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

M.Sc. APPLIED STATISTICS
CBCS - 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	STAS1-I	Linear Algebra and Linear Models (LA and LM)	4	3	80	20	4
II	STAS1-II	Probability Theory (PT)	4	3	80	20	4
III	STAS1-III	Distribution Theory and Estimation Theory (DT and ET)	4	3	80	20	4
IV	STAS1-IV	Sampling Theory and Surveys (STS)	4	3	80	20	4
PRACTICALS							
V	STAS1-V	Statistical methods using Python Programming	9	3	100	***	4
VI	STAS1-VI	Linear Algebra, Linear Models, Distribution Theory, Estimation Theory and Sampling Theory and surveys (LA, LM, DT, ET, ST)	9	3	100	***	4
Total			34	***	520	80	24
Semester Total					600		

M.Sc. (Applied Statistics) Semester I

STAS1- I : Paper I - Linear Algebra and Linear Models (LA and LM)

UNIT – I

Vector Spaces with an inner product, Gram –Schmidt orthogonalization 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–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 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

Formulation of a linear model through examples. Estimability of a linear parametric function. Guass-Markov linear model, BLUE for Linear functions of parameters, relationship between BLUEs and linear Zero-functions. Gauss Markov theorem, Aitkens generalized least squares. Concept of Multicollinearity.

UNIT – IV

Simple Linear regression – precision of the estimated regression, examining the regression equation - lack of fit and pure error. Analysis of multiple regression model, estimation and testing of regression parameters, Sub-hypothesis. Testing a general linear hypothesis, Multiple and partial correlations - derivation and testing. Use of dummy variables in multiple regression. Polynomial regression- Use of orthogonal polynomials

REFERENCES

1. Graybill, F.A. (1983) : Matrices with applications in Statistics, 2nd ed., Wards worth.
2. Searle, S.R.(1982) : Matrix Algebra useful for Statistics, John Wiley & Sons.
3. Rao, C.R. and Mithra, S.K.(1971) : Generalized inverse of matrices and its applications, John Wiley & Sons.
4. Rao, A.R. and Bhimasankaram, P. (1992) : Linear Algebra, Tata McGraw Hill Publishing Co. Ltd.
5. Draper and Smith:Applied Regression Analysis ,John Wiley
6. Montgomery :Introduction to Linear Regression Analysis .John Wiley.
7. Searle, S.R.(1982) : Linear models, John Wiley & Sons.
8. Kshirsagar.A.M. (1972) : A Course in Linear Models.

M.Sc. (Applied Statistics) Semester I

STAS1-II : Paper II - Probability Theory (PT)

UNIT – I

Review axiomatic approach to Probability, Probability as a measure, conditional probability (and Baye's Theorem). Random Variable, distribution function and its properties. Riemann – Stieltjes integration, Statement of properties of Riemann – Stieltjes integrals, Examples. Expectations of functions of random variables – moments. Conditional expectation and conditional variances, applications (A list model, random graph, uniform priors, Polyas' urn model and Bose-Einstein distribution, mean time for patterns, the compound Poisson identity, the k-record values of discrete random variables).

UNIT – II

Characteristic function and its properties, Uniqueness theorem and Inversion theorem, examples. (Functions which can not be Characteristic functions). Statement of Levy's continuity theorem. Probability and moment inequalities **:** Chebychev's, Markov, Cauchy-Schwartz, Holder, Minkowsky, Liapunov and Jensen Inequalities.

UNIT – III

Sequence of random variables – Borel-Cantelli Lemma; Borel 0-1 law. Convergence of sequence of random variables – convergence in law; convergence in probability; convergence in quadratic mean; convergence with probability one (almost sure convergence); Their implications and/or counter implications; Slutzky's theorem and its applications. Statement of Glivenko-Cantelli lemma.

