Process and Quality Control, Applied Stochastic Processes.	2	4	2	50	-
	STAS-406	Elective-I & Elective-II	2	4	2	50	-
VI	STAS-407	**Major Project**	4	8	2	**100**	-
Semester Total			34	***	***	480	120

M.SC. (APPLIED STATISTICS) SEMESTER IV

STAS-401: PAPER-I : STATISTICAL PROCESS AND QUALITY CONTROL (SPQC)

Unit–I

Basic concept of process monitoring – Basic principles, Choice of control limits, sample size and sampling frequency, rational subgroups, analysis of patterns on control charts, magnificent seven, nonmanufacturing applications of Statistical process control, Process capability and Process optimisation.

Unit–II

General theory and review of control charts for variable data and attributes : O.C. and A.R.L. functions of control charts, modified control charts for variables and Acceptance control charts for attributes, control by gauging.

Unit–III

Moving Average and exponentially weighted moving average charts, Cu-sum charts using V-Masks and decision intervals, Economic design of X bar chart. Concept of control chart for non-normal distributions, concept of Nonparametric control charts.

Unit–IV

Acceptance sampling plans for attribute inspection, single, double and sequential sampling plans and their properties; Rectifying sampling plans for attributes, AOQ, AOQL, designing of RSP. for specified AOQL and LTPD. Plans for inspection by variables for one–sided and two–sided specifications; Dodges Continuous sampling Plan–l and its properties modifications over CSP-1.

Unit–V

Process Capability Analysis: Capability indices Cp, Cpk and Cpm, estimation, confidence intervals and tests of hypotheses relating to capability indices for normally distributed characteristics. Multivariate quality control, use of control ellipsoid and of utility functions. Concept of TQM, Six sigma.

REFERENCES

1. Montgomery, D.C.(1985) : Introduction to Statistical Quality Control, Wiley
2. Wetherill, G.B. (1977): Sampling Inspection and Quality Control, Halsted Press.
3. Cowden, D. J. (1960) : Statistical Methods in Quality Control, Asia Publishing House.
4. Ott,E.R. (1975) : Process Quality Control, McGraw Hill
5. Phadke, M.S. (1989): Quality Engineering through Robust Design, Prentice Hall.
6. Wetherill, G.B., and Brown, D.W: Statistical Process Control: Theory and Practice, Chapman and Hall.

M.SC. (APPLIED STATISTICS) SEMESTER IV

STAS-402: PAPER II: APPLIED STOCHASTIC PROCESSES (ASP)

Unit–I

Markov Chains: Classification of states, canonical representation of transition probability matrix. Probabilities of absorption and mean times for absorption of the Markov Chain from transient states into recurrent classes. Limiting behaviour of Markov chain: Stationary distribution

Unit–II

Continuous–time Markov Processes: Kolmogorov–Feller differential equations, Poisson process and birth and death processes. Renewal Processes: Renewal process when time is discrete and renewal process with time is continuous, with examples. Renewal function, renewal density, limiting behaviour. Statement of elementary and basic renewal theorems.

Unit–III

Branching Processes: Examples of natural phenomena that can be modelled as a branching process. Probability of extinction; Statement of fundamental theorem of branching processes. Stochastic Processes in Biological Sciences: Markov models in population genetics; Recovery, relapse and death due to disease; cell survival after irradiation; compartmental analysis.

Unit-IV

Stochastic Processes in communication and information systems: Markov models in storage requirements for unpacked messages; buffer behaviour for batch arrivals; loop transmission systems; a probabilistic model for hierarchical message transfer.

Stochastic Processes in traffic–flow theory; some traffic flow problems; pedestrian traffic on a side–walk; free–way traffic; parking lot traffic; intersection traffic; left–turning traffic; pedestrian delay; headway distribution

Unit–V

Stochastic Processes in social and behavioural sciences; Markov chain models in the study of social mobility; industrial mobility of labour; educational advancement; labour force planning and management; diffusion of information.

Stochastic Processes in Business Management: Markov models in marketing and accounting; consumer behaviour; selecting a port–folio of credit–risks; term structure; human resource management; income determination under uncertainty.

REFERENCE

1. Bhat, U.N., (1984): Elements of Applied Stochastic Processes, John Wiley
2. Ross, S. (1996): Stochastic Processes, Second Edition, John Wiley.
3. J. Medhi: Stochastic Processes.

