Hindi Vidya Prachar Samiti's

# RAMNIRANJAN JHUNJHUNWALA COLLEGE OF ARTS, SCIENCE & COMMERCE

## (Autonomous)

Affiliated to

## University of Mumbai

## Syllabus for the M. Sc.

## Program: M. Sc. In Statistics

## Program Code: RJSPGSTA

**(Choice Based Credit System with effect from the academic year 2021 – 2022 for Semester I and II (Part I )and 2022 – 2023 for Semester III and IV (Part II)**

# Syllabus

| Sr. No. | Heading | Particulars |
|---|---|---|
| 1 | Title of course | M.Sc. in Statistics |
| 2 | Eligibility for admission | A candidate for being eligible for admission to the M.Sc. degree course in Statistics must have passed<br>The B.Sc. (Three Year Integrated course) degree examination of this University (or any other University recognized as equivalent there to) with at least Seven Units in Statistics (i.e. the minimum required for majoring in the subject).<br>**OR**<br>(a) With Three Units in Statistics at T.Y.B.Sc. in combination with Three Units of Mathematics at the T.Y.B.Sc.<br>The students will be admitted to the M.Sc. degree course in Statistics on the basis of marks obtained at the T.Y.B.Sc. examination. |
| 3 | Passing Marks | 40% |
| 4 | No. of Years, Semesters | 2 Years, 4 Semesters |
| 5 | Level | Post Graduate |
| 6 | Pattern | Semester |
| 7 | Status | Approved |
| 8 | To be implemented from Academic Year | Part I: 2021 – 2022,<br>Part II: 2022 – 2023 |

# Members of Board of Studies

| Sr. No. | Name of BOS Member | Designation | Signature |
|---|---|---|---|
| 1 | Mrs. Jayshree Vaze Chairperson | Associate Professor | |
| 2 | Mr. Ajit Limaye | Associate Professor | |
| 3 | Mrs. Chhaya Pinge | Associate Professor | |
| 4 | Mr. Rahul Tiwari | Assistant Professor | |
| 5 | Mr. Jaishankar Singh | Lecturer | |
| 6 | Dr. Ganesh Bala | Professor and Head, Department of Biostatistics and Epidemiology Subject Expert Other University | |
| 7 | Mr. Prasad Patki | Vice Principal, Head and Associate Professor in Statistics, Bhavan's College VC Nominee | |
| 8 | Dr. Leena Kulkarni | Assistant Professor NMIMS's Sunandana Divita School of Science | |
| 9 | Mrs. Pratima Parab | Vice President, Risk Analytics Unit HDFC Bank Industry expert | |
| 10 | Mr. Vikas Dubey | Senior Research Analyst, Nielson Meritorious alumnus | |

**Meeting of Board Studies: January 29, 2021**

**Academic Council: March 23, 2021**

Governing Body : 6th April, 2021

# Program Educational Objectives

1. To enable graduates to excel professionally by adapting to the dynamic needs of the academia, industry and research in the field of **Statistics** and Data Science.
2. To enable student to excel in the field of Data Analytics, Data Mining, Machine Learning, Visualization Techniques, Predictive Analysis and Statistical modeling.
3. To practice the problems of analysis and decision making using big data.
4. To gain practical, hands-on experience with programming languages, data analysis tools and frameworks through coursework.

# Program Outcomes

Students who have completed the M.Sc. in Statistics will be able to:

1. To apply statistical modeling and data analysis techniques to the solution of real-world business problems, effectively present results using data visualization techniques.
2. To test and train various machine learning algorithms for real world data and applications.
3. To create data warehouse and mine the data for analysis.
4. To analyse big data using various languages and tools.
5. To apply machine learning and deep learning algorithms to real-world problems.

