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

M.Sc. APPLIED STATISTICS III-SEMESTER

Paper	Sub. Code	Paper Title	Credits	Instruction Hours per Week	Semester end Exam duration	Max. Marks in Semester end Exam	Max. Marks in the Internal Assessment and Assignments
THEORY PAPERS							
I	STAS-30I	Operations Research (OR)	3	3	3	70	20+10
II	STAS-302	Forecasting Models (FM)	3	3	3	70	20+10
III	STAS-303 Elective – I	A. Reliability Theory (RT) B. Actuarial Science (ASC) C. Econometric Models (EM) D. Statistical Process for Data Science (SPDS)	3	3	3	70	20+10
IV	STAS-304 Elective – I	A. Data Mining Techniques (DM) B. Bayesian Inference (BI) C. Statistical Pattern Recognition **(SPR)** D. Advanced Machine Learning Techniques (AMLT)	3	3	3	70	20+10
PRACTICAL PAPERS							
V	STAS-305	Operations Research & Forecasting Models (OR + FM)	4	2	2	50	-
VI	STAS-306	Elective-I & II	4	2	2	50	-
VII	STAS-307	Statistical Analysis using R & TORA	4	2	2	50	-
VIII	STAS-308	Data Analysis Project (Mini project)	4	2	2	50	-
Semester Total			34	***	***	480	120

M.SC. (APPLIED STATISTICS) III-SEMESTER

STAS-301: PAPER-I : OPERATIONS RESEARCH (OR)

Unit–I

General Linear programming problem, Graphical method, Simplex method, Big-M and two phase methods, Concept of Duality, Primal-Dual relation; Dual Simplex algorithm;

Unit–II

Sensitivity Analysis: Introduction, definition of sensitivity analysis; discrete changes in requirement and cost vectors. Parametric Programming: Introduction, parameterization of cost and requirement vectors. Integer Programming Problem: Gomory's cutting plane algorithm for pure and mixed IPP; Branch and bound Technique. Fractional and O-1 Knapsack problems, Sequencing Problems: 2 machine n-job and 3 machine n-job problems;

Unit-III

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

Unit–IV

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.

Unit–V

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

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.

M.SC. (APPLIED STATISTICS) III-SEMESTER

STAS-302: 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).

Unit–II

Periodgram, power spectrum and spectral density functions. Simple examples of autocorrelation and spectral density functions. Link between sample spectrum and auto-correlation function.

Linear Stationary Models: Two equivalent forms for the general linear process. Autocovariance generating function and spectrum. Stationarity and invertibility conditions for a linear process.

Unit-III

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–IV

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–V

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

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE-I(A) : STAS-303(A): PAPER III (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.

Unit-V

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

M.SC. (APPLIED STATISTICS) III- SEMESTER IV

ELECTIVE I(B) : STAS-303(B): PAPER III (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.

Unit–II

Life table and its relation with survival function examples, assumptions of fractional ages, some analytical laws of mortality, select and ultimate tables**.** 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.

Unit–III

Elements of compound interest: Nominal and effective rates of interest, discount, accumulation factor and continuous compounding. Multiple life functions, joint life and last survivor status, insurance and annuity benefits through multiple life functions, evaluation for special mortality laws.

Unit-IV

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–V

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.

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE I(C) : STAS-303(C): PAPER-III (C): 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.

Unit–II

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

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

Unit–III

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

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.

Unit–V

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) :
2) G. S. Maddala – Econometrics
3) A. Koutsoyiennis – Theory of econometrics

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE I(D): STAS-303 (D): PAPER-III (D) : STATISTICAL PROCESS IN DATA SCIENCE (SPDS)

Unit – I

Data visualization: Data types, Measurement of scales, understanding data with descriptive statistics. Data visualization techniques: Pictogram, Pie Chart, Bar Chart, Histogram, Line plot, frequency curves& polygons, ogive curves, Scatter Plot, Gantt Chart, Heat Map, Box and Whisker Plot, Waterfall Chart, Area Chart, Stacked Bar Charts - Sub Plots – Matplotlib, Seaborn Styles, Box plot - Density Plot - Tree map - Graph Networks. Visual Perception and Cognition, Applications of Principles of Information Visualization, Dashboard Design.

