Department of Statistics, University College of Science, Osmania University
Two Year M.Sc. (Statistics) Programe for the Academic year 2018-2019
Scheme for Choice Based Credit System

Course : M.Sc. Statistics				Course : M.Sc. Statistics			
III Semester		**Teaching Hrs**	**Credits**	**IV Semester**		**Teaching Hrs**	**Credits**
1	Core (PI)	4	4	1	Core (NPI)	4	4
2	Core (OR)	4	4	2	Core (TS)	4	4
3	Elective - I	4	4	3	Elective – I	4	4
4	Elective - II	4	4	4	Elective - II	4	4
5	Practical – I (PI+OR) / (PI+OR+ E-I, **for students who select DMMLT as elective - II**)	9	4	5	Practical – I (NPI + TS+E-I+E-II) / (NPI+TS+E-I, **for students who select TA as elective - II**)	9	4
6	Practical – II (E-.I + E-II) / **Project in DMMLT, only for students who select DMMLT as elective - II**	9	4	6	Practical –II (SPSS) / **Project in TA, only for students who select TA as elective - II**	9	4
Total		**34**	**24**	**Total**		**34**	**24**

Paper Titles:

Semester III:

Paper - I : Parametric Inference **(PI)**
Paper - II: Operations Research **(OR)**
Elective – I:
1. Applied Regression Models **(ARM)**
2. Econometric Models **(EM)**

Elective – II:
1. Advance Design and Analysis of Experiments **(ADE)**
2. Reliability Theory **(RT)**
3. Data Modeling using Machine Learning Techniques **(DMMLT)**

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

Semester IV :

Paper - I : Non-Parametric Inference **(NPI)**
Paper - II: Time Series **(TS)**
Elective – I:
1. Statistical Quality Control **(SQC)**
2. Actuarial Science **(ASC)**

Elective – II:
1. Advanced Operations Research **(Adv. OR)**
2. Data Mining **(DM)**
3. Text Analytics **(TA)**

(*) Practical – I includes Elective-II practical's for those students who select Adv. OR /DM 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.***

Department of Statistics, University College of Science, Osmania University
Two Year M.Sc. (Applied Statistics) Programe for the Academic year 2018-2019
Scheme for Choice Based Credit System

Course : M.Sc. Applied Statistics				**Course : M.Sc. Applied Statistics**			
III Semester		**Teaching Hrs**	**Credits**	**IV Semester**		**Teaching Hrs**	**Credits**
1	Core (OR-I)	4	4	1	Core (SPR)	4	4
2	Core (SPQC)	4	4	2	Core (ASP)	4	4
3	Elective –I	4	4	3	Elective – I	4	4
4	Elective – II	4	4	4	Elective - II	4	4
5	Practical – I (OR-I+SPQC) / (OR-I+SPQC+E-I), **for students who select DMMLT as elective - II**)	9	4	5	Practical – I (SPR+ASP+ E-I + E-II) / (SPR+ASP+E-I, **for students who select TA as elective - II**)	9	4
6	Practical – II (E-I + E- II) / **Project in DMMLT, only for students who select DMMLT as elective - II**	9	4	6	Practical –II (SPSS) / **Project in TA, only for students who select TA as elective - II**	9	4
Total		**34**	**24**	**Total**		**34**	**24**

Paper Titles:

Semester III:

Paper - I: Operations Research **(OR-I)**
Paper - II: Statistical Process and Quality Control **(SPQC)**
Elective – I:
1. Forecasting Models **(FM)**
2. Econometric Models **(EM)**

Elective – II:
1. Reliability Theory **(RT)**
2. Data Modeling using Machine Learning Techniques **(DMMLT)**

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

Semester IV:

Paper - I : Statistical Pattern Recognition (SPR)
Paper - II: Applied Stochastic Process **(ASP)**
Elective – I:
1. Artificial Neural Networks **(ANN)**
2. Actuarial Science **(ASC)**

Elective – II:
1. Operations Research - II **(OR-II)**
2. Text Analytics **(TA)**

(*) Practical–I includes Elective-II practical's for those students who select OR-II 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.

UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD – 500 007

M.Sc. STATISTICS

CB - SCHEME OF INSTRUCTION AND EXAMINATION

WITH EFFECT FROM 2018 – 2019

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	Parametric Inference (PI)	4	3	80	20	4
II	STS3-II	Operations Research (OR)	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 (PI + OR) / (PI + OR) +Elective–I **for students who select DMMLT as elective - II)**	9	3	100	***	4
VI	STS3-VI	Practical – II Elective–I + Elective–II / **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. Advance Design of Experiments (ADE)
2. Reliability Theory (RT)
3. Data Modeling using Machine Learning Techniques (DMMLT)

(*) Practical–I includes Elective-I practical's for those students who select DMMLT 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 - 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 III

STS3 – II : Paper II - Operations Research (OR)

Unit–I

Definition and scope of Operations Research: Phases in OR; Models and their solutions.

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.

Unit–II

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

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 inter arrival. Distribution of death (departure) process, service time .Classification of Queuing models and solution of Queuing models; M/M/1:∞/FIFO and M/M/1:N/FIFO

Unit–III

Introduction to Simulation: Generation of random numbers from Uniform, Normal, Exponential, Cauchy and Poisson distributions; Estimating the reliability of the random numbers. Simulation to Queuing and Inventory problems.

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

Unit–IV

Inventory: Introduction; ABC analysis EOO Problem with and without shortage with (a) Production is instantaneous (b) Finite Constant rate (c) Shortages permitted.

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

REFERENCES

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

STS3 - III : 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: 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 (2^{nd} 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: Elective II(A) – 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 : Elective II (B) – 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–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 III

STS3 – IV : Elective II (C) - 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 – V : Paper V(A) – Practical (PI, OR)

Practical in Parametric Inference, Operations Research

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)

Operations Research

1. Dual by Simplex Method
2. Dual Simplex Method
3. Revised Simplex Method
4. Integer Programming Problem
5. Sensitivity Analysis
6. Parametric Programming Problem
7. Simulation
8. Simulation of Queuing and inventory problems
9. Evaluation of project time through CPM and PERT
10. Evaluation of Time cost analysis through CPM and PERT
11. Game theory

M.Sc.(Statistics) Semester III

STS3 – V : Paper V(B) – Practical (PI, OR, Elective-I)

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)

Operations Research

1. Dual by Simplex Method
2. Dual Simplex Method
3. Revised Simplex Method
4. Integer Programming Problem
5. Sensitivity Analysis
6. Parametric Programming Problem
7. Simulation
8. Simulation of Queuing and inventory problems
9. Evaluation of project time through CPM and PERT
10. Evaluation of Time cost analysis through CPM and PERT
11. Game theory

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.

(OR)

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

(*) Practical–I includes Elective-I practical's for those students who select DMMLT as Elective–II in Semester–III.

M.Sc.(Statistics) Semester III

STS3 – VI : Paper VI – Practical - II (Elective-I + Elective-II)

Practical in Elective-I + Elective-II

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 (A) 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 (B) 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.

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

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 IV

Paper	Sub. Code	Paper Title	Instruction Hrs/ Week	Duration of Exam (in Hrs)	Max. Marks	IA and Assign.	Credits
THEORY							
I	STS4-I	Non Parametric Inference (NPI)	4	3	80	20	4
II	STS4-II	Time Series Analysis (TS)	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, TS, 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. Statistical Quality Control (SQC)
2. Actuarial Science (ASC)

Elective – II:
1. Advanced Operations Research (Adv. OR)
2. Data Mining (DM)
3. Text Analytics (TA)

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

STS4 – 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

2) Ferguson, T.S. – Mathematical Statistics, A decision theoretic approach (Academic press, 1967)
3) Gibbons – Non-parametric Statistical Inference (1978)
4) Myles Hollander and Douglas A. Wolfe: Nonparametric statistical methods (John Wiley and Sons)
5) 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 IV
STS4 – II : Paper II - Time Series (TS)

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 : Elective I(A) Statistical Quality Control (SQC)

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. (Statistics) Semester IV

STS4 – III : 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 : Paper IV(A) Elective II (A) – 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(B) Elective II (B) - 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 IV

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

Unit-1

Introduction to Natural Language Processing Basics,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, Apress 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 III

STS4 – V : Paper V(A) – Practical- I (NPI, TS, Elective-I and Elective-II)

Practical on Non–Parametric Inference, Time Series, 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.

