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 $\overline{X}$ 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.

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 (4th 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.