M.SC. (APPLIED STATISTICS) SEMESTER IV

ELECTIVE-I: STAS-403(A): PAPER III (A): OPERATIONS RESEARCH-II (OR-II)

Unit–I

Non-linear Programming problem – Formulation Generalised Lagrange multiplier technique, Kuhn-Tucker necessary and sufficient conditions for optimality of an NLPP, Wolfe's and Beale's Algorithms for solving QPP. Separate Programming Problem; Piecewise linearization method.

Unit–II

Dynamic Programming, Principle of optimality, solution of LPP by Dynamic Programming technique, Knapsack problem by Dynamic Programming Technique. General goal Programming model and formulation of its objective function. Solutions to linear goal programming and linear integer goal programming.

Unit–III

Game Theory : 2 person zero sum game, pure strategies with saddle point, mixed strategies with saddle point, principles of dominance and games without saddle point, 2xm, mx2, mx n games

Unit-IV

Introduction to simulation, generation of random numbers for Uniform, Normal, Exponential, Cauchy and Poisson Distributions. Estimating the reliability of the random numbers, Simulation to Queuing and Inventory problem.

Unit–V

S-S policy for inventory and its derivation in the case of exponential demand; Models with variable supply and models for perishable Items. Replacement Problems; Introduction, block and age replacement policies, replacement of items with long life. Machine interference problems.

REFERENCES

1. Taha, H.A.(1982): Operations Research : An Introduction; McMillan
2. Kantiswarup;Gupta P.K. and Singh,M.N.(1985) : Operations Research; Sultan Chand.
3. Sharma,S.D.: Operations Research.

M.SC.(APPLIED STATISTICS) IV-SEMESTER

ELECTIVE I(B): STAS-403:: PAPER III (B): TEXT ANALYTICS (TA)

Unit - I

Introduction to Natural Language Processing Basic, Language Syntax and Structure (Words, Phrases, Clauses, & Grammar), Language Semantics Processing, (Lexical Semantic Relations, Homonyms, Homographs, and Homophones, Capitonyms, Hyponyms and Hypernyms),Text Corpora (Corpora Annotation and Utilities),Accessing Text Corpora (Brown Corpus, WordNet Corpus) and NLP Applications (Machine Translation, Text Summarization and Text categorization).

Unit – II

Concept of the Tokenization, Sentence Tokenization, Word Tokenization, Concept of the Text Normalization,(Cleaning Text, Removing Special characters, Removing stop words,..etc) correcting words using stemming and Lemmatization and Understanding text syntax and structure.(POS tagging and Parsing)

Unit – III

Concepts of feature extraction, Methods of Feature extraction (Bag of words Model, TF-IDF Models, Advanced word Factorization Models likes Word2vec), Strengths and weakness of models and Word cloud … etc, Concepts of Document term matrix, Term Document Matrix.

Unit – IV

Concepts of Topic Modelling, Algorithms of Topic Modelling (Latent Semantic Indexing (LSI) , Latent Dirichlet Allocation (LDA), Non Negative Matrix Factorization (NMF) and Similarity based text clustering models),

Unit-V

Text Classification using supervised methods (Like Multinomial Naïve Bayes, Support vector machines, Random Forest …), concept of Sentiment Analysis and its applications.

REFERENCES

1) Chapman & Hall : Handbook of Natural Language Processing, Second Edition.
2) CRC: Machine Learning & Pattern Recognition, 2nd Edition.
3) Christopher Manning and Hinrich Schuetze: Foundations of Statistical Natural Language Processing.
4) Dipanjan Sarkar: Text Analytics with Python, A press Publication.
5) Julia Silge: Text Mining with R: A Tidy Approach, 1st Edition.

M.SC. (STATISTICS) SEMESTER IV

ELECTIVE-I(C): STAS-404(C): PAPER III(C) : DEMOGRAPHY (DGY)

Unit–I

Coverage and content errors in demographic data, use of balancing equations and Chandrasekharan, Deming formula to check completeness of registration data.

Unit-II

Adjustment of age data - use of Whipple, Myer and UN indices. Population composition, dependency ratio.