Unit-II

Data Pre-processing: Understanding data with Descriptive statistics. Data pre-processing steps, Data transformations (Standardize, Normalize, converting data from one scale to other scales). Identification suitable basic statistical tools / tests Parametric tests (z-, χ^2, t-, F-tests), Nonparametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, K-S, Wald-Wolfowitz run test) for the data sets. Feature selection methods

Unit-III

Introduction to Data Modelling: Review of the modelling process, Concepts of Classification & Clustering, Supervised and Un-supervised Modelling, Concepts of Model evolution, Cross validation concepts, (train/test, K fold and Leave out one approaches), Model Performance evaluation for Qualitative and Quantitative data, Model improvement and saving models for future use (classification matrix, Precision and Recall, F1 score, Sensitivity, Specificity, ROC curve) and Model performance concepts for regression (MSE, RMSE, R^2, adj R^2, MAPE),

Unit-IV

Concepts of Model improvement (Tuning parameters using manual search, Manual grid search, random search) and saving models for future use. Simple linear regression and its analysis (model fitting, regression ANOVA, testing lack of fit, MSE, RMSE, R^2, adj R^2, testing regression coefficients and confidence limits).

Unit-V

Basic concepts on Multivariate data; Simple, Partial & Multiple correlations; Multi collinearity; Multiple linear regression and its analysis; Selection of best linear regression (over fitting & under fitting) & its methods in outline (all possible, forward, backward, step-wise and stagewise). Simple and Multiple Logistic models fitting and its analysis.

REFERENCES

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) Foster Provost & Tom Fawcett, Data science for Business, O'REILLY Publications
4) Henrik Brink, Joseph W. Richards. Mark Fetherolf, Real World Machine Learning, Manning Publications
5) Brett Lantz , Machine Learning with R, Packt Publications

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE-II (A): STAS-304(A): PAPER-IV(A) : 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, Perceptron classifier, Support vector machine, Ensemble methods, Class imbalance problem – Multiclass problem

Unit-IV

Cluster Analysis: Agglomerative hierarchical clustering, K-means, DBSCAN, C4.5, CART 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.

Unit-V

Evaluation of Association patterns, Effect of Skewed support distribution; Handling categorical attributes. Handling continuous attributes, Handling a concept hierarchy.

REFERENCES

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2008): “Introduction to Data Mining”, Pearson Education.
2. Arun K Pujari, Data Mining Techniques, University Press, 2nd Edn, 2009.
3. K.P. Soman, Shyam Diwakar, V.Ajay, Insight into Data Mining Theory and Practice, PHI, 2010.
4. Vikram pudi P. Radha Krishna, Data Mining, Oxford University Press, 1st Edition 2009
5. Galit S, Nitin RP, Peter C Bruce. Data Mining for Business Intelligence. Wiley India Edition,2007.

M.SC. (APPLIED STATISTICS) III-SEMESTER

ELECTIVE-II (B): STAS-304(B): PAPER IV(B): 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.

Unit-III

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-IV

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-V

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.
4. Box, G. P. and Tiao, G. C. Bayesian Inference in Statistical Analysis, Addison - Wesley.

M.SC.(APPLIED STATISTICS) III-SEMESTER

ELECTIVE-II (C): STAS-304 (C): PAPER-IV (C) : 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.

Unit-V

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 Johnson Baugh and Steve Jost (2009): Pattern Recognition and Image Analysis, PHI.
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) III-SEMESTER

ELECTIVE-II(D) : STAS-304(D): PAPER-IV(D) : ADVANCED MACHINE LEARNING TECHNIQUES (AMLT)

Unit – I

Basic Concepts to Statistical Pattern Recognition, Pattern Recognition System, Fundamental problems in Pattern Recognition. Linear classifies, Multiple Linear regression, Logistic regression, Linear Discriminant Function (for binary outputs) with minimum squared error, Naïve Bayes classifier, Support Vector Machines, KNN algorithm

Unit – II

Decision Tree algorithms, Random Forest algorithm, Bagging Gradient boosting, Ada-Boosting and XG-Boosting algorithm, Market-Basket Analysis.

Unit – III

Cluster Analysis: Introduction, similarities and dissimilarities, Hierarchical clustering, Single linkage method, k-means and k-Nearest Neighbourhood (KNN) clustering,

Unit – IV

Introduction to Artificial Neuron Networks and its characteristics; Algorithms of Perceptron Learning; Multi-layer Perceptron Learning, Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning. Back-Propagation and their applications.