# Course Structure

## Semester I

| Course Code | Course Name | Group | Teaching Scheme (Hrs/Week) | | Credits |
|---|---|---|---|---|---|
| | | | Lectures | Practical | |
| RJSPGSTA101 | Distribution Theory and Its Application | | 4 | - | 4 |
| RJSPGSTA102 | Applied Statistical Inference | | 4 | - | 4 |
| RJSPGSTA103 | Sampling Theory | | 4 | - | 4 |
| RJSPGSTA104 | Linear models and Regression Analysis | | 4 | | 4 |
| RJSPGSTAPA101 | Statistical Computing I (Practical based on Paper I and II in relevant software) | | - | 6 | 3 |
| RJSPGSTAPA102 | Statistical Computing II (Practical based on Paper III and IV in relevant software) | | - | 6 | 3 |
| RJSPGSTA1I1 | Career Advancement Course Organizational Behavior | | 2 | - | 2 |
| | **Total** | | **18** | **12** | **24** |

## Semester II

| Course Code | Course Name | Group | Teaching Scheme (Hrs/Week) | | Credits |
|---|---|---|---|---|---|
| | | | Lectures | Practical | |
| RJSPGSTA201 | Multivariate Analysis and Its Application | | 4 | - | 4 |
| RJSPGSTA202 | Design of Experiments | | 4 | - | 4 |
| RJSPGSTA203 | Stochastic Processes | | 4 | - | 4 |
| RJSPGSTA204 | Time Series Analysis | | 4 | | 4 |
| RJSPGSTAPA201 | Statistical Computing I (Practical based on Paper I and II in relevant software) | | - | 6 | 3 |
| RJSPGSTAPA202 | Statistical Computing II (Practical based on Paper III and IV in relevant software) | | - | 6 | 3 |
| RJSPGSTA2I1 | Career Advancement Course Leadership Development | | 2 | - | 2 |
| | **Total** | | **18** | **12** | **24** |

# Semester III

| Course Code | Course Name | Group | Teaching Scheme (Hrs/Week) | | Credits |
|---|---|---|---|---|---|
| | | | Lectures | Practical | |
| RJSPGSTA301 | Machine Learning | | 4 | - | 4 |
| RJSPGSTA302 | Big Data Technology | | 4 | - | 4 |
| RJSPGSTA303 | Data Mining | | 4 | - | 4 |
| RJSPGSTA304 | Data Warehousing | | 4 | - | 4 |
| RJSPGSTAPA301 | Statistical Computing I (Practical based on Paper I and II in relevant software) | | - | 6 | 3 |
| RJSPGSTAPA302 | Statistical Computing II (Practical based on Paper III and IV in relevant software) | | - | 6 | 3 |
| RJSPGSTA3I1 | Career Advancement Course Business Ethics and CSR | | 2 | - | 2 |
| | **Total** | | **18** | **12** | **24** |

# Semester IV

| Course Code | Course Name | Group | Teaching Scheme (Hrs/Week) | | Credits |
|---|---|---|---|---|---|
| | | | | | |
| RJSPGSTA4INT | Major Project Dissertation | | - | | 12 |
| | Industrial Internship | | | | 12 |
| | **Total** | | **-** | | **24** |

# Scheme of Examination

**Continuous Evaluation with 40 marks Internal and 60 marks external**

Passing Standards 40% separately for Internal and External

Modes of Internal Assessment: Seminar, problem based, Assignments, mini projects, case studies 40 marks

External Assessment: Question Paper will cover the entire syllabus to assess various levels of learning

Duration of two and half hour 60 marks each paper

Practical assessment to assess the application skills of student two practicals of 100 marks each with continuous evaluation internal assessment 50 marks and term end assessment of 50 marks of two hours.

Career advancement course : 50 marks continuous evaluation

## Semester I S I / P I

| Course code | Course name | Group | Teaching Scheme (Hrs/Week) | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA101** | **Distribution Theory and Its Application** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for Statistical Distributions.
2. To expose students to Statistical Distributions, its theoretical aspects and applications.