Unit–III

Measures of fertility; stochastic models for reproduction, distributions of time to first birth, inter-live birth intervals and of number of births (for both homogeneous and nonhomogeneous groups of women), estimation of parameters; estimation of parity progression ratios from open birth interval data.

Unit–IV

Measures of Mortality; construction of abridged life tables. Distributions of life table functions and their estimation. Stable and quasi-stable populations, intrinsic growth rate. Models for population growth and their fitting to population data. Stochastic models for population growth.

Unit–V

Stochastic models for migration and for social and occupational mobility based on Markov chains. Estimation of measures of mobility. Methods for population projection. Use of Leslie matrix.

REFERENCES

1. Bartholomew, D. J. (1982). Stochastic Models for Social Processes, John Wiley.
2. Benjamin, B. (1969). Demographic Analysis, George, Allen and Unwin.
3. Chiang, C. L. (1968). Introduction to Stochastic Processes in Biostatistics; John Wiley.
4. Cox, P. R. (1970). Demography, Cambridge University Press.
5. Keyfitz, N. (1977). Applied Mathematical Demography; Springer Verlag.

M.SC.(APPLIED STATISTICS) SEMESTER IV

ELECTIVE-II(A): STAS-404(A): ARTIFICIAL NEURAL NETWORKS (ANN)

Unit – I

Basics of Artificial Neural Networks (ANN), Human vs Computers, Organization of the Brain, Biological Activations of Neuron; Artificial Neuron Models: McCulloch-Pitts, Perceptron, Adaline, Hebbian Models; Historical Developments of ANN, Characteristics of ANN, Types of Neuron Activation Function, Signal functions and their properties, monotonicity. ANN Architecture, Classification Taxonomy of ANN, Un-supervised and Reinforcement learning; Learning tasks, Memory, Adaptation, Statistical nature of the learning process. Statistical learning theory**.** Gathering and partitioning of data for ANN and its pre and post processing.

Unit – II

Perceptron Learning Algorithm, Derivation, Perceptron convergence theorem (statement); Multi-layer Perceptron Learning rule, limitations. Applications of the Perceptron learning. Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning. Feed-forward and Fedd-back Back-Propagation Algorithms and derivation, learning rate, Momentum, Difficulties and Improvements. Bias and Variance. Under- Fitting and Over-Fitting

Unit-III

Radial Basis Function Networks: Introduction, Regularization theory, Regularization Networks, Generalized Radial Basis Function Networks, Approximation properties of Radial Basis Function Networks, Comparison with Multi-layer Perceptron, Applications.

Unit-IV

Hebbian Learning, Competitive learning**.** Self Organizing Maps**:** Two basic feature mapping models, Self-Organizing Map, SOM algorithm, properties of feature map, computer simulations, Vector quantization, Learning vector quantization, Hierarchical Vector Quantization,

Unit-V

Boltzman Machine and its learning rule, Hopfield model and its learning. Sigmoid belief network learning procedure, Stochastic machines. Applications of ANN in Classification, Clustering, Regression, Time series forecasting.

REFERENCES

1. Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. New York: Macmillan Publishing. A comprehensive book and contains a great deal of background theory
2. Yagnanarayana, B. (1999): “Artificial Neural Networks” PHI
3. Bart Kosko(1997): Neural Networks and Fuzzy systems, PHI
4. Jacek M. *Zurada(1992): Artificial Neural Systems*, West Publishing Company.
5. Carling, A. (1992). Introducing Neural Networks. Wilmslow, UK: Sigma Press.
6. Fausett, L. (1994). Fundamentals of Neural Networks. New York: Prentice Hall.

M.SC. (APPLIED STATISTICS) SEMESTER IV

ELECTIVE II(C): STAS-404 (C): PAPER IV(C): CLINICAL TRIALS (CT)

Unit–I

Introduction to clinical trials : The need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I-IV trials, multi-center trials. Data management: data definitions, case report forms, database design, data collection systems for good clinical practice.

Unit-II

Determination of sample size: for two independent samples of Dichotomous Response variables, for two independent samples of Continuous Response variables and for repeated variables.

Unit–III

Design of clinical trials : parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, review of 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, design of bioequivalence trials.

Unit–IV

Reporting and analysis: analysis of categorical outcomes from Phase I - III trials, analysis of survival data from clinical trials.

