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

M.Sc. STATISTICS
CB - SCHEME OF INSTRUCTION AND EXAMINATION
WITH EFFECT FROM 2018 – 2019

SEMESTER I

Paper	Sub. Code	Paper Title	Instruction Hrs/ Week	Duration of Exam (in Hrs)	Max. Marks	IA and Assign.	Credits
THEORY							
I	STS1-I	Mathematical Analysis and Linear Algebra (MA and LA)	4	3	80	20	4
II	STS1-II	Probability Theory (PT)	4	3	80	20	4
III	STS1-III	Distribution Theory (DT)	4	3	80	20	4
IV	STS1-IV	Theory of Estimation (ET)	4	3	80	20	4
PRACTICALS							
V	STS1-V	Practical – I Statistical methods using Python Programming	9	3	100	***	4
VI	STS1-VI	Practical – II Linear Algebra, Distribution Theory and Theory of Estimation (LA, DT, ET)	9	3	100	***	4
Total			34	***	520	80	24
Semester Total					600		

M. Sc. (Statistics) Semester I

STS1- I : Paper I - Mathematical Analysis and Linear Algebra (MA and LA)

UNIT–I

Functions of Bounded Variation (BV), Total variation and its additive Property. Functions of BV expressd as the difference of increasing functions.

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

UNIT–II

Complex derivatives. Cauchy-Riemann equations. Analytic functions. Statements of Cauchy theorem and integral formula. Power, Taylor's and Laurent's series. Zeroes and poles. Statement of Cauchy residue theorem. Cantour integration. Evaluation of real valued integrals by means of residues.

Functions of several variables-concepts of limit, continuity, directional derivatives, partial derivatives, total derivative, extreme and saddle points with examples. Taylor's expansion. Multiple Integration. Application of Jacobians in the evaluation of multiple integrals.

UNIT – III

Vector spaces with an inner product, Gram-Schmidt orthogonolization process, orthonormal basis and orthogonal projection of a vector.

Moore-Penrose and generalized inverses and their properties. Solution of matrix equations. Sufficient conditions for the existence of homogeneous and non-homogeneous linear equations.

UNIT – IV

Characteristic roots and vectors, Caley-Hamilton theorem, algebraic and geometric multiplicity of a characteristic root and spectral decomposition of a real symmetric matrix.

Real quadratic forms (QFs), reduction and classification of QFs, index and signature. Simultaneous reduction of two QFs. Extreme form of a QF. Cauchy-Schwartz and Hadamard inequalities for matrices.

REFERENCES

1. Apostol,T.M. (1985) : Mathematical Analysis, Narosa, Indian Ed.
2. Malik,S.C. (1984) : Mathematical Analysis, Wiley – Eastern.
3. Rudin,W. (1976) : Principles of Mathematical Analysis, McGraw Hill.
4. Graybill,F.A. (1983) : Matrices with applications in statistics, 2nd ed, Wadsworth.
5. Rao,C.R. (1973) : Linear Statistical inference and its applications, 2nd Ed, John Wiley & Sons Inc.
6. Searle,S.R. (1982) : Matrix algebra useful for statistics, John Wiley and Sons Inc.
7. Rao,C.R., Mithra,S.K. (1971) : Generalised inverse of matrices and its applications, John Wiley & Sons Inc.
8. Rao, A.R. and Bhimasankaram, P(1992) : Linear algebra, Tata – McGrawhill Publishing Co. Ltd.

M. Sc. (Statistics) Semester I

STS1- II : Paper II - Probability Theory (PT)

UNIT – I

Classes of sets, fields, sigma-fields, minimal sigma-fields, Borel sigma-fields in R, Measure, Probability Measure, Properties of a Measure, Caratheodory extension theorem (Statement only), measurable function, random variables, distribution function and its properties, expectation, statements and applications of monotone convergence theorem, Foatou's lemma, dominated convergence theorem.

UNIT – II

Expectations of functions of rv's, conditional expectation and conditional variance, their applications. Characteristic function of a random variable and its properties. Inversion theorem, uniqueness theorem (Functions which cannot be Characteristic functions). Levy's continuity theorem (Statement only). Chebychev, Markov, Cauchy-Schwartz, Jenson, Liapunov, Holder's and Minkowsky's inequalities.

UNIT – III

Sequence of Random variables, convergence in Probability, convergence in distribution, almost sure convergence, convergence in quadratic mean and their interrelationships, Slutskey's theorem, Borel-Cantelli lemma Borel 0-1 law, Kolmogorov 0-1 law (Glevenko – Cantelli Lemma - Statement only).

UNIT – IV

Law of large numbers, Weak law of large numbers, Bernoulli and Khintchen's WLLN's, Kolomogorov Inequality, Kolmogorov SLLN for independent random variables and statement only for i.i.d. case, statements of three series theorem.