UNIT – V

Reinforcement learning, Markov Decision Process, Hidden Markov Model, Convolutional Neural Networks, Recurrent Neural Networks, Long-Short Term Memory Networks.

REFERENCES

1. Shai Shalev-Shwartz, Shai Ben-David Understanding Machine Learning: From Theory to Algorithms, Cambridge University press.
2. Marc Peter Deisenroth, A Aldo faisal, Cheng soon Ong: “Mathematics for Machine Learning”, Cambridge University Press, First Edition.
3. Hayes: Artificial Neural networks
4. Henrik Brink, Joseph W. Richards. Mark Fetherolf, Real World Machine Learning, Manning Publications
5. Trevor Hastie & Robert Tibshirani, An introduction to statistical learning with R, Springer Publications
6. Brett Lantz , Machine Learning with R, Packt Publications

M.SC. (APPLIED STATISTICS) III-SEMESTER

STAS-305 : PAPER V – OPERATIONS RESEARCH AND FORECASTING METHODS

PRACTICAL-I

SECTION-A: OPERATIONS RESEARCH

1. Solving an LPP by Simplex Method
2. Solving an LPP by Big M Method
3. Solving an LPP by Two Phase Method
4. Solving an LPP by Revised Simplex
5. Solving an LPP by using its Duality.
6. Solving an LPP by Dual Simplex Method
7. Sensitivity Analysis for cost and requirement vectors.
8. Parametric Programming for cost and requirement vectors.
9. Sequencing problem with 2 machines n- jobs and 3 machines n- jobs
10. Integer Programming Problem- Gomery's cutting plane method.
11. Evaluation of project time through CPM and PERT
12. Time cost Analysis for CPM and PERT

SECTION-B: 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.

M.SC. (APPLIED STATISTICS) SEMESTER III

STAS-306 : PAPER-VI: SECTION-A (ELECTIVE-I) & SECTION-B (ELECTIVE-II)

PRACTICAL-II

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. Assurances payable at the moment of death.

ELECTIVE – I(C): 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-I(D): STATISTICAL PROCESS FOR DATA SCIENCE

Data sets of Kaggle.com can be used for practice. For example, few of the them are: Iris Dataset; flights.csv Dataset; Sustainable Development Data; Credit Card Fraud Detection; Employee dataset; Heart Attack Analysis & Prediction Dataset; Dataset for Facial recognition; Covid_w/wo_Pneumonia Chest Xray Dataset; Groceries dataset; Financial Fraud and Non-Fraud News Classification; IBM Transactions for Anti Money Laundering
List of Practicals in Data Handling with Python:

1. Understanding data with Data types, Measurement of scales, descriptive statistics and data pre-processing steps.
2. Data transformations (Standardize, Normalize, converting data from one scale to other scales).
3. Parametric tests (z-, χ^2, t-, F-tests, ANOVA), Correlation & Regression etc.
4. Non-Parametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, Run test).
5. Applying the modelling process, Model evolution, over fitting, under fitting, cross validation concepts, (train/test, K fold and Leave out one approaches),
6. Evaluation of Model Performance for classification techniques for qualitative and Quantitative data.
7. Drawing One dimensional diagrams (Pictogram, Pie Chart, Bar Chart,).
8. Drawing two-dimensional (Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot)
9. Drawing Gantt Chart, Heat Map, Box - Whisker Plot, Correlation Matrices.

ELECTIVE–II(A): 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.

ELECTIVE–II(B): BAYESIAN INFERENCE

1. Simulation of the data based on the sample drawn from Normal, Binomial, Poisson, Beta, Exponential, Gamma.
2. Estimation of parameters in Bayesian approach based on the sample using Gibbs sampler under various priors and posterirs (Normal, Binomial, Poisson, Beta with conjugate priors)
3. Estimation of parameters in Bayesian approach based on the sample using Metro Pollis hasting under various priors and posterirs (Normal, Binomial, Poisson, Beta)

ELECTIVE-II(C): 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(D): ADVANCED MACHINE LEARNING TECHNIQUES

1. Implement and demonstrate the use of set of training data samples. Read the training data from a .CSV file.
2. Write a program to demonstrate the working of the decision tree-based ID-3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
3. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
4. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
5. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.
6. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.
7. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.
8. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.
9. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Note: Consider the data sets in https://www.kaggle.com/datasets