Unit–V

Surrogate endpoints: selection and design of trials with surrogate endpoints, analysis of surrogate endpoint data. (2L) Meta-analysis of clinical trials.

REFERENCES

1. S. Piantadosi (1997). Clinical Trials : A Methodological Perspective. Wiley and Sons.
2. C. Jennison and B. W. Turnbull (1999). Group Sequential Methods with Applications to Clinical Trials, CRC Press.
3. L. M. Friedman, C. Furburg, D. L. Demets (1998). Fundamentals of Clinical Trials, Springer Verlag.
4. J. L. Fleiss (1989). The Design and Analysis of Clinical Experiments. Wiley and Sons.
5. E. Marubeni and M. G. Valsecchi (1994). Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley and Sons.

M.SC. (APPLIED STATISTICS) SEMESTER IV

ELECTIVE-II: STS-404(B): PAPER IV(B): DESIGN AND ANALYSIS OF ALGORITHMS

UNIT I

Introduction to Algorithms: Algorithm, Time & space complexity, Asymptotic Notations. Writing pseudocode, Design Techniques.

Divide and Conquer: Control Abstraction, Binary Search, Finding the Maximum and Minimum, Merge Sort; Quick Sort, Selection sort, Strassen's Matrix Multiplication, Convex Hull.

UNIT-II

Greedy Method: Control Abstraction, Knapsack Problem, Job Sequencing with Deadlines, Minimum-Cost Spanning Trees (Kruskal's & Prim's), Single Source Shortest Paths (Dijkstra's).

Dynamic Programming: Control Abstraction, Multistage Graphs, All-Pairs Shortest Paths, Single-Source Shortest Paths, Optimal Binary Search Trees, 0/1 Knapsack, Traveling Salesperson Problem.

UNIT-III

Basic Traversal and Search Techniques: Techniques for Binary Trees, Techniques for Graphs, Connected Components and Spanning Trees, Biconnected Components and DFS.

UNIT -IV

Back Tracking: Control Abstraction, , 8-Queens Problem, Sum of Subsets, Graph Colouring, Hamiltonian Cycles, Knapsack Problem.

Branch-Bound: Control Abstraction, 0/1 Knapsack Problem, Traveling Sales Person problem.

UNIT -V

NP-Hard and NP-Complete Problems: Basic Concepts, Cook's Theorem, NP-Hard. Graph Problems, NP-Hard Scheduling Problems, NP-Hard Code Generation, Some Simplified NP-Hard Problems.

REFERENCE BOOKS

1. E Horowitz, S Sahni, S Rajasekaran, (2007): Fundamentals of Computer Algorithms, 2/e, Universities Press.
2. T.H. Cormen, CE Leiserson, R.L Rivert, C Stein, (2010): Introduction to Algorithms, 3/e, PHI.
3. R. Pannerselvam (2007): Design and Analysis of Algorithms, PHI.
4. Hari Mohan Pandey, (2009): Design, Analysis and Algorithm, University Science Press.

M.SC. (APPLIED STATISTICS) IV-SEMESTER

STAS-405: PAPER-V : STATISTICAL PROCESS AND QUALITY CONTROL, APPLIED STOCHASTIC PROCESSES

PRACTICAL-I (CONVENTIONAL)

SECTION-A: STATISTICAL PROCESS AND QUALITY CONTROL

1. Construction of $\overline{X}$, R and σ - charts and OC curves for $\overline{X}$ and R charts
2. Construction of p – chart (with constant and variable sample size) – OC curve for constant sample size
3. Construction of C–chart and U–chart and OC curve for C–Chart
4. Construction of Simple and Exponentially weighted moving average control chart and simple moving range control chart.
5. Construction of CUSUM chart using tabular approach.
6. Construction of CUSUM charts V – Mark and ARL curves
7. Designing Single Sampling Plans for specified p_1,p_2,α and β
8. OC, ASN Curves for double sampling plans – designing for specified p_1,p_2,α and β
9. Construction of AOQ and AFI curves for CSP–I
10. Computation of process capability indices

SECTION-B: APPLIED STOCHASTIC PROCESSES

1. Classification of states of a Markov chain, determination of periods of states and mean recurrence times of recurrent states.
2. Computation of higher order transition probability matrix in a two–state Markov chain using spectral decomposition
3. Probabilities of absorption and mean time for absorption from each transient state into recurrent class.
4. Determination of stationary distribution(s) and evaluation of the same.