Central Limit theorems : Demoviere - Laplace CLT, Lindberg-Levy CLT, Liapounou's CLT, Statement of Lindberg-Feller CLT, simple applications, statement of Cramer-Wald theorem, Asymptotic distribution of sample quantiles.

REFERENCES

1. Ash Robert (1972) : Real analysis and Probability, Academic Press
2. Bhat,B.R. : Modern probability Theory, 3rd Edition, New Age India
3. Rohatgi,V.K. : Introduction to Probability Theory and Mathematical Statistics
4. Milton and Arnold – Introduction to probability and Statistics (4th Edition)-TMH publication.

ADDITIONAL REFERENCES

1. Kingman,J.F.C. and Taylor, S.J. (1966) : Introduction of measure and probability, Cambridge University press
2. Basu,A.K. : Probability and Measure,Narosa (PHI)
3. W.Feller : An Introduction to Probability theory and its Applications Vol I and II, John Wiely.

M. Sc. (Statistics) Semester I

STS1- III : Paper III - Distribution Theory (DT)

UNIT – I

Normal, Lognormal, Weibull, Pareto and Cauchy distributions and their properties. Joint, Marginal and conditional pmf's and pdf's.

UNIT – II

Families of Distributions: Power series distributions, Exponential families of distributions. Functions of Random variables and their distributions (including transformation of rv's). Bivariate Normal, Bivariate Exponential (Marshall and Olkins form), Compound Binomial - Poisson, Gamma(α,β). Truncated (Binomial, Poisson, Normal and Lognormal) and mixture distributions - Definition and examples.

UNIT – III

Sampling Distributions of sample mean and variance, independence of $\overline{X}$ and s^2. Central and Non-central χ^2,t and F distributions.

UNIT – IV

Distributions of quadratic forms under normality and related distribution theory. Order statistics, their distributions and properties. Joint and marginal distributions of order statistics and Distribution of Range. Extreme values and their asymptotic distributions (statements only) with applications.

REFERENCES

1. Rohatgi,V.K.(1984) : An introduction to probability theory and mathematical Statistics, Wiley Eastern.
2. Rao,C.R. (1972) : Linear Statistical Inference and its applications, 2/e, Wiley Eastern
3. Milton and Arnold – Introduction to probability and Statistics (4th Edition)-TMH publication.

ADDITIONAL REFERENCES

1. Pittman,J. (1993) : Probability, Narosa Publishing House
2. Johnson,S. and Kotz,(1972) : Distributions in Statistics,Vol. I,II and III, Houghton and Miffin.
3. Cramer, H. (1946) : Mathematical methods of statistics, Princeton.
4. Dudewicz,E.J., and Mishra, S.N. (1988) : Modern Mathematical statistics, Wiley International Students edition.

M.Sc. (Statistics) Semester I
STS1- IV : Paper IV - Estimation Theory (ET)

UNIT – I

Point Estimation Vs. Interval Estimation, Advantages, Sampling distribution, Likelihood function, exponential family of distribution.

Desirable properties of a good estimator: Unbiasedness, consistency, efficiency and sufficiency - examples. Neyman factorization theorem (Proof in the discrete case only), examples. UMVU estimation, Rao-Blackwell theorem, Fisher Information, Cramer-Rao inequality and Bhattacharya bounds.

UNIT II

Completeness and Lehmann-Scheffe theorem. Median and modal unbiased estimation. Estimation of bias and standard deviation of point estimation by the Jackknife, the bootstrap methods with examples.

UNIT III

Methods of estimation, method of moments and maximum likelihood method, examples. Properties of MLE. Consistency and asymptotic normality of the consistent solutions of likelihood equations. Definition of CAN and BAN, estimation and their properties, examples.

UNIT IV

Concept of U statistics 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.

REFERENCES

1. Goon, Gupta and Das Gupta : Outlines of Statistics, Vol. 2, World Press, Calcutta.
2. Kale, B.K. (1999): A first course on parametric inference, Narosa publishing house.
3. Rohatgi, V.K.: An introduction to Probability theory and mathematical statistics, Wiley Eastern.

ADDITIONAL REFERENCES

1. Rao, C.R.: Linear Statistical Inference and its applications, John Wiley
2. Gray and Schucany : Generalized Jackknife; Marcel Decker
3. Bradely Efron and Robert J. Tibshirani : An Introduction to the Bootstrap, Chapmen and Hall.
4. Lehman, E.L. (1983) : Theory of point estimation, John Wiley
5. Gray, Schncory and Watkins : Generalized Jacknife, Dovenpul

M.Sc. (Statistics) Semester I
STS1- V : Paper V - Practical – I Statistical methods using Python Programming

Topics to be covered (25% Weight for Theory): 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 Practical's using Python programming: (75% including Execution of Program)
A. Basic Programs (2 out of 6)