**Learning Outcomes**

Upon completion of this course, the student should be able to understand probability **distributions** and its application.

| Unit | Topic | Lectures |
|---|---|---|
| I | Concept of random variable, Expectation, moments, moment generating function, Distribution function, decomposition of distribution function (Jordan's decomposition), transformation, Mixture probability models, probability models of truncated random variable, Leibnitz's rule , Standard discrete distributions, Standard continuous distributions. | 15 |
| II | Bivariate random variable, joint and marginal probability distributions, joint distribution function, conditional distribution and independence, Bivariate transformation, variance and covariance matrix, conditional expectation and variance, Bivariate normal distribution. Multiple and partial correlation coefficient. | 15 |
| III | Exact sampling distribution – Chi square, t , F and z probability models Convergence in distribution, convergence in probability, almost sure convergence . Chebycheve's inequality. Probability models of quadratic form. | 15 |
| IV | Order statistics - Joint distribution of order statistics, Distribution of $r^{th}$ order statistics, joint distribution of rth and sth order statistics(r<s). Distribution of sample median and sample range. Non parametric test- Krushkal Wallis tests, Chi square goodness of fit test, Kolmogorov- Smirnov one sample and two sample test. | 15 |

References:

1. Cassela G. and Berger R . Statistical Inference ($2^{nd}$ edition), Duxbury Resource Centre.
2. Bhat B.R. (1999): Modern Probability Theory: An introductory test book $3^{rd}$ edition. New Age International
3. Robert V. Hogg, Allen Craig and Joseph McKean : Introduction to Mathematical statistics($6^{th}$ edition), Pearson Prentice Hall.
4. Rohatgi V.K. : Introduction to Probability and Statistics, paperback, Wiley, New York.
5. Wayne W. Daniel : Applied Nonparametric statistics ($2^{nd}$ edition) Duxbury Thomas Learning.

**S I / PII**

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA102** | **Applied Statistical Inference** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for estimation theory.
2. To expose students to Bayesian inference and its application.

**Learning Outcomes**

Upon completion of this course, the student should be able to understand applicability of estimation theory in real life.

| Unit | Topic | Lectures |
|---|---|---|
| I | Estimation - Sufficiency, Neyman factorization theorem, Unbiasedness , Bounds on variance(Cramer-Rao Lower Bound, Bhattacharya Bound), Maximum likelihood estimation, Asymptotic properties of M.L.E, Consistency, Ancillary and complete statistics, P-value.<br>Testing of Hypothesis - N-P lemma, Most Powerful test, Uniformly Most Powerful test, Likelihood Ratio Test. | 15 |
| II | Bootstrap methods –parametric simulation, Non-parametric simulation, Simple confidence interval 's, Reducing Error, Statistical Issues.<br>EM algorithm. Jackknife estimator. | 15 |
| III | Bayesian Inference- Bayesian inference from the numerical posterior, Bayesian Inference from Posterior Random Sample.<br>Bayesian Statistics Using Conjugate Priors - One Dimensional Exponential Family of Densities, Distribution for Count data, Distributions for Waiting Times, Normally distributed Observations with known Variance, Normally distributed Observations with known Mean. | 15 |
| IV | Markov Chain Monte Carlo Sampling from Posterior<br>Metropolis- Hastings algorithm for a single parameter<br>Metropolis- Hastings algorithm for a Multiple Parameters<br>Block wise Metropolis –Hasting Algorithm<br>Gibbs Sampling. | 15 |

References:

1. Cassela G. and Berger R . Statistical Inference (2nd edition), Duxbury Resource Centre.
2. Ulhas Jayram Dixit: Examples in parametric inference with R, Springer.
3. A.C. Davison and D.V. Hinkley: Bootstrap Methods and their application, Cambridge University Press.
4. Geoffrey J. McLachlan and Thriyambakam Krishnan: The EM Algorithm and Extension, Wiley.
5. William M. Bolstad: Understanding Computational Bayesian Statistics, Wiley.
6. Robert V. Hogg, Allen Craig and Joseph McKean : Introduction to Mathematical statistics(6th edition), Pearson Prentice Hall.
7. E.L. Lehman and George Cassella: Theory of Point Estimation (2nd Edition), Springer, New York.
8. Rohatgi V.K. : Introduction to Probability and Statistics, paperback, Wiley, New York.

