M.Sc. STATISTICS COURSE DETAILS

Eligibility : B.A./B.Sc. with Mathematics & Statistics

Fee : Rs.8000/-

Duration : 4 – Semester / 2 years

CHECK LIST FOR ADMISSION :

1. Online Registration Form
2. Transfer Certificate (T.C. ORIGINAL)
3. Migration Certificate (Other than O.U.)
4. S.S.C. Memorandum of Marks
5. Intermediate Memorandum of Marks
6. Degree Memorandum of Marks, Consolidated Memo
7. Degree Provisional and Convocation Certificate
8. Caste Certificate for SC / ST / BC candidates
9. PHC, O.U. Employee certificate if applicable
10. Aadhaar card

All the certificate has to be produced in original for verification along with one Xerox set at the counter

Note : Candidates belonging to the other state Universities has to obtain the EQUIVALENCE CERTIFICATE from Deputy Registrar (Academic) Section, OU.

Website : oucde.net

IVRS Phone No.8886111690

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

M.Sc. (STATISTICS) III-SEMESTER

SCHEME OF INSTRUCTIONS AND EXAMINATION w.e.f. 2024-2025

Paper	Sub. Code	Paper Title	Credits	Semester end Exam duration	Max. Marks in the Internal Assessment and Attendance	Max. Marks in Semester end Exam
THEORY PAPERS						
I	STS-301	Operations Research-I (OR-I)	3	2	30	70
II	STS-302	Applied Regression Models (ARM)	3	2	30	70
III	STS-303	Time Series Analysis (TSA)	3	2	30	70
IV	STS-304	Statistical Quality Control (SQC)	3	2	30	70
PRACTICAL PAPERS						
V	STS-305	OR-I & ARM Lab	2	2	-	50
VI	STS-306	TSA & SQC Lab	2	2	-	50
VII	STS-307	Data Analysis Project -I	4	-	-	100
Semester Total			**20**	-	**-**	600

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

M.Sc. (STATISTICS) IV-SEMESTER

SCHEME OF INSTRUCTIONS AND EXAMINATION w.e.f. 2024-2025

Paper	Sub. Code	Paper Title	Credits	Semester end Exam duration	Max. Marks in Internal Assessment and attendance	Max. Marks in Semester end Exam
		THEORY PAPERS				
I	STS-401	Stochastic Process (SP)	3	2	30	70
II	STS-402	Reliability Theory (RT)	3	2	30	70
III	STS-403	Operations Research-II (OR-II)	3	2	30	70
IV	STS-404	Advanced Machine Learning Techniques (AMLT)	3	2	30	70
		PRACTICAL PAPERS				
V	STS-405	OR-II & RT Lab	2	2	--	50
VI	STS-406	OR-II & AMLT Lab	2	2	--	50
VII	STS-407	Data Analysis Project-II	4	-	--	100
Total			20	-	--	600

M.SC. (STATISTICS) III-SEMESTER

CDESTS-301: PAPER I: OPERATIONS RESEARCH - I

UNIT–I

Operations Research: Meaning and scope of OR. Convex sets and their properties. General linear Programming Problem (LPP). Formulation of LPP. Statements of Fundamental theorem of LPP and other related theorems. Optimal Solution of LPP by graphical, Simplex and Charner's & two-phase methods. Concept of degeneracy and resolving it Concept of duality of LPP. Dual Primal relationship, Complementary slackness theorems, Fundamental Theorem of Duality. Dual simplex Algorithm.

UNIT-II

Sensitivity Analysis: Discrete changes requirement and cost vectors; parametric programming: Parameterization of cost and requirement vectors Definition of transportation problem, TPP as a special case of LPP, Initial basic feasible solutions by North-West Corner Rule, Matrix minimum method and VAM. Optimal solution through MODI method for balanced and unbalanced transportation problem. Description of Assignment problem, Assignment problem as special case of TP and LPP. Unbalanced assignment problem, optimal solution using Hungarian method and traveling salesman problem and its solution. Problem of Sequencing. Optimal sequence of N jobs on two and three machines without passing.

UNIT-III

Integer Programming Problem: Gomory's cutting plane Algorithm for pure and mixed IPP Branch and bound Technique. Basic concepts of Networks constraints; Construction of Network and critical path; PERT and CPM; Network flow problems. Time Cost Analysis. Inventory Models: Introduction; ABC analysis and Deterministic Inventory models with and without shortages.