UNIT – IV

Weak law of large numbers – Bernoulli and Khintchine's WLLNs. Kolmogorov inequality. Strong law of large numbers – Borel's SLLNs. Kolmogorov's SLLNs for independent random variables and i.i.d. random variables, examples.

Central Limit Theorem – Demoviere-Laplace form of CLT, Levy-Lindeberg form of CLT, Liapunov's form of CLT and Statement of Lindberg – Feller form of CLT – examples.

REFERENCES

1. Bhat, B.R. (1985) : Modern Probability Theory – Wiley Eastern.
2. Rohatgi, V.K. (1993): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern
3. Ross, S.M (2004) : Introduction to Probability Models, 8th Edition (Chapter 3) – Academic Press
4. Chandra, T.K. and Chatterji D (2001) : A First Course in Probability, Narosa Publishing House
5. Milton and Arnold – Introduction to probability and Statistics (4th Edition)-TMH publication.

ADDITIONAL REFERENCE

1. Karlin, S and Taylor, S.J. (1975) : A First course in Stochastic Processes, Academic Press.

M.Sc. (Applied Statistics) Semester I

STAS1- III : Paper III - Distribution Theory and Estimation Theory (DT and ET)

UNIT – I

Cauchy, Lognormal, Weibull, Pareto, Laplace distributions and their properties. Compound distributions (Binomial and Poisson only). Truncated distributions (Poisson, Exponential and Normal distributions). Mixture Distributions. Bivariate Normal distribution.

UNIT – II

Functions of random variables and their distributions using Jacobian of transformations and Characteristic function. Sampling Distributions of Sample mean and variance, independence of $\overline{X}$ and S^2. Central t, F and χ^2 distributions and their properties. Non-central χ^2, t and F distributions and their properties (Statements only). Distributions of Quadratic forms under normality. Joint and Marginal Distributions of order statistics. Distributions of sample range.

UNIT – III

Concepts of point estimation – Criterion for good estimator, MVUE, Fisher's information, Cramer-Rao lower bound and its applications. Rao–Blackwell theorem, completeness, Lehmann – Scheff's theorem. Estimation of bias and standard deviation of point estimation by the Jackknife and Bootstrap methods with examples.

UNIT – IV

Method of moments, 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, 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.

REFERENCES

1. Rohatgi,V.K. (1984) : An Introduction to Probability theory and Mathematical Statistics, Wiley Eastern.
2. Dudewicz,E.J. and Mishra,S.N. (1988) : Modern Mathematical Statistics, Wiley International, Students Edition.
3. Parimal Mukhopadhya: Mathematical Statistics.
4. Milton and Arnold – Introduction to probability and Statistics (4^{th} Edition)-TMH publication.

ADDITIONAL REFERENCES

1. Ferguson, T.S. (1967) : Mathematical Statistics, A decision theoretic approach, Academic Press.
2. Rao,C.R.(1973) : Linear Statistical Inference and its applications,2/e, Wiley Eastern.
3. Johnson,S. and Kotz (1972) : Distribution in Statistics, Vol. I,II and III.
4. Lehman, E.L. (1983) : Theory of Point Estimation, John Wiley and Sons.

M.Sc. (Applied Statistics) Semester I

STAS1- IV : Paper IV - Sampling Theory and Surveys (ST)

UNIT – I

Review of SRSWR, SRSWOR, Stratified random sampling and Systematic Sampling.

Unequal probability Sampling – Probability proportional to size (PPS) sampling with and without replacements (ppswr / wor) methods - drawing samples using Cumulative total and Lahiri's methods. Horwitz -Thompson, Hansen – Horwitz and Yates and Grundy estimators for population mean, total and their variances.

UNIT – II

Ratio Method of Estimation - Concept of ratio estimators, Ratio estimators in SRS, their bias, variance/MSE. Ratio estimators in Stratified random sampling – Separate and combined estimators, their variances/MSE.