1. Program to examine the given number is a prime number or not.
2. Program to find the Factorial of positive integer.
3. Program to find the largest among the given three numbers.
4. Program to generate Fibonacci sequence up to given number n.
5. Program to construct a Pascal Triangle.
6. Program to find the value of e^x, Sin x and Cos x using series expansion
7. Program to find the sum of two matrices $[A]_{mxp}$ and $[B]_{mxp}$
8. Program to find the product of two matrices $[A]_{mxp}$ and $[B]_{pxr}$.
9. Program to sort the given set of numbers using bubble sort and finding median.
10. Program with a function that accepts a string as an argument and returns the no. of vowels that the string contains. Another function to return number of consonants.
11. Program that opens specified text file and then displays list of all unique words found in the file.
12. Program to find the Median, Mode for the given of array of elements.
13. Program to find the first four Central & Non-central moments to the given array of elements.
14. Program to generate random numbers from Uniform, Binomial, Poisson, Normal, Exponential.
15. Program for preparation of frequency tables and computing mean, median, mode, variance and standard deviation of the frequency distribution.
16. Program to Fitting of Binomial distribution for the given frequency distribution (recursive)
17. Program to Fitting of Poisson distribution for the given frequency distribution (recursive)
18. Program to Fitting of Negative Binomial distribution for the given frequency distribution (recursive).
19. Program to Fitting of Exponential Distribution for the given frequency distribution (recursive)
20. Program for finding the Correlation and regression lines for the given Bi-variate data.
21. Program for finding the roots of a quadratic equation.
22. Solution to simultaneous equations by Gauss - Siedal method (minimum 3 variables)

References:

1. Tony Gaddis, Starting Out With Python (3e)
1. Kenneth A. Lambert, Fundamentals of Python
2. Clinton W. Brownley, Foundations for Analytics with Python
3. James Payne, Beginning Python using Python 2.6 and Python 3
4. Charles Dierach, Introduction to Computer Science using Python
5. Paul Gries, Practical Programming: An Introduction to Computer Science using Python 3

M.Sc. (Statistics) Semester I
STS1-VI : Paper VI Practical (LA, DT, ET)

PRACTICALS IN LINEAR ALGEBRA, DISTRIBUTION THEORY AND ESTIMATION

LINEAR ALGEBRA

1. Inverse of a matrix by partition method
2. Solutions of linear equations by sweep-out method
3. Solutions of linear equations by Doolittle Method
4. Computation of Moore-Penrose inverse by Penrose method
5. Computation of generalized inverse of a matrix.
6. Formation of characteristic equation by using traces of successive powers
7. Spectral decomposition of a square matrix of third order
8. Simultaneous reduction of a pair of quadratic forms to diagonal and canonical forms.
9. Finding orthonormal basis by Gram – Schmidt process.

DISTRIBUTION THEORY

1. Discrete Bivariate distributions
2. Fitting of Cauchy distributions
3. Fitting of Gamma distribution with two parameters
4. Fitting of Lognormal Distribution
5. Fitting of Weibull Distribution
6. Fitting of Pareto distribution.

ESTIMATION THEORY

1. Computation of Jackknife estimates
2. Computation of Boot-strap estimates
3. MLE by Scoring method
4. Confidence limits for parameters of normal population
5. Large sample confidence limits in case of Binomial, Poisson, Exponential distributions.

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

M.Sc. STATISTICS
CB - SCHEME OF INSTRUCTION AND EXAMINATION
WITH EFFECT FROM 2018 – 2019

SEMESTER II

Paper	Sub. Code	Paper Title	Instruction Hrs/ Week	Duration of Exam (in Hrs)	Max. Marks	IA and Assign.	Credits
THEORY							
I	STS2-I	Sampling Techniques (ST)	4	3	80	20	4
II	STS2-II	Parametric Inference (PI)	4	3	80	20	4
III	STS2-III	Linear Models and Design of Experiments (LM and DOE)	4	3	80	20	4
IV	STS2-IV	Multivariate Analysis (MVA)	4	3	80	20	4
PRACTICALS							
V	STS2-V	Practical – I Sampling Techniques and Parametric Inference (ST, PI)	9	3	100	***	4
VI	STS2-VI	Practical – II Linear Models, Design of Experiments and Multivariate Analysis (LM, DOE, MVA)	9	3	100	***	4
Total			34	***	520	80	24
Semester Total					600		

M.Sc. (Statistics) Semester II

STS2 - I : Paper I - Sampling Techniques (ST)

UNIT – I

Review of SRSWR/WOR, Stratified random sampling and Systematic Sampling.

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.

UNIT – II

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.

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.

UNIT – III

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.

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.

UNIT – IV

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.

ADDITIONAL REFERENCES

1. Des Raj (1976) : Sampling Theory, Tata McGraw Hill, New Delhi.
2. Sukhatme etal (1984) : Sampling Survey methods and its applications, Indian society of Agricultural Statistics.
3. Cochran, W.C. (1977) : Sampling Techniques, Third Edition, Wiley Eastern.