**S I / P III**

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA103** | **Sampling Theory** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for sampling theory.
2. To expose students to sampling and its applications.

**Learning Outcomes**

Upon completion of this course, the student should be able to understand how sampling theory is applicable in different organizations.

| Unit | Topic | Lectures |
|---|---|---|
| I | Systematic Sampling: Linear Systematic Sampling and Circular Systematic Sampling. Estimation of Population Mean, its Variance, Variance in terms of intra-sample correlation coefficient. Comparison with SRSWOR. Problem in Estimation of Variance using one systematic sample. Use of interpenetrating sub samples in estimation of variance. Other methods of estimation of variance. | 15 |
| II | Ratio and Regression Methods of Estimation and Two-phase Sampling Ratio and Regression Estimation of Population mean/total using SRSWOR. Comparison with Mean per Unit Estimator. Separate and combined ratio and regression estimators in stratified sampling. Mean Square error of Estimators. Unbiased type ratio estimator. Hartley-Ross Estimator.<br>Two phase sampling in stratification. | 15 |
| III | Cluster Sampling, Two Stage Sampling and Adaptive Sampling Cluster sampling: For equal and unequal cluster sizes. Estimation of population mean/total, its variance and estimation of variance. Ratio to size estimator, Mean of Unit Means Estimator. Comparison with SRSWOR. Cluster Sampling for Proportions.<br>Two Stage Sampling: With and Without Replacement at both the stages. Estimation of Population mean per second stage unit, its variance, estimation of variance. Optimum sampling.<br>Adaptive Sampling: Adaptive Cluster Sampling, Systematic and strip adaptive cluster sampling. Stratified Adaptive Cluster Sampling. | 15 |
| IV | Probability Proportional to size sampling and Network Sampling: Probability proportional to size sampling With Replacement (PPSWR): Hansen-Hurwitz Estimator of population total, its variance and estimator of the variance. Comparison with SRSWR. Cumulative Total Method and Lahiri's method of drawing PPSWR. PPSWOR: Horvitz-Thompson Estimator of Population Total, its variance and estimator of variance. Desraj ordered estimator, its expectation and variance, estimation of variance.<br>Network Sampling: Multiplicity Estimators. Horvitz-Thompson Estimator. Stratification in Network Sampling. | 15 |

| | | |
|---|---|---|
| | Non-sampling Errors: Response and Non-response Errors, Effect of Non-Response in simple random sampling. | |

References
1. Cochran W.G. (2007) sampling Techniques, 3rd Ed., Wiley.
2. P. Mukhopadhyay (2008) Theory and Methods of Survey Sampling, 2nd Ed.
3. Des Raj and Chandok P. (1998) Sampling Theory, Narosa Publication.
4. Singh D. and Chaudhary F.S. (1986) Theory and Analysis of Sample Survey Designs, New Age International Publishers.
5. Sukhatme P.V., Sukhatme B.V., Sukhatme S. and Ashok (1984) Sampling theory of Surveys with Applications, ICAR publication.
6. Bansal A, (2017): survey Sampling, Narosa.

**S I / P IV**

| Course code | Course name | Group | Teaching scheme | | Credits |
|---|---|---|---|---|---|
| **RJSPGSTA104** | **Linear models and Regression analysis** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for linear models.
2. To expose students to matrix theory and linear model.