Regression method of estimation – Concept Regression estimators, Regression estimators in SRS with pre–assigned value of regression coefficient (Difference Estimator) and estimated value of regression coefficient, their bias, variance/MSE, Regression estimators in Stratified Random sampling – Separate and combined regression estimators, their variances/ MSE.

UNIT – III

Cluster Sampling - Cluster sampling with clusters of equal sizes, estimator of mean per unit, its variance in terms of intracluster correlation coefficient, determination of optimum sample and cluster sizes for a given cost. Cluster sampling with clusters of unequal sizes, estimator of population mean and its variance/MSE.

Sub sampling (Two–Stage only) - Equal first stage units – Estimator of population mean, variance/MSE, estimator of variance. Determination of optimum sample size for a given cost. Unequal first stage units – estimator of population mean and its variance/MSE.

UNIT – IV

Planning of Sample Surveys - Methods of data collection, problem of sampling frame, choice of sampling design, pilot survey, processing of survey data.

Non-sampling errors - Sources and treatment of non-sampling errors. Non – sampling bias and variance.

REFERENCES

1. Parimal Mukhopadhyay (1998) : Theory and methods of Survey sampling, Prentice – Hall of India, New Delhi.
2. Cochran, W.C. (1977) : Sampling Techniques, Third Edition, Wiley Eastern.
3. Daroga Singh and Chowdary (1986) : Theory and Analysis of Sample Survey Designs – Wiley Eastern Ltd.

ADDITIONAL REFERENCES

1. Des Raj (1976) : Sampling Theory, Tata McGraw Hill, New Delhi.
2. Sukhatme et. Al (1984): Sampling Survey methods and its applications, Indian society of Agricultural Statistics.
3. Murthy, M.N. (1967) : Sampling theory, Tata McGraw Hill, New Delhi.

M.Sc. (Applied Statistics) Semester I
STAS1- 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.(Applied Statistics) Semester I
STAS1-VI : Paper VI - Practical (LA, LM, DT, ET, ST)

PRACTICALS IN LINEAR ALGEBRA, LINEAR MODELS, DISTRIBUTION THEORY, ESTIMATION AND SAMPLING

LINEAR ALGEBRA

1. Inverse of a matrix by partition method.
2. Solutions of linear equations by sweep-out method.
3. Computation of Moore-Penrose inverse by Penrose method.
4. Computation of Generalized inverse of a matrix.
5. Formation of characteristic equation by using traces of successive powers.
6. Spectral decomposition of a square matrix of third order.

LINEAR MODELS

1. Fitting of a simple linear regression model - Computation of Pure error and lack of fit.
2. Fitting of Multiple Regression models with Two and Three Independent variables. and testing of regression parameters
3. Computation and Testing of Multiple Correlation coefficient.
4. Computation and Testing of Partial Correlation Coefficients.

DISTRIBUTION THEORY AND ESTIMATION

1. Distributions: Fitting of
    (i) Lognormal
    (ii) Weibull
    (iii) Cauchy
    (iv) Gamma with parameters
2. Estimation:
    a. Computation Jackknife estimator
    b. Computation of Bootstrap estimator
    c. Method of MLE (Scoring Method)
    d. Computation of Bayes estimator (Binomial)

SAMPLING THEORY

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

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

M.Sc. APPLIED STATISTICS
CBCS - 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	STAS2-I	Statistical Inference (SI)	4	3	80	20	4
II	STAS2-II	Applied Regression Analysis (ARA)	4	3	80	20	4
III	STAS2-III	Multivariate Data Analysis (MDA)	4	3	80	20	4
IV	STAS2-IV	Design of Experiments (DOE)	4	3	80	20	4
PRACTICALS							
V	STAS2-V	Statistical Inference and Applied Regression Analysis (SI, ARA)	9	3	100	***	4
VI	STAS2-VI	Multivariate Data Analysis and Design of Experiments (MDA, DOE)	9	3	100	***	4
Total			34	***	520	80	24
Semester Total					600		

M.Sc. (Applied Statistics) Semester II
STAS2 – I : Paper I - Statistical Inference (SI)

UNIT – I

Concepts of Hypothesis, Types of errors, Statistical test, critical region, test functions, randomized and non–randomized tests. Concepts of MP and UMP tests, Neymann – Pearson lemma and its applications to one parameter exponential family of distributions.