M.Sc. (Statistics) Semester II

STS2 - II : Paper II - Parametric Inference (PI)

Unit–I

Fundamental notions of hypothesis testing–Statistical hypothesis, statistical test, Critical region, types of errors, test function, randomised and non–randomised tests, level of significance, power function, Most powerful test, Neyman–Pearson fundamental lemma, MLR families and Uniformly most powerful tests for one parameter exponential families.

Unit–II

Concepts of consistency, unbiased and invariance of tests. Likelihood Ratio tests, statement of the asymptotic properties of LR statistics with applications (including homogeneity of means and variances). Relation between confidence interval estimation and testing of hypothesis. Concept of robustness in estimation and testing with example. ML Estimation and testing of Transition Probability Matrix.

Unit–III

Concept of sequential estimation, sequential estimation of a normal population. Notions of sequential versus fixed sample size techniques. Wald's sequential probability Ratio test (SPRT) procedure for testing simple null hypothesis against simple alternative. Termination property of SPRT. SPRT procedures for Binomial, Poisson, Normal and Exponential distributions and associate OC and ASN functions. Statement of optimality of SPRT.

Unit–IV

Concepts of loss, risk and decision functions, admissible and optimal decision functions, Estimation and testing viewed as decision problems, apriori and aposteriori distributions, conjugate families, Bayes and Minmax decision functions with applications to estimation with quadratic loss.

REFERENCES:

1. Rohatgi,V.K. : An Introduction to probability theory and Mathematical Statistics (Wiley Eastern Ltd)
2. Wald, A : Sequential Analysis, Dover Publications
3. Ferguson, R.S. : Mathematical Statistics, a decision theoretic approach (Academic Press)
4. Rao,C.R. : Linear Statistical Inference and its applications, John Wiley
5. Medhi, J : Stochastic Processes – New age Publications

ADDITIONAL REFERENCES

1. Lehman, E.L.: Testing statistical Hypothesis, John Wiley
2. Mark Fisz: Probability theory and Mathematical Statistics
3. Parimal Mukhopadhyay: Mathematical Statistics

M.Sc. (Statistics) semester II

STS2 - III : Paper III - Linear Models and Design of Experiments (LM & DOE)

UNIT– I (LM)

Formulation of a linear model 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.

UNIT– II

Simple linear regression, examining the regression equation, Lack of fit and pure error. Analysis of Multiple regression models. Estimation and testing of regression parameters, sub-hypothesis. Introduction of residuals, overall plot, time sequence plot, plot against Yi, Predictor variables X_{ij}, Serial correlation among the residual outliers. The use of dummy variables in multiple regression, Polynomial regressions –use of orthogonal polynomials. Derivation of Multiple and Partial correlations, tests of hypothesis on correlation parameters.

UNIT– III (DOE)

Analysis of Covariance: One-way and Two-way classifications.
Factorial experiments: Estimation of Main effects, interaction and analysis of 2^k, factorial experiment in general with particular reference to k = 2, 3 and 4 and 3^2 factorial experiment. Multiple Comparisons: Fishers least significance difference (LSD) and Duncan's Multiple Range test (DMR test).

UNIT – IV

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.

REFERENCES

1. Searles S.R.(1971):Linear statistical Models.
2. Draper and Smith: Applied Regression Analysis
3. Montogomery,D.C.:Design and Analysis of Experiments, John Wiley
4. Giri, N.C.:Analysis of Variance

ADDITIONAL REFERENCES

1. Kshirasagar A.M.(1972): A course in Linear Models.
2. Graybill F.A(1966): An introduction to linear statistical models- Vol.I
3. Gultman (1982): Linear Models - An Introduction.
4. Rao A.R and Bhimsankaram P: Linear Algebra – Hindustan Agency.
5. Kempthrone: Design and Analysis of Experiments.
6. Cochran and Cox: Experimental Designs.

M.Sc. (Statistics) Semester II

STS2 - IV : Paper IV - Multivariate Analysis (MVA)

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.

UNIT – II

Wishart matrix – its distribution and 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. 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 – IV

Principal components, Dimension reduction, graphical of Principal Components, canonical variables and canonical correlation – definition, use, estimation and computation.

Concepts of cluster analysis and multi – dimensional scaling.

Introduction to Factor analysis, orthogonal factor model.

REFERENCES

1. Anderson, T.W. (1983) : An Introduction to multivariate statistical analysis, 2nd Edition, Wiley.
2. Kshirasagar, A.M. (1972) : Multivariate Analysis, Marcel Decker.
3. Johnson, R.A.W.: Applied Multivariate Analysis.

ADDITIONAL REFERENCES

1. Giri, N.C. (1977): Multivariate statistical inference, Academic Press
2. Morrison, D.F. (1976): Multivariate Statistical Methods, 2nd Edition, McGraw Hill
3. Muirhead, R.. (1982) : Aspects of multivariate statistical theory, J. Wiley.

M.Sc. (Statistics) Semester II
STS2 - V : Paper V Practical - I (ST and PI)

PRACTICALS IN SAMPLING TECHNIQUES AND PARAMETRIC INFERENCE

SAMPLING TECHNIQUES

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.