M.SC. (APPLIED STATISTICS) III-SEMESTER

STAS-307: PAPER-VII: DATA ANALYSIS USING R-SOFTWARE PRACTICAL-III

Practical with R-software using proper data sets available in https://www.kaggle.com/datasets

1. **Data visualization**: Pie Chart, Bar Chart, Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot, Gantt Chart, Heat Map, Box and Whisker Plot, Waterfall Chart, Area Chart, Box plot, Density Plot, Tree map, Graph Networks.
2. **Descriptive Statistics:** Measures of Central Tendencies, Dispersions, Relative measures of Dispersions, Percentile, Deciles, Moments, Skewness, Kurtosis. (Grouped and ungrouped data sets)
3. **Design & Analysis of Experiments:** Analysis of variances for CRD, RBD, LSD, Factorial experiments (2^2, 2^3, 2^4, 2^5 FE with and without confounding), BIBD, PBIBD.
4. **Multivariate Analysis:** estimation of Mean vector and covariance matrix, Hotelling T^2, Mahanobis D^2, linear Discriminant analysis, Principal Component analysis, Factor analysis, multi-dimensional scaling, Cluster analysis (Agglomerative and k-means),
5. **Time Series Analysis:** Computation of auto covariance, auto correlations, Smoothening, model fitting, AR, MA, ARMA and ARIMA models and their analysis
6. **Parametric tests**: Testing means, variances, proportions for single, two and more than two sample cases; Testing significance of correlation coefficient and regression coefficients,
7. **Non–Parametric tests:** Sign test, Wilxon Sign Rank test, Mann-Whitney U-test, Run test, Kolmogorov Smirnov test, chisquare test for goodness of fit and independence, Kendal tau,
8. **Regression Analysis:** Fitting simple and multiple linear regression models, Analysis, model Lack of fit, R^2, Adj R^2, selection best linear regression using all possible, forward, backward, stepwise and stage wise methods. Simple and multiple logistic regression models, Probit analysis.

M.SC. (APPLIED STATISTICS) SEMESTER III

STAS-308: PAPER-VIII: DATA ANALYSIS MINI PROJECT

Objectives and Outcomes:

1. To familiarize tools and techniques and content for presentation
2. To enhance practical presentation, effective communication and professional skills
3. To expose the students to answer the queries raised on the topic of presentation.
4. To encourage students to work with innovative and entrepreneurial ideas
5. Demonstrate the ability to synthesize and apply the knowledge and skills acquired in the academic program to real-world problems
6. Evaluate different solutions based on economic and technical feasibility
7. Effectively plan a project and confidently perform all aspects of project management
8. Demonstrate effective written and oral communication skills

PROJECT GUIDELINES:

1. The Head of Department will appoint Internal supervisor to Guide the students in each group.
2. Each group should consist of Five students.
3. Each student in the group must actively participate and report to the internal supervisor.
4. Each group has to search for the internship from any industry/ institution, if not found they have to choose a project with the help of supervisor allotted such that, the aim of project work is to develop solutions to realistic problems applying the knowledge and skills obtained on the courses studied with specializations, new technologies and current industry practices.
5. Each student has to give minimum two seminars, one in the second week ("Project Design Seminar") another on 8^{th} week (project progress seminar).
6. Submit Title of the project and one page abstract /synopsis about the project in the first week to the Head, forwarded by the internal supervisor.
7. Each project should give a 30 minutes presentation using power point presentation and followed by 10 minutes of discussion.
8. Project seminar presentations should contain, source of the data, Sample data, data description, literature survey on the similar studies, objectives of the study, Methodology, statistical techniques, work plan etc. and details of progress of the work, individual roles and their work distribution and their plan etc.
9. Each group Project Report should follow the Ph.D. thesis norms with Plagiarism report and each group has to submit two copies duly signed by the Students, Supervisor, industry certificate (if exists) and Head of the Department on before the last instruct date of the semester.
10. Project Marks will be awarded based on all stages of the project and the topic chosen, seminar presentation, communication skills, role/ contribution of the student in the project etc and viva-voce conducted by the internal & External examiners.