UNIT – II

Concepts of unbiased and consistent tests. Likelihood Ratio Criterion with simple applications (including homogeneity of variances). Statements of asymptotic properties of LR test. Confidence Intervals (based on fixed sample size and distributions for the parameters of Normal, exponential, Binomial, Poisson distributions). Relationship between confidence intervals and hypothesis testing. The concept of robustness in testing.

UNIT – III

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

UNIT – IV

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

REFERENCES

1. Rohatgi, V.K.: An Introduction to Probability Theory and Mathematical Statistics (Wiley Eastern)
2. Gibbons : Non Parametric Statistical Inference,(Tata Mc Graw Hill)
3. Myles Hooander and Douglas A. Wolfe – Non parametric Statistical methods (John Wiley and sons)
4. Wald,A. : Sequential Analysis (Dover Publications)
5. Milton and Arnold – Introduction to probability and Statistics (4^{th} Edition)-TMH publication.
6. Lehman, E. L. : Testing of hypothesis, John Wiey
7. Goon, Gupta and Das Gupta : Outlines of Statistics, Vol. II, World Press.

ADDITIONAL REFERENCES

1. C.R. Rao – Linear Statistical Inference (John Wiley)
2. W.J. Conovar – Practical Non parametric Statistics (John Wiley)

M.Sc (Applied Statistics) Semester II

STAS2 – II : Paper II - Applied Regression Analysis (ARA)

UNIT – I

Review of the general regression situation, extra sum of squares principle, orthogonal columns in the X – matrix, partial and sequential F-tests. Bias in regression estimates. Weighted least squares. Introduction to examination of residuals, overall plot, time sequence plot, plot against Y_i, predictor variables X_{ij} .Correlations and serial correlations among the residuals, Durbin Watson Test. Concept of outliers, Detecting of outliers, standardized residuals. Testing of outliers in linear models.

UNIT – II

Introduction of selecting the best regression equation, all possible regressions: backward, stepwise regression procedures. Variations on these methods. Stagewise regression procedures. Polynomial regression –use of orthogonal Polynomials. Ridge regression: Introduction, basic form of ridge regression, ridge regression on a selection procedure.

Robust regression: Introduction, Least absolute deviation regression(L_1-regression),M-Estimation Procedure, Least Median squares regression, ranked residuals regression(RREG).

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

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 and S), Statements of asymptotic properties, Non–linear growth models – Types of models – the Logistic model, the Gompertz model.

REFERENCES

1. Draper and Smith: Applied Regression Analysis- John Wiley
2. Dennis Cook. R and Sanford Weisberg (1999) Applied Regression Including Computing and Graphics –John Wiley
3. Galton: Applied Regression Analysis
4. Regression Analysis: Concepts and Applications, Franklin A. Graybill and Hariharan K. Iyer
5. Applied Regression Analysis, linear models and related methods: John Fox
6. Non–linear Regression Analysis and its Applications: Douglas M. Bates and Donald G. Watts
7. Applied Logistic Regression: David W. Hosme and Stanley Lemeshow.
8. Linear Models for unbalanced Data: Shayler Searle
9. Residuals and Influence in Regression: R. Dennis Cook and Sanford Weisberg
10. Log–linear models and Logistic Regression: Ronald Christensen.

M.Sc (Applied Statistics) Semester II

STAS2 – III : Paper III - Multivariate Data Analysis (MDA)

UNIT – I

Concept of Bivariate and multivariate random variables, concept of random vector, its expectation, and variance-covariance matrix, marginal and joint distributions, stochastic independence of random vectors, conditional distributions.