PARAMETRIC INFERENCE

1. Type I and Type II error probabilities
2. MP and UMP tests
3. Likelihood Ratio tests
4. Large Sample tests for means, proportions and correlation coefficient
5. Sequential probability Ratio test and Computation of OC and ASN function (Binomial, Poisson, Normal, Exponential)
6. Determination of Bayes and Minimax decision rules (Finite no. Of actions and finite no. of states of n atoms)

M.Sc. (Statistics) Semester II

STS2 – VI : Paper VI Practical - II (LM & DOE and MVA)

Practicals in Linear Models, Designs of Experiments and Multivariate Analysis

Linear Models and Designs of Experiments

1. Computation of BLUE and testing their parameters.
2. Computation of Pure error and Lack of fit.
3. Computation of residuals and their plots for two and three variables.
4. Computation of Multiple Correlation coefficient
5. Computation of Partial Correlation coefficient
6. Testing of Multiple and Partial Correlation Coefficients.
7. Analysis of 2^3, 2^4 and 3^2 factorial experiments.
8. Analysis of total confounding and partial confounding of 2^3 design.
9. Analysis of one–half fraction of 2^4 designs and one–quarter fraction of 2^5 designs.
10. Analysis of Split-Plot design.

Multivariate Analysis

1. MLE of Mean vector and variance covariance Matrix from Normal population.
2. Hotelling's T^2 and Mahalanobi's D^2.
3. Computation of Principal components.
4. Classification between two normal populations by discriminant analysis.
5. Cluster analysis.
6. Computation of Canonical variables and correlation.

DEPARTMENT OF STATISTICS

UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD – 500 007

M.Sc. STATISTICS

CB - SCHEME OF INSTRUCTION AND EXAMINATION

WITH EFFECT FROM 2019 – 2020

SEMESTER III

Paper	Sub. Code	Paper Title	Instruction Hrs/ Week	Duration of Exam (in Hrs)	Max. Marks	IA and Assign.	Credits
THEORY							
I	STS3-I	Non-Parametric Inference (NPI)	4	3	80	20	4
II	STS3-II	Quality Control and Optimization Techniques (QCOT)	4	3	80	20	4
III	STS3-III	Elective - I	4	3	80	20	4
IV	STS3-IV	Elective - II	4	3	80	20	4
PRACTICALS							
V	STS3-V	Practical – I Non-Parametric Inference, Quality Control and Optimization Techniques, Elective – I & II (NPI, QCOT, Elective – I & II / NPI + QCOT+E-I **for students who select DMMLT as elective - II**)	9	3	100	***	4
VI	STS3-VI	Practical – II (R+ TORA) / **E-II Project****	9	3	100	***	4
Total			34	***	520	80	24
Semester Total					600		

Electives to be offered in Semester III **:**

Elective – I:
1. Applied Regression Models **(ARM)**
2. Econometric Models **(EM)**

Elective – II:
1. Data Modeling using Machine Learning Techniques **(DMMLT)**
2. Advanced Design of Experiments **(ADE)**
3. Data Mining **(DM)**
4. Bayesian Inference **(BI)**

(*) Practical–I includes Elective-II practical's for those students who select **ADE/DM/BI** as Elective–II in Semester–III.
(**) Students who select **DMMLT** as Elective–II have **Project** instead of Practical–II in Semester-III.

M. Sc. (Statistics) Semester III

STS3 – I : Paper I - Non Parametric Inference (NPI)

Unit–I

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. Nonparametric methods for one-sample problems based on sign test, Wilcoxon signed Rank test, run test and Kolmogorov – Smirnov test.

Unit–II

Two sample problems based on sign test, Wilcoxon signed rank test for paired comparisons, Wilcoxon Mann-Whitney test, Kolmogorov – Smirnov Test, (Expectations and variances of above test statistics, except for Kolmogorov – Smirnov tests, Statements about their exact and asymptotic distributions), Wald–Wolfowitz Runs test and Normal scores test.

Unit–III

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. Ansari–Bradley test for two sample dispersions. Kruskal–Wallis test for one-way layour (K-samples). Friedman test for two-way layout (randomised block).

Unit–IV

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) Ferguson, T.S. – Mathematical Statistics, A decision theoretic approach (Academic press, 1967)
2) Gibbons – Non-parametric Statistical Inference (1978)
3) Myles Hollander and Douglas A. Wolfe: Nonparametric statistical methods (John Wiley and Sons)
4) Silverman: Density estimation for statistics and data analyses.

ADDITIONAL REFERENCE

1) W.J. Conover – Practical Non parametric Statistics (John Wiley)
2) Sidney Siegel – Non-parametric Statistics for Behavioural Science, Mc. Graw Hill.

M.Sc. (Statistics) Semester III

STS3 – II : 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.