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
9.

Elective – I (A) Statistical Process and Quality Control

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

(OR)

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 (A) Advanced Operations Research

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

(OR)

Elective – II (B) 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

M.Sc.(Statistics) Semester IV
STS4 – VI : Paper VI – Practical - II (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. Operations Research (TORA Package)
9. Statistical Quality Control
10. Regression Analysis

(*) 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.

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 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	Statistical Process and Quality Control (SPQC)	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	Practical – I (OR–I,SPQC) / (OR–I, SPQC + Elective–I **for students who select DMMLT as elective - II**)	9	3	100	***	4
VI	STAS3-VI	Practical – II Elective–I + Elective–II / **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. Forecasting Models (FM)
2. Econometric Models (EM)

Elective – II:

1. Reliability Theory (RT)
2. Data Modeling using Machine Learning Techniques DMMLT)

(*) Practical–I includes Elective-I practical's for those students who select DMMLT 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 : 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 III

STAS3 - III : Paper III(A) Elective I (A) – 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: 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 III

STAS3 – IV: Paper IV(A) Elective II (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–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 III

STAS3 – IV : Elective II (B) - 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.(Applied Statistics) Semester III

STAS3 – V : Paper V(A) – Practical - I (OR-I, SPQC)

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.

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

M.Sc.(Applied Statistics) Semester III

STAS3 – V : Paper V(B) – Practical-II (OR-I, SPQC, Elective-I)

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.

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

Elective-I (A) 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.

(OR)

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

(*) Practical–I includes Elective-I practical's for those students who select DMMLT as Elective–II in Semester–III.

M.Sc.(Applied Statistics) Semester III

STAS3 – VI : Paper VI – Practical - II (Elective-I + Elective-II)

Practical in Elective-I + Elective-II

Elective-I (A) Forecasting Models

10. Moving Averages and exponential smoothing.
11. Generation of Time series by means of simple time series models.
12. Sample and theoretical correlograms.
13. Periodogram analysis.
14. Writing the models in B notation and stationarity and invertability of the models.
15. Classification of ARIMA models and computation of weights.
16. Identification AR, MA and ARMA models.
17. Estimation of parameters in AR, MA and ARMA models.
18. Computation of forecasts, updating and probability limits for forecasts.

(OR)

Elective – I (B) Econometric Models

9. Use of dummy variables (dummy variable trap) and seasonal adjustment
10. GLS estimation and predictors
11. Tests for heteroscedasticity.
12. Tests for Autocorrelations
13. Instruments variable estimation
14. Estimation with lagged dependent variable
15. Identification problems – Checking rank and order condition
16. Two SLS estimation

Elective – II (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.

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

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 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 Pattern Recognition (SPR)	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 Pattern Recognition, Applied Stochastic Processes, Elective – I & II (SPR, 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. Artificial Neural Networks (ANN)
2. Actuarial Science (ASC)

Elective – II:
1. Operations Research – II (OR – II)
2. Text Analytics(TA)

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

STAS4 – I: Paper – I : Statistical Pattern Recognition (SPR)

Unit–I

Basic concepts of pattern recognition. Fundamental problems in pattern recognition. Linear classifiers (Statistical approximation), Linear discriminant function for minimum squared error, L.D.F. for binary outputs; perception learning algorithm.

Unit–II

Nearest neighbour decision rules: description convergence, finite sample considerations, use of branch and bound methods.