Multinomial Distribution and its properties (Marginal, Conditional, MGF, Ch.F, Correlation), Multivariate normal distributions and its properties (Marginal, Conditional, MGF, Ch.F, Correlation), Distribution of sample mean vector, Independence of sample mean vector and sample variance-covariance matrix, Maximum likelihood estimates of parameters (Mean vector and covariance matrix).

UNIT – II

Sample dispersion matrix, statement of Wishart distribution and its simple properties; Wilk's λ criterion and its distribution, statements of its properties; Hotelling's T^2 statistic, Null distribution of Hotellings' T^2 and Applications of Hotelling T^2 Statistic (single and two sample mean vector cases), Mahalanobis D^2 statistic,

UNIT - III

Concepts of Discriminant analysis, Computation of linear discriminant function using likelihood ratios based on Multivariate normal populations and Fisher's Linear Disciminent Function, Bayes Mis-classification, Relationship between Discriminant function and Mahalanobis D^2 statistic.

Path analysis and computation of path coefficients. Correspondence analysis. Introduction to multidimensional scaling, some related theoretical results, similarities, metric and non-metric multidimensional scaling methods.

UNIT – IV

Principal component analysis: Introduction, Derivation of Principal components and statements of its properties; Factor analysis: Introduction, simple factor model, Orthogonal factor model construction. Canonical variables and canonical correlations, Cluster analysis: Introduction, similarities and dissimilarities, Single, Complete and average linkage methods.

REFERENCES

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

M.Sc. (Applied Statistics) Semester II

STAS2 – IV : Paper IV - Design of Experiments (DOE)

UNIT – I

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

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

UNIT – III

Balanced incomplete block design (BIBD) – parametric relations, intra-block analysis, recovery of inter-block information. Partially balanced incomplete block design with two associate classes PBIBD (2) – Parametric relations, intra block analysis. Simple lattice design and Youden-square design.

UNIT – IV

Concept of Response surface methodology (RSM), the method of steepest ascent. Response surface designs. Design for fitting first – order and second – order models. Variance of estimated response. Second order rotatable designs (SORD), Central composite designs(CCD): Role of CCD as an alternative to 3^k design, Notatability of CCD.

REFERENCES

1. Das, M.N. and Giri,N.: Design and Analysis of Experiments, Wiley Eastern.
2. Montogomery, D.C. : Design and Analysis of Experiments, John Wiley.
3. Draper and Smith : Applied Regression Analysis, John Wiley.
4. Parimal Mukhopadhyay : Applied Statistics, New Central Book Agency.

ADDITIONAL REFERENCES

1. Cochran and Cox : Experimental designs, John Wiley.
2. Kempthrone : Desing and Analysis of Experiments, John Wiley.
3. Kapoor and Gupta : Applied Statistics, Sultan Chand.
4. Alok Dey : Theory of Block Desings, Wiley Eastern.

M.Sc. (Applied Statistics) Semester II
STAS2 – V : Paper V Practical (SI and ARA)

PRACTICALS IN STATISTICAL INFERENCE AND APPLIED REGRESSION ANALYSIS

STATISTICAL INFERENCE

1. Type I and Type II errors
2. MP tests
3. UMP tests
4. L.R. Tests
5. Wilcoxon Signed rank test
6. Wilcoxon Mann-Whitney test
7. Kolmogorov – Smirnov one sample, two sample tests
8. Ansari – Bradley test for two sample dispersion
9. Krusakal Walli's test for one way layout
10. Friedman test for two way layout
11. Normal Scores test
12. Kendall's Tau
13. SPRT procedures for
    - (i) Binomial
    - (ii) Poisson
    - (iii) Normal and computation of their OC function.