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

M.Sc. APPLIED STATISTICS IV- SEMESTER

Paper	Sub. Code	Paper Title	Credits	Instruction Hours per Week	Semester end Exam duration	Max. Marks in Semester end Exam	Max. Marks in the Internal Assessment and Assignments
THEORY							
I	STAS4-I	Statistical Process and Quality Control (SPQC)	3	3	3	70	20+10
II	STAS4-II	Applied Stochastic Processes (ASP)	3	3	3	70	20+10
III	STAS4-III Elective - I	A) Operations Research-II (OR-II) B) Text Analytics (TA**)** C) Demography (DGY)	3	3	3	70	20+10
IV	STAS4-IV Elective - II	**A)** Artificial Neural Networks (ANN) B) Design & Analysis of Algorithms (DAA) C) Clinical Trails (CT)	3	3	3	70	20+10
PRACTICALS							
V	STAS-405	Statistical Process and Quality Control, Applied Stochastic Processes.	2	4	2	50	-
	STAS-406	Elective-I & Elective-II	2	4	2	50	-
VI	STAS-407	**Major Project**	4	8	2	**100**	-
Semester Total			34	***	***	480	120

M.SC. (APPLIED STATISTICS) SEMESTER IV

STAS-401: 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.

Unit–II

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–III

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–IV

Acceptance sampling plans for attribute inspection, single, double and sequential sampling plans and their properties; Rectifying sampling plans for attributes, AOQ, AOQL, designing of RSP. for specified AOQL and LTPD. Plans for inspection by variables for one–sided and two–sided specifications; Dodges Continuous sampling Plan–l and its properties modifications over CSP-1.

Unit–V

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.
4. Ott,E.R. (1975) : Process Quality Control, McGraw Hill
5. Phadke, M.S. (1989): Quality Engineering through Robust Design, Prentice Hall.
6. Wetherill, G.B., and Brown, D.W: Statistical Process Control: Theory and Practice, Chapman and Hall.

M.SC. (APPLIED STATISTICS) SEMESTER IV

STAS-402: 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.

Unit–III

Branching Processes: Examples of natural phenomena that can be modelled as a branching process. Probability of extinction; Statement of fundamental theorem of branching processes. Stochastic Processes in Biological Sciences: Markov models in population genetics; Recovery, relapse and death due to disease; cell survival after irradiation; compartmental analysis.

Unit-IV

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–V

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

M.SC. (APPLIED STATISTICS) SEMESTER IV

ELECTIVE-I: STAS-403(A): PAPER III (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, 2xm, mx2, mx n games

Unit-IV

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–V

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.

M.SC.(APPLIED STATISTICS) IV-SEMESTER

ELECTIVE I(B): STAS-403:: PAPER III (B): 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),

Unit-V

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

REFERENCES

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.

M.SC. (STATISTICS) SEMESTER IV

ELECTIVE-I(C): STAS-404(C): PAPER III(C) : DEMOGRAPHY (DGY)

Unit–I

Coverage and content errors in demographic data, use of balancing equations and Chandrasekharan, Deming formula to check completeness of registration data.

Unit-II

Adjustment of age data - use of Whipple, Myer and UN indices. Population composition, dependency ratio.

Unit–III

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–IV

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–V

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

ELECTIVE-II(A): STAS-404(A): 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.

Unit-IV

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 Quantization,

Unit-V

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

ELECTIVE II(C): STAS-404 (C): PAPER IV(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.

Unit-II

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–III

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–IV

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

Unit–V

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

ELECTIVE-II: STS-404(B): PAPER IV(B): DESIGN AND ANALYSIS OF ALGORITHMS

UNIT I

Introduction to Algorithms: Algorithm, Time & space complexity, Asymptotic Notations. Writing pseudocode, Design Techniques.

Divide and Conquer: Control Abstraction, Binary Search, Finding the Maximum and Minimum, Merge Sort; Quick Sort, Selection sort, Strassen's Matrix Multiplication, Convex Hull.

UNIT-II

Greedy Method: Control Abstraction, Knapsack Problem, Job Sequencing with Deadlines, Minimum-Cost Spanning Trees (Kruskal's & Prim's), Single Source Shortest Paths (Dijkstra's).

Dynamic Programming: Control Abstraction, Multistage Graphs, All-Pairs Shortest Paths, Single-Source Shortest Paths, Optimal Binary Search Trees, 0/1 Knapsack, Traveling Salesperson Problem.

UNIT-III

Basic Traversal and Search Techniques: Techniques for Binary Trees, Techniques for Graphs, Connected Components and Spanning Trees, Biconnected Components and DFS.