Unit–III

Probability of errors: Two classes, Normal distribution, equal covariance matrix assumptions, Chernoff bounds and Bhattacharya distance, estimation of probability of error. Introduction to Hidden Markov Models (H.M.M.) and its applications.

Unit–IV

Feature selection and extraction: Interclass distance measures, discirmanant analysis, Probabilistic distance measures, Principal Components.

REFERENCES

1) R.O. Duda & H.E. Hart(1978): Pattern Recognition and scene analysis, Wiley
2) J.T. Ton and R.C. Gonzalez (1974) : Pattern Recognition Principles, Addison Wesley Publishing Company
3) G.J. McLactilan (1992): Discriminant Analysis and Statistical Pattern Recognition, Wiley
4) B.D. Ripley (1996) : Pattern Recognition & Neural Networks, Cambridge University Press.
5) Duda, Hast & Strok: Pattern Recognition.

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 : Elective I (A) – Artificial Neural Networks (ANN)

Unit – I

Biological Neuron, Biological and Artificial Neuron Models, Characteristics of ANN, McCulloch-Pitts Model, Historical Developments, Potential Applications of ANN. Types of Neuron Activation Function, ANN Architectures, Classification Taxonomy of ANN, Learning Strategy (Supervised, Unsupervised, Reinforcement) and Learning Rules.

Unit – II

Gathering and partitioning of data for ANN and its pre and post processing. Single Layer Feed Forward Neural Networks: Perceptron Models, Hebbian Learning and Gradient Descent Learning. Limitations and applications of the Perceptron Model. Multilayer Feed Forward Neural Networks: Generalized Delta Rule, Back propagation (BP) Training Algorithm, Learning rate, Momentum and Conjugate Gradient Learning, Difficulties and Improvements. Bias and Variance. Under- Fitting and Over-Fitting.

Unit – III

Radial Basis Function Networks: Introduction, Algorithms and Applications. . Approximation properties of RBF. Self Organizing Maps: Fundamentals, Algorithms and Applications.

Unit – IV

Applications of ANN in classification, clustering, regression, time series forecasting, variable selection and dimensionality reduction.

REFERENCES

1) Bishop, C. (1995). Neural Networks for Pattern Recognition. Oxford: University Press. Extremely well-written but requires careful reading, putting neural networks firmly into a statistical context.
2) Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. New York: Macmillan Publishing. A comprehensive book and contains a great deal of background theory.
3) Ripley, B.D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press. A very good advanced discussion of neural networks, firmly putting them in the wider context of statistical modeling.
4) Neural Networks Chapter in www.statsoft.com

ADDITIONAL REFERENCES

1) Carling, A. (1992). Introducing Neural Networks. Wilmslow, UK: Sigma Press.
2) Fausett, L. (1994). Fundamentals of Neural Networks. New York: Prentice Hall.
3) Patterson, D. (1996). Artificial Neural Networks. Singapore: Prentice Hall.
4) Kishan Mehrotra, Chilukuri K. Mohan and Sanjay Ranka(1996). Elements of Artificial Neural Networks:The MIT Press.

M.Sc. (Applied Statistics) Semester IV

STAS4 - III : 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 IV

STAS4 – IV : Elective II (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.(Statistics) Semester IV

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

Unit-1

Introduction to Natural Language Processing Basics,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, Apress 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 III

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

Practical in Statistical Pattern Recognition, Applied Stochastic Processes, Elective-I and Elective-II

Elective – I (A) Statistical Pattern Recognition

1. Linear Classifiers using LDF
2. Binary outputs using LDF
3. Probability of Errors – Normal distribution with equal covariance matrix
4. Hidden Markov Model
5. Feature relation using P.C.A.

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) Artificial Neural Networks

1. Forward propagation
2. Backward propagation
3. Classification
4. Clustering
5. Regression
6. Time Series

Note : 1and 2 by manual computations and 3 to 6 by using Neuro Solutions/SPSS

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 (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

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

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. Operations Research (TORA Package)
9. Statistical Quality Control
10. Regression Analysis

(*) 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.