APPLIED REGRESSION ANALYSIS

1. Computation of residuals and their plots.
2. Computation and testing of Serial Correlation.
3. Computation of all possible regression for three variables using R^2.
4. Probit and Logit analysis
5. Step wise Regression for four variables
6. Forward selection for four variables
7. Backward elimination for four variables

M.Sc.(Applied Statistics) Semester II
STAS2 – VI : Paper VI Practical (MDA and DOE)

PRACTICALS IN MULTIVARIATE DATA ANALYSIS AND DESIGN OF EXPERIMENTS

MULTIVARIATE DATA ANALYSIS

1. MLE of parameters of multivariate normal distribution.
2. Computation of Hotellings T^2 and Mahalanobis D^2.
3. Computation Path coefficients.
4. Classification between two normal populations by discriminant analysis.
5. Computation of Principle Components.
6. Computation of canonical correlations
7. Estimating the factor loading in single factor model.
8. Computation of single linkage method.
9. Single linkage dendogram for dissimilarity matrix.

DESIGN OF EXPERIMENTS

1. Analysis of 2^3 and 2^4 factorial experiments.
2. Analysis of 3^2 factorial experiment.
3. Analysis of Total and partial confounding of 2^3 factorial design.
4. Analysis of one-half fraction of 2^4 design and one-quarter fraction of 2^5 design.
5. Analysis of Split-plot Design
6. Intra-block analysis of BIBD
7. Intra-block analysis of PBIBD(2)
8. Analysis of Youden-square design
9. Analysis of Simple Lattice design

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

M.Sc. APPLIED STATISTICS
CBCS - 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	STAS3-I	Operations Research–I (OR-I)	4	3	80	20	4
II	STAS3-II	Forecasting Models (FM)	4	3	80	20	4
III	STAS3-III	Elective - I	4	3	80	20	4
IV	STAS3-IV	Elective - II	4	3	80	20	4
PRACTICALS							
V	STAS3-V	Operations Research–I and Forecasting Models Elective – I & II (OR-I + FM Elective – I & II / OR-I + FM +E-I **for students who select DMMLT as elective - II**))	9	3	100	***	4
VI	STAS3-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. Reliability Theory **(RT)**
2. Actuarial Science **(ASC)**

Elective – II:

1. Data Modeling using Machine Learning Techniques **(DMMLT)**
2. Statistical Pattern Recognition **(SPR)**
3. Data Mining **(DM)**
4. Bayesian Inference **(BI)**

(*) Practical–I includes Elective-II practical's for those students who select **SPR/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. (Applied Statistics) Semester III

STAS3 - I : Paper – I : Operations Research–I (OR-I)

Unit–I

Definition and scope of OR: Phases in O.R.; Models and their solutions; decision making under uncertainty and risk.

Duality and complementary slackness theorem, primal dual relation; dual simplex algorithm;

Sensitivity Analysis: Introduction, definition of sensitivity analysis; discrete changes in requirement and cost vectors. Parametric Programming: Introduction, parameterization of cost and requirement vectors.

Unit–II

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

Sequencing and scheduling Problems: 2 machine n-job and 3 machine n-job problems with identical machine sequence for all jobs; 2-job n-machine problem with different machine problem with different routings.

Unit–III

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

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

Unit–IV

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

Stochastic Programming problem; analysis of chance constrained linear programming under zero order, non randomised decision rule, deterministic equivalents of chance constraints with reference to Normal and Cauchy distributions.

REFERENCES

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

ADDITIONAL REFERENCES

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

M.Sc. (Applied Statistics) Semester III

STAS3 - II : Paper II - Forecasting Models (FM)

Unit–I

Forecasting: The role of forecasting in decision-making, forecasting techniques. Smoothing Techniques: Simple Moving Averages, exponential smoothing and Winter's linear and seasonal exponential smoothing.

Stationary stochastic processes, Autocovariance and Autocorrelation functions and their estimation. Standard error of autocorrelation estimates. Bartlett's approximation (without proof). Periodgram, power spectrum and spectral density functions. Simple examples of autocorrelation and spectral density functions. Link between sample spectrum and auto-correlation 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 for the ARIMA models viz., difference equation, random shock and inverted forms.