Process Capability Analysis: Capability indices Cp, Cpk and Cpm, estimation, confidence intervals and tests of hypotheses relating to capability indices for normally distributed characteristics.

Unit–II

Acceptance sampling plans for attributes, single, double and sequential sampling plans and their properties; 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–I and its properties, modifications over CSP–I.

Unit–III

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.

ADDITIONAL REFERENCES

1. Ott,E.R. (1975) : Process Quality Control, McGraw Hill
2. Phadke, M.S. (1989): Quality Engineering through Robust Design, Prentice Hall.
3. Wetherill, G.B., and Brown, D.W: Statistical Process Control: Theory and Practice, Chapman and Hall.
4. Hillier F.S. and Leiberman,G.J.(1962) : Introduction to Operations Research; Holdon Day
5. Philips, D.T., Ravindran, A. and Solberg, J.(2000) : Operations Research principles and practice.

M.Sc. (Statistics) Semester III

STS3 -III (A) : Paper III (A) Elective I (A) – Applied Regression Models (ARM)

Unit–I

Introduction of selecting the best regression equation, all possible regression, backward and forward, stage, stepwise regression. 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), Statements of asymptotic properties, Non–linear growth models – Types of models – the Logistic model, the Gompertz model.

Unit–III

Logistic regression model – Introduction, Fitting the Logistic regression model, testing for the significance of the coefficients, Introduction to multiple Logistic regression, the multiple Logistic regression models, fitting the multiple logistic regression model, testing for the significance of the model.

Interpretation of the fitted Logistic regression model – Introduction, Dichotomous independent variable. Probit Analysis: Introduction, Analysis of Biological data, sigmoid curve, fitting a Probit Regression line through least squares method.

Unit–IV

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 (rreg).

Generalized Linear Models (GLIM)–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) Semester III

STS3 – III: Paper III (B) Elective I (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.

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

Unit–II

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

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

Unit–III

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.

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) :
   Chapter 1, Chapter 7: Section 7-1,7-3, Chapter 9 : Section 9-3, 9-4, Chapter 12 : Section 12-2,12-3, Chapter 13, Section 13-2,13-6
2) G. S. Maddala – Econometrics
   Chapter 1,chapter 9: Section 9-2,9-6, Chapter 10 : Section 10-1,10-2, Chapter 16 : Section 16-1,16-2
3) A. Koutsoyiennis – Theory of econometrics
   Chapter 9: Section 9-3.1,9-3.3,9-3.4,9-3.5, Chapter 10: Section 10-1,10-2, 10-3, 10-4, 10-5, 10-6.2,10-7,10-8.3,10-8.4, Chapter 11 : Section 11-4.2, Chapter 12 : 12-1,12-1.3,12-1.4, Chapter 16 : Section 16-1.1,16-1.216-3.1,16-3.2

M.Sc.(Statistics) Semester III

STS3 – IV : Paper IV(A) Elective II (A) - Data Modeling using Machine Learning Techniques (DMMLT)

Unit – 1

Introduction to data types, Measurement of scales, Understanding data with descriptive statistics and understanding the data with Visualization and data pre-processing (data cleaning, Outlier identification/outliers treatment, Identifying missing values/ missing value treatment, transformation)

Unit – 2

Introduction to statistical hypothesis concepts, Understanding relationship between variables using Parametric / Non Parametric tests (Correlations, Chi square , t-tests for proportions, t test for means and F tests. Non parametric tests like sign, Wilcoxon sign, rank test, Kruskal-Wallis test**,** Friedman test) ,data transformations (Standardize, Normalize, converting data from one scale to other scales) and Feature Selection Methods

Unit – 3

Introduction to Modeling concepts, review of the modeling process, Concepts of unsupervised and Supervised Modeling, detail approaches of unsupervised models (Hierarchical cluster analysis, K means cluster Analysis, data reduction techniques) and details approaches of supervised models (Linear regression, Multiple regression, Logistic, Multinomial logistic, DT(Decision Tress), NN (Neural Networks), SVM (Support vector Machine) and concepts of ensemble methods and detail approaches of Random forest, XG boosting

Unit – 4

Concepts of Model evolution, over fitting, under fitting, cross validation concepts, (train/test, K fold and Leave out one approaches), Model Performance concepts for classification techniques (classification matrix, Precision and Recall, F1 score, Sensitivity, Specificity, ROC curve) and Model performance concepts for regression (MSE, RMSE, R2, MAPE), Concepts of Model improvement (Tuning parameters using manual search, Manual grid search, random search) and saving models for future use.

Reference Books:-

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) Charu C Agrawal, Data Mining, Springer Publications
4) Trevor Hastie & Robert Tibshirani, An introduction to statistical learning with R, Springer Publications
5) Brett Lantz , Machine Learning with R, Packt Publications

Practicals:- Hands on training will be given on the techniques covered in theory with real life data.

M.Sc. (Statistics) Semester III

STS3 – IV: Paper IV(B) Elective II(B) – Advanced Design of Experiments (ADE)

Unit–I

Concept of General block design and its information matrix(c). 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. Youden Square design and its analysis.