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.
4. Hillier F.S. and Leiberman, G.J.(1962) : Introduction to Operations Research; Holdon Day

M.SC. (STATISTICS) III-SEMESTER

CDESTS-302: PAPER-II: APPLIED REGRESSION MODELS (ARM)

UNIT–I

Selection of Best linear Regression: Simple & Multiple linear Regression Analysis, Estimation of parameters and Computation of R^2, Adjusted R^2, MSE, lack of fit of the model, testing significance of regression coefficient(s). Methods for selection of best linear regression: all possible regression, backward, forward, step-wise, stage-wise regressions. 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). Growth models: Logistic and Gompertz growth models.

Logistic regression model: Introduction to simple Logistic model, Fitting the model, testing for the significance of the coefficients, Logistic model for Dichotomous independent variable; Introduction to multiple Logistic regression, fitting the multiple logistic regression model, testing for the significance of the model. **Probit Analysis:** Introduction, Analysis of Biological data, sigmoid curve, fitting a Probit Regression line through least squares method.

Unit-III

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. **Generalized Linear Models:** 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. Applied Regression Analysis: Norman R. Draper and Harry Smith
2. Regression Analysis: Concepts and Applications, Franklin A. Graybill and Hariharan K. Iyer
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) III-SEMESTER
CDESTS-303: PAPER-III: TIME SERIES ANALYSIS (TSA)

UNIT–I

Stationery stochastic processes. The autocovariance and Auto correlation functions and their estimation. Standard errors of autocorrelation estimates. Bartlett's approximation (without proof). The periodogram, the power spectrum and spectral density functions. Link between the sample spectrum and autocorrelation function. 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-II

Auto Regressive (AR) and Moving Average (MA) processes: Evaluation of Auto-correlation function (ACF), Partial autocorrelation function (PACF). Spectrum for the process up to order 2, Stationarity and Invertibility conditions for AR(p) and MA(q), Duality between AR and MA processes. Mixed AR and MA (ARMA) process, Stationarity and invertibility properties, ACF and spectrum of mixed processes, The ARMA (p, q) process and its properties.

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.

UNIT–III

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. 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. 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.(STATISTICS) III-SEMESTER
CDESTS-304: PAPER-IV : STATISTICAL QUALITY CONTROL

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.

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–l and its properties modifications over CSP–l.

UNIT–III

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.(STATISTICS) III-SEMESTER
CDESTS-305: PAPER-V: PRACTICAL -I (OR-I & ARM)

SECTION-A: OPERATIONS RESEARCH-I (CONVENTIONAL & USING TORA/R)

1. Formulation of LPP
2. Optimal Solution to a L.P.P. by (i) Simplex method (ii) Charners Method (iii) Two-phase simplex method (iv) Dual Simplex Method (v) Duality
3. IBFS for a transportation problem by North-West corner rule, Matrix minimum method and Vogel's approximation method and also Optimum solution to balanced and unbalanced transportation problem by MODI method.
4. Optimum solution to balanced and unbalanced Assignment problem by Hungarian method and also Solution of traveling salesman problem.
5. Computation of Optimal Sequence and idle time for N jobs on 2 and 3 machines
6. Sensitivity Analysis
7. Parametric Programming Problem
8. Integer Programming Problem
9. Evaluation of project time through CPM and PERT
10. Evaluation of Time cost analysis through CPM and PERT

SECTION-B: APPLIED REGRESSION MODELS (USING R)

1. Analysis of Simple linear Regression.
2. Analysis of Multiple linear Regression (in conventional upto 3 variables, using software upto 10 variables)
3. Selection of Best linear Regression using All possible Regression, Forward, Backward, Step-wise and Stage wise Regression methods.
4. Simple Logistic regression for Dichotomous data, upto m, n = 4.
5. Analysis of Multiple Logistic regression (in conventional upto 2 variables).
6. Analysis of probit model.
7. Variance component analysis for one factor (balanced data) and two factors.
8. Computation of mean and variance for exponential family of distributions.

M.SC.(STATISTICS) III-SEMESTER

CDESTS-306: PAPER-VI: PRACTICAL -II (TSA & SQC)

SECTION-A: FORECASTING MODELS (USING R)

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.