**Learning Outcomes**

Upon completion of this course, the student should be able to understand how regression is used in different corporate sectors.

| Unit | Topic | Lectures |
|---|---|---|
| I | Linear Algebra - Matrix inverses and determinants, Solving systems of equations with matrices, Eigenvalues and eigenvectors, Orthogonal matrices, Positive definite matrices, Linear transformations, Linear dependence and independence.<br>Linear parametric function and its estimability, Solving linear equations, generalized inverse. Gauss markoff theorem, Interval estimates and test of hypothesis, fundamental theorems on conditional error ss, Test of $\Lambda\beta = d$, generalized least squares. | 15 |
| II | Analysis of variance, fixed effect models: i. One-way classification ii. Two-way classification model with and without interaction effect, one observation per cell. Tukey's test for non additivity. Two-way classification model with and without interaction effect with unequal number of observations per cell. | 15 |
| III | Linear regression models, subset selection, Stepwise regression: Forward selection, backward elimination and stepwise. Orthogonal polynomials. Assumptions and box-cox transformations in the Analysis of Variance: q-q plot, use of skewness and kurtosis, Bartlett's test for equality of variances, Levene's test. | 15 |
| IV | Ridge regression: Eigen values and Eigen vectors of a matrix. Conditioned matrix, need of ridge regression, biased estimator and Mean square error. Bias and MSE of ridge estimator, ridge trace method. Logistic regression: Example, model, MLE of parameters, Iterative procedure to solve likelihood equations, multiple regressors. Multinomial, ordinal, Poisson Analysis of Categorical data: Log linear models, contingency tables. | 15 |

References

1. Kshirsagar A.M. : A course in Linear Models
2. Draper N.R & Smith H : Applied Regression Analysis.
3. Song GUI Wang and S.C Chow: Advanced Linear Models.
4. Agresthi: Categorical data analysis.
5. Chattterjee and Haddi: Sensitivity Analysis

6. David W Hosmer and Stanley Lemeshow: Applied Logistic regression.
7. Healy M. J. R. : Matrices for Statistics
8. Shantinarayan : Textbook of Matrices
9. Bishop: discrete data analysis.
10. Cox, D. R. : Analysis of binary data.
11. Chaterjee and Price: Regression Analysis with examples

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTAPA101** | **Statistical Computing-I** | **CC** | **Lectures** | **Practical** | |
| | | | **-** | **6** | **3** |

**List of Practical's:**

**Practical 1:** Exact sampling distribution- chi-square, t, z
**Practical 2:** Non- parametric test
**Practical 3:** Estimation
**Practical 4:** Testing of hypothesis
**Practical 5:** Jackknife estimator
**Practical 6:** Bayesian estimation
**Practical 7:** MCMC

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTAPA102** | **Statistical computing-II** | **CC** | **Lecture** | **Practical** | |
| | | | **-** | **6** | **3** |

**List of Practical's:**

**Practical 1:** Systematic Sampling
**Practical 2:** Two Phase Sampling
**Practical 3:** Cluster and Two Stage Sampling
**Practical 4:** Varying Probability Sampling
**Practical 5:** Matrix Theory-I( Determinant, Rank of Matrix , Inverse of matrix)
**Practical 6:** Linear Model
**Practical 7:** ANOVA
**Practical 8:** Ridge and Logistic Regression

## Semester-II S II / P I

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA201** | **Multivariate Analysis and Its Application** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for multivariate analysis.

**Learning Outcomes**

Upon completion of this course, the student should be able to understand application of Principal Component Analysis, Factor Analysis and Cluster Analysis.