Unit–II

Partially balanced incomplete block design with two–associate classes PBIBD(2)– Parametric relations, intra–block analysis, different association schemes. Lattice designs– Balanced lattice design, simple lattice design and their analysis.

Unit–III

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–IV

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. Norman Draper and Harry Smith: Applied Regression Analysis

ADDITIONAL REFERENCE

1. Joshi, D.D. : Linear Estimation and Design of Experiments
2. Myers, R.H. : Response Surface Methodology
3. Aloke Dey : Theory of Block Designs
4. Cornell, M : Mixture Experiments
5. Gardiner,W.P. and Gettinlsy,G. : Experimental Design Techniques in statistical Practice.

M.Sc. (Statistics) Semester III

STS3 – IV : Paper IV(C) Elective II (C) - 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, Support vector machine, Ensemble methods, Class imbalance problem – Multiclass problem

Cluster Analysis: Overview – K-means – Agglomerative hierarchical clustering, DBSCAN, 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, Evaluation of Association patterns, Effect of Skewed support distribution; Handling categorical attributes. Handling continuous attributes, Handling a concept hierarchy

Text Book:

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2008): “Introduction to Data Mining”, Pearson Education. (**Ch.1:** 1.2 to 1.4; **Ch. 2:** 2.1 to 2.4, **Ch. 3:** 3.3, 3.4; **Ch. 4:** 4.1 to 4.6; **Ch. 5:** 5.1 to 5.8; **Ch. 6:** 6.1 to 6.8; **Ch. 7:** 7.1 to 7.3, **Ch. 8:** 8.1 to 8.5)

References:

1. Arun K Pujari, Data Mining Techniques, University Press, 2nd Edn, 2009.
2. K.P. Soman, Shyam Diwakar, V.Ajay, Insight into Data Mining Theory and Practice, PHI, 2010.
3. Vikram pudi P. Radha Krishna , Data Mining, Oxford University Press, 1st Edition 2009
4. Galit S, Nitin RP, Peter C Bruce. Data Mining for Business Intelligence. Wiley India Edition,2007.

M.Sc. (Statistics) Semester III

STS3 – IV : Paper IV(D) Elective II (D) – 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. 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-III

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-IV

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.

ADDITIONAL REFERENCES

1. DeGroot M. H. Optimal Statistical Decisions. McGraw Hill.
2. Bernando J. M. and Smith, A. F. M. Bayesian Theory, John Wiley and Sons.
3. Robert, C. P. The Bayesian Choice: A decision Theoretic Motivation, Springer.
4. Gemerman, D. Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference, Chapman Hall.
5. Box, G. P. and Tiao, G. C. Bayesian Inference in Statistical Analysis, Addison - Wesley.

Note: (i) If this Elective is offered, Practical paper V nine hours will be distributed among remaining three papers equally

(ii) If this Elective is offered, Practical Paper V will have three sections. Each section consists of two questions from each paper. Candidate has to answer 3 questions choosing one from each section.

M.Sc.(Statistics) Semester III
STS3 – V : Paper V – Practical (NPI, QCOT, Elective-I, Elective-II)

Practical in Non-Parametric Inference, Quality Control Optimization Techniques, Elective-I and Elective-II

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.

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

Elective – I (A) 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) 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 – II (B) 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. Analysis of Mixture Experiments.

Elective – II (C) 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

(*) Practical–I includes Elective-II practical's for those students who select **ADE/DM** as Elective–II in Semester–III.
(**) Students who select **DMMLT** as Elective–II have **Project** instead of Practical–II in Semester-III.

M.Sc.(Statistics) Semester III
STS3 – VI : Paper VI – Practical (R, TORA)

Practical with R Package for the following topics.

1. Charts and Diagrams
2. Basic Statistics
3. Design of Experiments
4. Multivariate Analysis
5. Time Series Analysis
6. Parametric tests
7. Non–Parametric tests
8. Operations Research (TORA Package) (Practical's not covered at UG level in TORA)
9. Statistical Quality Control
10. Regression Analysis

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

M.Sc. STATISTICS
CB - SCHEME OF INSTRUCTION AND EXAMINATION
WITH EFFECT FROM 2019 – 2020

SEMESTER IV

Paper	Sub. Code	Paper Title	Instruction Hrs/ Week	Duration of Exam (in Hrs)	Max. Marks	IA and Assign.	Credits
THEORY							
I	STS4-I	Stochastic Processes (SP)	4	3	80	20	4
II	STS4-II	Time Series Analysis (TSA)	4	3	80	20	4
III	STS4-III	Elective - I	4	3	80	20	4
IV	STS4-IV	Elective - II	4	3	80	20	4
PRACTICALS							
V	STS4-V	Practical – I Stochastic Processes, Time Series Analysis and Elective – I & II (SP, TSA, Elec. I & II)	9	3	100	***	4
VI	STS4-VI	Practical – II SPSS / **E-II Project****	9	3	100	***	4
		Total	34	***	520	80	24
		Semester Total			600		