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

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

Unit–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, probability limits of the forecasts at any lead time.

REFERENCES

1) Weel Wright, S.C. and Makridakis,S. (1973) : Forecasting methods for Management, John–Wiley & sons, New York.
2) Box, G.E.P. and Jankins,G.M.(1970) : Time series Analysis (Forecasting and control), Holden day publication.

ADDITIONAL REFERENCES

1. Anderson, T.W.(1971) : The statistical analysis of Time series, John Wiley, New York.
2. Brockwell,P.J. and Davis, R.A. : Time Series : Theory and methods(Second Edition), Springer-Verlag.

M.Sc. (Applied Statistics) Semester III

STAS3 -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 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. (Applied Statistics) Semester IV

STAS4 -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.(Applied Statistics) Semester III

STAS3 – IV : Elective II (A) - Data Modeling using Machine Learning Techniques (DMMLT)

Unit – I

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 – II

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 – III

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 – IV

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. (Applied Statistics) Semester III

STAS3 – IV: Elective II(B) – Statistical Pattern Recognition (SPR)

Unit - I

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

Unit – II

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

Unit – III

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

Unit – IV

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

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

REFERENCES

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

M.Sc. (Applied Statistics) Semester III

STAS3 – 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. (Applied 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 one paper. Candidate has to answer 2 questions choosing one from each section.

M.Sc. (Applied Statistics) Semester III
STAS3 - V : Paper V – Practical (OR-I, FM, Elective-I, Elective-II)

Operations Research–I

1. Solving an LPP by Dual Simplex Method
2. Solving an LPP by Revised Simplex
3. Sensitivity Analysis for cost and requirement vectors.
4. Parametric Programming for cost and requirement vectors.
5. Sequencing problem with 2 jobs n machine problem by graphical method.
6. Evaluation of project time through CPM and PERT
7. Time cost Analysis for CPM and PERT
8. Integer Programming Problem- Gomery's cutting plane method.

Forecasting Models

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

Elective – I (A) Reliability Theory

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

Elective – I (B) Actuarial Science

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

Elective – II (B) Statistical Pattern Recognition

1. Computation of Linear discriminant classifier function for two- multivariate normal classes.
2. Computation of Linear discriminant classifier function using Minimum Squared function for Binary data
3. Bayes classifier and the computation of its Error rate (Probability of Error)
4. Nearest Neighbor Classifier and computation of its error rate
5. Classification using Hidden Markov Model
6. Feature selection using Principal Component Analysis

Elective – II (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 **SPR/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.(Applied Statistics) Semester III
STAS3 – 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. APPLIED STATISTICS
CBCS - 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	STAS4-I	Statistical Process and Quality Control (SPQC)	4	3	80	20	4
II	STAS4-II	Applied Stochastic Processes (ASP)	4	3	80	20	4
III	STAS4-III	Elective - I	4	3	80	20	4
IV	STAS4-IV	Elective - II	4	3	80	20	4
PRACTICALS							
V	STAS4-V	Practical – I Statistical Process and Quality Control, Applied Stochastic Processes Stochastic Processes, and Elective – I & II (SPQC, ASP, Elec. I & II)	9	3	100	***	4
VI	STAS4-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. Operations Research – II **(OR – II)**
2. Econometric Models **(EM)**

Elective – II:
1. Text Analytics **(TA)**
2. Artificial Neural Networks **(ANN)**
3. Clinical Trials **(CT)**
4. Demography **(DGY)**

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

STAS4 - I : Paper - I : Statistical Process and Quality Control (SPQC)

Unit–I

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

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

.

Unit–II

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

Unit–III

Acceptance sampling plans for attribute inspection, single, double and sequential sampling plans and their properties; Rectifying sampling plans for attributes, AOQ, AOQL, designing of R.S.P. for specified AOQL and LTPD. 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–IV

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

Multivariate quality control, use of control ellipsoid and of utility functions. Concept of TQM, Six sigma.