| Unit | Topic | Lectures |
|---|---|---|
| I | Multivariate Distribution<br>Random vector and its properties , Multivariate Normal Distribution and its properties , Distribution of Quadratic Form, Wishart Distribution and its properties, Hotelling's $T^2$- Distribution, Wilks $\Lambda$- Distribution, Test For Mean Vector. | 15 |
| II | Principal Component Analysis_ Introduction, Method of Extraction of Principal Component, Graphical Representation of Principal Component , Properties of Principal Components, Decision Regarding Number of Principal Components, The Effect of Ignoring Some Components. | 15 |
| III | Factor Analysis- Introduction, The model for factor analysis, Estimation of Factor Loading, Estimation of Factor Loadings from correlation Matrix, Factor extraction, Interpretation of factors, Factor Score, Factor rotation. Canonical Correlation Analysis- Introduction, Population correlation analysis, Sample Canonical correlation Analysis, Interpretation from Canonical Correlation Analysis, Score and Prediction, Method of Test. | 15 |
| IV | Cluster Analysis- Introduction, Basic Steps of Cluster Analysis, Forming Clusters, and Test regarding Clustering.<br>Discriminant Analysis- Scope of Discriminant analysis, Method of Discrimination, Probability of Misclassification, Test of Discriminant Function. | 15 |

References

1. K.C. Bhuyan: Multivariate Analysis and Its Application.
2. Johnson Richard A and Wicheren D.W. (1998): Applied Multivariate Statistical Analysis (4th Edition).
3. Giri Narayan C. (1995): Multivariate Statistical Analysis
4. Parimal Mukhopadhyay: Multivariate Statistical Analysis
5. Dillon William R & Goldstein Mathew (1984): Multivariate Analysis : Methods and Applications.

**S II / P II**

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA202** | **Design of Experiments** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for design of experiments.
2. To expose students to understand application of different designs.

**Learning Outcomes**

Upon completion of this course, the student should be able to understand how different designs are used in agricultural and government organizations.

| Unit | Topic | Lectures |
|---|---|---|
| I | Randomized Block design, Latin Square Design , Graeco-Latin Square Design, Balance Incomplete Block Design, C matrix , Statistical analysis of BIBD, Estimation of Parameters Model adequacy checking. | 15 |
| II | Factorial design – An example. The advantage of factorial designs. $2^2$ factorial designs.<br>General $2^k$ factorial experiment. Blocking, Confounding and partial confounding. | 15 |
| III | Experiments with Random Factors<br>Random Effects Models, The Two-Factor Factorial with Random Factors, The Two-Factor Mixed Model.<br>The Two-Stage Nested Design, The Split-Plot Design. | 15 |
| IV | Response Surface Methods: Introduction to Response Surface Methodology, The Method of Steepest Ascent, Analysis of a Second-Order Response Surface, Experimental Designs for Fitting Response Surfaces. | 15 |

References:

1. Montgomery D. C. (2017). Design and analysis of experiments Wiley.
2. Das, M.N. and Giri N. C. (1986): Design and analysis of experiments, New Age International.
3. M.C. Chakrabarti : Mathematica of Design and Analysis of Experiments, Asia Publishing House.
4. Cochran W. G. and Cox G.M. (1959): Experimental Design, Asia publishing House.
5. Fisher R. A. (1935): The Design of Experiments, Olive and Boyd.

**S II / P III**

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA203** | **Stochastic Processes** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for stochastic Process.
2. To expose students to understand Markov chain.

**Learning Outcomes**

Upon completion of this course, the student should be able to understand how stochastic models are used in different model building.

| Unit | Topic | Lectures |
|---|---|---|
| I | Markov Chains- Chapman – Kolmogorov equations, Classification of states , Mean time spent in Transient States, Limiting Probabilities, Gambler's ruin problem, random walk, Branching Process. | 15 |
| II | Poisson process, Generalization of Poisson process, Random intensity functions and Hawkes Processes.<br>Renewal theory and its applications, Distribution of N(t), Limit theorems and their application, Renewal reward processes, Regenerative processes, Computing the renewal function, Applications to patterns and insurance. | 15 |
| III | Continuous-time Markov chains, Birth and Death Processes, The transition probability function Pij(t), Limiting probabilities, Time reversibility, The reversed chain, Computing the transition probabilities. Queuing theory, Queuing models and Network of queues, M/G/1 and its variations, G/M/1, Multi server queues. | 15 |
| IV | Brownian motion and stationary processes, Hitting times, Maximum variable, Variations on Brownian motion, Geometric Brownian motion, Application to pricing of stock options, The maximum of Brownian motion with drift, White noise, Gaussian processes. | 15 |