Electives to be offered in Semester IV **:**

Elective – I:
1. Reliability Theory **(RT)**
2. Actuarial Science **(ASC)**

Elective – II:
1. Text Analytics **(TA)**
2. Advanced Operations Research **(Adv.OR)**
3. Clinical Trials **(CT)**
4. Demography **(DGY)**

(*) Practical–I includes Elective-II practical's for those students who select **Adv. OR / CT/DGY** as Elective–II in Semester–
(**) Students who select **TA** as Elective–II have **Project** instead of Practical–II in Semester-IV.
(***) Foreign students will do project instead of Practical – II (SPSS) in Semester – IV.

M.Sc.(Statistics) Semester IV
STS4 – I : 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; classification of states of a Markov chain – recurrent, positive recurrent, null - recurrent and transient states. Period of a state.

UNIT – II

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. Stationary distribution of a Markov chain. Existence and evaluation of stationary distribution. Random walk and gambler's ruin problem.

UNIT – III

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.

Weiner process as limit of random walk. First passage time of the process.

UNIT – IV

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.

ADDITIONAL REFERENCES

1. Bhat, B.R. (2000): Stochastic Models: Analysis and applications – New Age International India.
2. Basu, A.K. (2003): Introduction to Stochastic Process, Narosa Publishing House.

M.Sc. (Statistics) Semester IV
STS4 – II : 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). 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–III

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.

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–IV

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

ADDITIONAL REFERENCE

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

M.Sc. (Statistics) Semester IV

STS4 -III (A) : Paper III (A) Elective I (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 theses cases.

Unit–III

Life Distribution: Survival function – Notion of ageing, 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.

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.

Chapter I – Section 1 to 4
II – Section 1 to 4
III – Section 1,2,4 and 5
IV – Section 1 to 4
VI – Section 1 to 3
VII – Section 1 to 3, Section 4.1,4.2

ADDITIONAL REFERENCES

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

M.Sc. (Statistics) Semester IV

STS4 -III (B) : Paper III (B) Elective I (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.
Life table and its relation with survival function examples, assumptions of fractional ages, some analytical laws of mortality, select and ultimate tables**.**

Unit–II

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. Elements of compound interest – Nominal and effective rates of interest, discount, accumulation factor and continuous compounding.

Unit–III

Multiple life functions, joint life and last survivor status, insurance and annuity benefits through multiple life functions, evaluation for special mortality laws.
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–IV

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.
7. Federation of Insurance Institutes study courses: mathematical basis of Life Assurance F.I.21 (Published by Federation if Insurance Institutes, Bombay).

M.Sc.(Statistics) Semester IV

STS4 – IV : Elective II (A) - Text Analytics (TA)

Unit-1

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 – 2

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 – 3

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 – 4

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

Reference Books:

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.

Practicals:- *Hands on training will be given on the techniques covered in theory with real life data.*

M.Sc. (Statistics) Semester IV

STS4 – IV : Paper IV(B) Elective II (B) – Advanced Operations Research (Adv. OR)

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.

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

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.

Linear Fractional Programming Problem and its applications.

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.

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

ADDITIONAL REFERENCES

1. Hillier F.S. and Leiberman,G.J.(1962) : Introduction to Operations Research; Holdon Day
2. Philips, D.T.,Ravindran,A. and Solberg,J.(2000) : Operations Research principles and practice.

M.Sc. (Statistics) Semester IV

STS4 – IV : Paper IV(C) Elective II (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. 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–II

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–III

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

Unit–IV

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

STS4 – IV : Paper IV(D) Elective II (D) – 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. Adjustment of age data - use of Whipple, Myer and UN indices. Population composition, dependency ratio.

Unit–II

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–III

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–IV

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 mateix.

REFERENCES

1. Bartholomew, D. J. (1982). Stochastic Models for Social Processes, John Wiley.
2. 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.

STS4 – V : Paper V – Practical – I (SP, TSA, Elective-I, Elective-II)

Practical in Stochastic Processes, time Series Analysis, Elective-I and Elective-II

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.

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

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 and Office premium
13. Assurances payable at the moment of death.

Elective – II (B) 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 – 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

Elective – II (D) Demography

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

(*) Practical–I includes Elective-II practical's for those students who select **Adv. OR / CT / DGY** as Elective–II in Semester–IV.

(**) Students who select **TA** as Elective–II have Project instead of Practical–II in Semester-IV.

Practical with SPSS Package for the following topics.

1. Charts and Diagrams
2. Basic Statistics
3. Design of Experiments
4. Multivariate Analysis
5. Time Series Analysis
6. Parametric tests
7. Non–Parametric tests
8. Statistical Quality Control
9. Regression Analysis

(***) Foreign students will do project instead of Practical – II (SPSS) in Semester – IV.