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, 2^{nd} 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