M.SC. (APPLIED STATISTICS) IV-SEMESTER

STAS-406: PAPER-VI: SECTION-A (ELECTIVE-I) & SECTION-B (ELECTIVE-II)

PRACTICAL-II (CONVENTIONAL)

ELECTIVE-I (A): ADVANCED OPERATIONS RESEARCH

1. Wolfe and Beale's methods for QPP
2. Separable Programming problem
3. Dynamic Programming Problem
4. Goal Programming Problem
5. Game Theory
6. Simulation

ELECTIVE – I (B): TEXT ANALYTICS

1. Perform data collection by web scrapping with python and Perform following tasks (i) Find the URL that you want to scrape (ii) Inspecting the Page (iii) Find the data you want to extract (iv) Write the code (v) Run the code and extract the data (vi) Store the data in the required format.
2. Perform following Data Pre-processing tasks in Python using Scikit-learn. standardization, normalization, encoding, discretization, imputation of missing values. Use your own dataset to perform all pre-processing tasks as suggested in given reference.
   (i) https://www.analyticsvidhya.com/blog/2016/07/practical-guide-datapreprocessing-python-scikit-learn/
   (ii) https://scikit-learn.org/stable/modules/preprocessing.html
3. Answer the following question in your blog (As per dataset taken by you): Dataset Description: Task to be performed: How to decide variance threshold in data reduction? Code Snapshot, Output Snapshot, Task-2, Code Snapshot, Output Snapshot.
   Perform following Data Pre-processing tasks using python
   Data reduction using variance threshold, univariate feature selection,
   recursive feature elimination, PCA, correlation
   Reference:
   1. https://medium.com/analytics-vidhya/feature-selection-using-scikit-learn5b4362e0c19b
   2. https://machinelearningmastery.com/rfe-feature-selection-in-python/
   3. https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60
   4. https://towardsdatascience.com/feature-selection-using-python-for-classificationproblem-b5f00a1c7028
   5. https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/
   Answer the following question in your blog (As per dataset taken by you):
   Dataset Description:
   Task to be performed:
   Why feature selection is important?? Its advantages/disadvantages.
   Code Snapshot
   Output Snapshot
   What is the impact on accuracy, with or without data reduction?
   Code Snapshot
   Output Snapshot
   Amongst all methods, which method avoids overfitting and improves model performance

ELECTIVE-I (C): DEMOGRAPHY

1. Construction of Abridged life tables
2. Fitting of population growth models
3. Estimation of population projection
4. Estimation of Life table functions

ELECTIVE – II (A) ARTIFICIAL NEURAL NETWORKS

1. Perceptron learning rule (single layer)
2. Gradient Descent Learning
3. Multilayer Perceptron Learning
4. Widrow-Hoff Learning
5. Hebbian Learning
6. Competitive Learning
7. Back-Propagation Algorithm (Forward & Backward)
8. Radial Basis Function

ELECTIVE-II(B): DESIGN & ANALYSIS OF ALGORITHMS (USING PYTHON)

1. Write a program for sorting the given list using: Merge Sort, Quick Sort, Heap Sort.
2. Write a program to find the given number in a list using Binary Search.
3. Write a program to find the minimal spanning tree using Kruskal's and Prims Algorithms.
4. Write a program to find the shortest path using Dijkstra's Algorithm.
5. Write a program to solve using dynamic programming technique for Travelling sales man problem. Multistage Graph problem, Optimal Binary Search Trees.
6. Write a program to solve Knapsack problem using Back tracking

ELECTIVE – II (C): CLINICAL TRIALS

1. Determination of Sample size
2. Multiple Logistic Regression with two or Three variables
3. Analysis of Clinical trial data using Cross over design
4. Analysis of Clinical trial data using Parallel design
5. Meta-analysis of Clinical trials
6. Analysis of Clinical data using Factorial Experiments

M.SC. (APPLIED STATISTICS) IV-SEMESTER

STAS-407: PAPER-VII: MAJOR PROJECT

Note: Follow the guidelines of the project specified in STS-308, Mini project.

****