SECTION-B: STATISTICAL QUALITY CONTROL (CONVENTIONAL & USING R)

1. Construction of charts and OC, ASN curves for Mean and Range charts
2. Construction of charts and OC, ASN curves for charts for attributes (p, d, c; for a fixed n)
3. Construction of simple and exponentially weighted moving average control chart and simple moving range control chart.
4. Construction of CUSUM charts V – Mark and Tabular methods
5. Construction of AOQ and AFI curves for CSP–I

M.SC.(STATISTICS) III- SEMESTER

CDESTS-307: PAPER VII: DATA ANALYSIS PROJECT-I

GUIDELINES

1. Each Student has to do one standalone Statistical data analysis project in III-Semester (Minor Project) to familiarize the practical usage of all statistical techniques covered in UG & PG level using Statistical software's Python and R, from any one industry/ institution under a recognized Supervisor (supervisor must be eligible for teaching at PG level Statistics in any University as per UGC norms).
2. Suggested to collect Live large scale data set (may be primary or secondary source data set with minimum sample size 2000 with minimum 10 data variables with different measurement of scales of variables)
3. Each student has to submit the project reports in two copies (minimum 100 pages) hard bound copy along with the assignments follow the Ph.D. thesis norms as per Osmania University duly signed by the Students on Declaration, Certificate from Industry /institution and certified by recognized Supervisor supported by Plagiarism report (not exceeding 10% of similarity index).
4. The Project Report should fulfill the norms of statistical data analysis report. (It should contains (a) Literature collected related to the study and its review (research articles minimum 10) & Data Domain Description (b) National & International significance on the problem (c) Detailed data variable description (d) Data set objectives/ objectives of the study (e) Formation of statistical hypothesis (f) Data visualization techniques applied on each/combination of variables (g) Descriptive statistics (h) Exploratory Data Analysis on each variable understudy, (i) Advanced statistical tools applied (minimum 5), (j) Results analysis and its interpretation. (k) sample data set (min 100), Python / R-programs written for implementation should be placed in appendix (l) Bibliography; Analysis report. Usage of Python programs should be presented in Appendix.
5. Projects will be evaluated by two subject experts by conducting Project viva-voce along with other practical's. May require presentation on the Project.
6. Project Marks will be awarded based on 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.

M.SC.(STATISTICS) IV-SEMESTER
STS-402: PAPER-I: STOCHASTIC PROCESSES (SP)

UNIT – I

Introduction to stochastic processes; classification of stochastic process according to state-space and time-domain. Finite and countable state Markov chains; time-homogeneity; Chapman-Kolmogorov equations; marginal distribution and finite – dimensional distribution. classification of states of a Markov chain – recurrent, positive recurrent, null - recurrent and transient states. Period of a state. Canonical form of transition probability matrix of a Markov chain. Fundamental matrix; probabilities of absorption from transient states into recurrent classes, in a finite Markov Chain; mean time for absorption. Ergodic state and ergodic chain.

UNIT-II

Stationary distribution of a Markov chain. Existence and evaluation of stationary distribution. Random walk and gambler's ruin problem. Weiner process as limit of random walk. First passage time of the process. Discrete state-space, continuous time Markov Processes – Kolmogorov difference - differential equations. Poisson process and its properties. Birth and Death Process, application in queuing. Pure Birth and pure Death processes.

UNIT – III

Renewal process, elementary renewal theorem and its applications. Statement and uses of Key – renewal theorem. Residual life time. Branching process – Galton-Watson branching process, mean and variance of size of n^{th} generation; probability of ultimate extinction of a branching process – fundamental theorem of Branching process – Examples.

REFERENCES

1. Medhi,J. (1982) : Stochastic Processes – Wiley Eastern
2. Karlin, S. and Taylor, H.M. (1975): A First Course in Stochastic Processes, Vol. I, Acad. Press.
3. Bhat, B.R. (2000): Stochastic Models: Analysis and applications-New Age International India.
4. Basu, A.K. (2003): Introduction to Stochastic Process, Narosa Publishing House.

M.SC. (STATISTICS) IV-SEMESTER

CDESTS-402: PAPER-II: 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. 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–II

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–III

Life Distribution: Survival function: Notion of ageing, IFR, DFR, DFRA, NBU and NBUE classes, Exponential distributions and its no-ageing property, ageing properties of other common life distribution, closures under formation of coherent structures, convolutions and mixtures of these cases. 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.
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.(STATISTICS) IV-SEMESTER

STS-403: PAPER-III: 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. Dynamic Programming, Principle of optimality, solution of LPP by Dynamic Programming technique, Knapsack problem by Dynamic Programming Technique.

UNIT–II

General goal Programming model and formulation of its objective function. Solutions to linear goal programming and linear integer goal programming. 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-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

S-S policy for inventory and its derivation in the case of exponential demand; Models with variable supply and models for perishable Items.