UNIT -IV

Back Tracking: Control Abstraction, , 8-Queens Problem, Sum of Subsets, Graph Colouring, Hamiltonian Cycles, Knapsack Problem.

Branch-Bound: Control Abstraction, 0/1 Knapsack Problem, Traveling Sales Person problem.

UNIT -V

NP-Hard and NP-Complete Problems: Basic Concepts, Cook's Theorem, NP-Hard. Graph Problems, NP-Hard Scheduling Problems, NP-Hard Code Generation, Some Simplified NP-Hard Problems.

REFERENCE BOOKS

1. E Horowitz, S Sahni, S Rajasekaran, (2007): Fundamentals of Computer Algorithms, 2/e, Universities Press.
2. T.H. Cormen, CE Leiserson, R.L Rivert, C Stein, (2010): Introduction to Algorithms, 3/e, PHI.
3. R. Pannerselvam (2007): Design and Analysis of Algorithms, PHI.
4. Hari Mohan Pandey, (2009): Design, Analysis and Algorithm, University Science Press.

M.SC. (APPLIED STATISTICS) IV-SEMESTER

STAS-405: PAPER-V : STATISTICAL PROCESS AND QUALITY CONTROL, APPLIED STOCHASTIC PROCESSES

PRACTICAL-I (CONVENTIONAL)

SECTION-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

SECTION-B: 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.

M.SC. (APPLIED STATISTICS) IV-SEMESTER

STAS-406: PAPER-VI: SECTION-A (ELECTIVE-I) & SECTION-B (ELECTIVE-II)

PRACTICAL-II (CONVENTIONAL)

ELECTIVE-I (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. Game Theory
6. Simulation

ELECTIVE – I (B): TEXT ANALYTICS

1. Perform data collection by web scrapping with python and Perform following tasks (i) Find the URL that you want to scrape (ii) Inspecting the Page (iii) Find the data you want to extract (iv) Write the code (v) Run the code and extract the data (vi) Store the data in the required format.
2. Perform following Data Pre-processing tasks in Python using Scikit-learn. standardization, normalization, encoding, discretization, imputation of missing values. Use your own dataset to perform all pre-processing tasks as suggested in given reference.
   (i) https://www.analyticsvidhya.com/blog/2016/07/practical-guide-datapreprocessing-python-scikit-learn/
   (ii) https://scikit-learn.org/stable/modules/preprocessing.html
3. Answer the following question in your blog (As per dataset taken by you): Dataset Description: Task to be performed: How to decide variance threshold in data reduction? Code Snapshot, Output Snapshot, Task-2, Code Snapshot, Output Snapshot.
   Perform following Data Pre-processing tasks using python
   Data reduction using variance threshold, univariate feature selection,
   recursive feature elimination, PCA, correlation
   Reference:
   1. https://medium.com/analytics-vidhya/feature-selection-using-scikit-learn5b4362e0c19b
   2. https://machinelearningmastery.com/rfe-feature-selection-in-python/
   3. https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60
   4. https://towardsdatascience.com/feature-selection-using-python-for-classificationproblem-b5f00a1c7028
   5. https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/
   Answer the following question in your blog (As per dataset taken by you):
   Dataset Description:
   Task to be performed:
   Why feature selection is important?? Its advantages/disadvantages.
   Code Snapshot
   Output Snapshot
   What is the impact on accuracy, with or without data reduction?
   Code Snapshot
   Output Snapshot
   Amongst all methods, which method avoids overfitting and improves model performance

ELECTIVE-I (C): DEMOGRAPHY

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

ELECTIVE – II (A) 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(B): DESIGN & ANALYSIS OF ALGORITHMS (USING PYTHON)

1. Write a program for sorting the given list using: Merge Sort, Quick Sort, Heap Sort.
2. Write a program to find the given number in a list using Binary Search.
3. Write a program to find the minimal spanning tree using Kruskal's and Prims Algorithms.
4. Write a program to find the shortest path using Dijkstra's Algorithm.
5. Write a program to solve using dynamic programming technique for Travelling sales man problem. Multistage Graph problem, Optimal Binary Search Trees.
6. Write a program to solve Knapsack problem using Back tracking

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

M.SC. (APPLIED STATISTICS) IV-SEMESTER

STAS-407: PAPER-VII: MAJOR PROJECT

Note: Follow the guidelines of the project specified in STS-308, Mini project.

****