REFERENCES

1) Montgomery, D.C.(1985) : Introduction to Statistical Quality Control, Wiley
2) Wetherill, G.B. (1977): Sampling Inspection and Quality Control, Halsted Press.
3) Cowden, D. J. (1960) : Statistical Methods in Quality Control, Asia Publishing House.

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.

M.Sc. (Applied Statistics) Semester IV

STAS4 – II : Paper II - Applied Stochastic Processes (ASP)

Unit–I

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

Unit–II

Continuous–time Markov Processes: Kolmogorov–Feller differential equations, Poisson process and birth and death processes.

Renewal Processes: Renewal process when time is discrete and renewal process with time is continuous, with examples. Renewal function, renewal density, limiting behaviour. Statement of elementary and basic renewal theorems.

Branching Processes: Examples of natural phenomena that can be modelled as a branching process. Probability of extinction; Statement of fundamental theorem of branching processes.

Note: Emphasis is only on statements of theorems and results and their applications.

Unit–III

Stochastic Processes in Biological Sciences: Markov models in population genetics; Recovery, relapse and death due to disease; cell survival after irradiation; compartmental analysis.

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

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

Unit–IV

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

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

REFERENCE

1. Bhat, U.N., (1984): Elements of Applied Stochastic Processes, John Wiley

ADDITIONAL REFERENCE

1. Ross, S. (1996): Stochastic Processes, Second Edition, John Wiley.
2. J. Medhi: Stochastic Processes.

M.Sc. (Applied Statistics) Semester IV

STAS4 – III : Paper III (A) Elective I (A) - Operations Research – II (OR – II)

Unit–I

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

Unit–II

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

Unit–III

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

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

Unit–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. U. N. Bhat: Introduction to Applied Stochastic Process.

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 (Applied Statistics) Semester IV
STAS4 – 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.(Applied Statistics) Semester IV

STAS4 – IV : Paper IV (a) Elective II (A) - Text Analytics (TA)

Unit - I

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

Unit – II

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

Unit – III

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

Unit – IV

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

STAS4 – IV : Paper IV (B) Elective II(B) – Artificial Neural Networks (ANN)

Unit – I

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

Unit – II

Perceptron Learning Algorithm, Derivation, Perceptron convergence theorem (statement); Multi-layer Perceptron Learning rule, limitations. Applications of the Perceptron learning. Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning.

Feed-forward and Fedd-back Back-Propagation Algorithms and derivation, learning rate, Momentum, Difficulties and Improvements. Bias and Variance. Under- Fitting and Over-Fitting

Unit – III

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

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

Unit – IV

Boltzman Machine and its learning rule, Hopfield model and its learning. Sigmoid belief network learning procedure, Stochastic machines,

Applications of ANN in Classification, Clustering, Regression, Time series forecasting

REFERENCES

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

M.Sc. (Applied Statistics) Semester IV

STAS4 – 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. (Applied Statistics) Semester IV

STAS4 – IV : Paper IV(D) Elective II (D) – Demography (DGY)

Unit–I

Coverage and content errors in demographic data, use of balancing equations and Chandra Sekharan - 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 matrix.

REFERENCES

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

M.Sc.(Applied Statistics) Semester IV

STAS4 – V : Paper V – Practical (SPQC, ASP, Elective-I, Elective-II)

Practical in Statistical Process and Quality Control, Applied Stochastic Processes, Elective-I and Elective-II

Statistical Process and Quality Control

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

Applied Stochastic Processes

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

Elective – I (A) Operations Research–II

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

Elective – I (B) 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) Artificial Neural Networks

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

Elective – II (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 **ANN / CT / DGY** as Elective–II in Semester–IV.

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

M.Sc.(Applied Statistics) Semester IV
STAS4 – VI : Paper VI Practical (SPSS)

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.