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.(STATISTICS) IV-SEMESTER

STS-404: PAPER-IV: ADVANCED MACHINE LEARNING TECHNIQUES (AMLT)

UNIT-I

Understanding of data and identification of appropriate statistical tools related to descriptive statistics, probability models, Parametric and Non-parametric tests, simple, multiple and logistic regression models, appropriate multivariate data analysis techniques,

Classification: Support Vector Machines and its construction, Naïve Bayes classifier, k-Nearest Neighbor. Decision Tree classification.

UNIT–II

Ensemble methods: Random Forest, Bagging, Boosting (Gradient, Ada-, XG), Market-Basket Analysis. Association Rules, Apriori and Frequent Pattern-Growth Algorithms

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

UNIT-III

Deep learning techniques: Perceptron Learning Algorithm, Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning, Back-Propagation Learning Algorithms and their real time applications in classification. Radial Basis Function Networks, Reinforcement learning, 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.

M.SC.(STATISTICS) IV-SEMESTER
STS-305: PAPER-V: PRACTICAL -I: (SP & RT)

SECTION-A: STOCHASTIC PROCESSES (CONVENTIONAL)

1. Formulation of problems as Markov chain models
2. Computation of finite dimensional and marginal distributions; higher dimensional transition probabilities.
3. Classification of states, identification of recurrent classes and reduction to canonical form of TPM.
4. Probabilities of absorption into recurrent classes (from transient states)
5. Computation of stationary distribution (unique case)
6. Computation of stationary distribution (non-unique case)
7. M|M|1 queue – operating characteristics
8. Mean and variance of n^{th} generation size and probability of extinction of Branching processes.

RELIBILITY THEORY (CONVENTIONAL)

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.

M.SC. (STATISTICS) IV-SEMESTER

STS-306: PAPER-VI: PRACTICAL-II (OR-II & AMLT)

SECTION-A: OPERATIONS RESEARCH –II (CONVENTIONAL & USING R)

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

SECTION-B: ADVANCED MACHINE LEARNING TECHNIQUES USING PYTHON

1. Analysis of Simple and Multiple linear regression Models
2. Analysis of Simple and Multiple Logistic regression Models
3. Construction of Support Vectors and classification using SVM
4. Classification using Naïve Bayes classier
5. Classification using Ensemble methods: Random Forest, Bagging, Boosting, Ada & XG
6. K-Nearest Neighbor, K-means, C 4.5, CART, DBSCAN clustering
7. Association Rule algorithms: Apriori and FP-tree growth.
8. Implementation of Perceptron Learning
9. Implementation of Multi-layer Perceptron Learning
10. Implementation of Back-Propagation Algorithms.
11. Implementation of Markov Decision Process,
12. Implementation of Hidden Markov Model,
13. Implementation of Convolutional Neural Networks,
14. Implementation of Recurrent Neural Networks,
15. Implementation of Long-Short Term Memory.

M.SC.(STATISTICS) IV- SEMESTER

CDESTS- 407: PAPER VII: DATA ANALYSIS PROJECT-II

PROJECT GUIDELINES

1. Each Student has to do one standalone Statistical data analysis project in IV semester (Major project) (it should be different with III sem project) to familiarize the practical usage of all statistical techniques covered in UG & PG level using Statistical software's Python and R, from any one industry/ institution under a recognized Supervisor (supervisor must be eligible for teaching at PG level Statistics in any University as per UGC norms).
2. Suggested to collect Live large scale data set (may be primary or secondary source data set with minimum sample size 2000 with minimum 10 data variables with different measurement of scales of variables)
3. Each student has to submit the project reports in two copies (minimum 100 pages) hard bound copy along with the assignments follow the Ph.D. thesis norms as per Osmania University duly signed by the Students on Declaration, Certificate from Industry /institution and certified by recognized Supervisor supported by Plagiarism report (not exceeding 10% of similarity index).
4. The Project Report should fulfill the norms of statistical data analysis report. (It should contains (a) Literature collected related to the study and its review (research articles minimum 10) & Data Domain Description (b) National & International significance on the problem (c) Detailed data variable description (d) Data set objectives/ objectives of the study (e) Formation of statistical hypothesis (f) Data visualization techniques applied on each/combination of variables (g) Descriptive statistics (h) Exploratory Data Analysis on each variable understudy, (i) Advanced statistical tools applied (minimum 5), (j) Results analysis and its interpretation. (k) sample data set (min 100), Python / R-programs written for implementation should be placed in appendix (l) Bibliography; Analysis report. Usage of Python programs should be presented in Appendix.
5. Projects will be evaluated by two subject experts by conducting Project viva-voce along with other practical's. May require presentation on the Project.
6. Project Marks will be awarded based on 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.