Reference

1. Ross, S. M. (2014). Introduction to Probability Models, 11th Ed, Academic Press, New York.
2. Medhi, J. (2017). Stochastic Processes, Paperback, 4th Ed, New Age International.
3. Ross S. M. (2011). An elementary Introduction to Mathematical Finance, 3rd Ed, Cambridge University Press, London.
4. Bhat, B. R. (2000). Stochastic Models: Analysis and Applications, New Age International.
5. Cinlar, E. (2013). Introduction to Stochastic Processes, Paperback, Dover Publications Inc.
6. Hoel, P. G., Port, S. C. and Stone, C. J. (1986). Introduction to Stochastic Processes, Waveland Pr Inc.
7. Pinsky, M. A. and Karlin, S. (2010). An Introduction to Stochastic Modeling, 4th Ed, Academic Press.

**S II / P IV**

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA204** | **Time Series Analysis** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for Time series
2. To expose students to understand different processes in Time Series.

**Learning Outcomes**

Upon completion of this course, the student should be able to understand how time series models are used for forecasting.

| | Topic | Lectures |
|---|---|---|
| I | Time-series as discrete parameter stochastic process. Exploratory Time Series Analysis: Tests for trend and seasonality, Exponential and Moving average smoothing. Hot Winters smoothing. Forecasting based on smoothing, adaptive smoothing. | 15 |
| II | Auto covariance and auto correlation functions and their properties. Auto covariance and auto correlation functions and their properties, invertibility. Stationary processes: a) moving average) ( MA), b) Auto Regressive (AR), c) ARMA and (d) AR integrated MA (ARIMA) models, Box-Jenkins models, Discussion of estimation of mean, auto covariance and auto correlation functions under large sample theory ( without proof). | 15 |
| III | Choice of AR and MA periods, Estimation of ARIMA models parameters. Forecasting, Residual analysis and diagnostic checking. | 15 |
| IV | Spectral analysis of weakly stationary process, Periodogram and Correlogram analysis. Computations based on Fourier transform. Spectral Decomposition of weakly AR process and representations as a one-sided MA process- necessary and sufficient conditions | 15 |

References

1. Anderson, T. W (1971): The Statistical Analysis of Time Series, Wiley, N.Y.
2. Brockwell, P.J. and Davis, R. A. Time-Series: Theory and Methods (Second Edition), Springer-Verlag. 19
3. Box, G.E.P. and Jenkins, G.M. (1976): Time Series Analysis-Forecasting and control Hodlen-day, San Franciscor.
4. Kendall, Sir Maurice and Ord. J. K. (1990): Time Series (Third Edition) Edward Arnold.
5. Montgomery, D. C. and Johnson, L. A. (1977): Forecasting and Time Series Analysis, McGraw Hill.

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTAPA201** | **Statistical Computing-I** | **CC** | **Lectures** | **Practical** | |
| | | | **-** | **6** | **3** |

**List of Practical's:**

**Practical 1:** Multivariate Distribution
**Practical 2:** Principal Component Analysis
**Practical 3:** Factor Analysis
**Practical 4:** Cluster Analysis
**Practical 5:** Design of Experiment -I
**Practical 6:** Design of Experiment - II
**Practical 7:** Response Surface Methodology

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTAPA202** | **Statistical computing-II** | **CC** | **Lecture** | **Practical** | |
| | | | **-** | **6** | **3** |

**List of Practical's:**

**Practical 1:** Markov Chain -I
**Practical 2:** Markov Chain -II
**Practical 3:** Queuing Theory
**Practical 4:** Time Series -I
**Practical 5:** Time Series -II
**Practical 6:** Time Series -III

## Semester -III S III / P I

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA301** | **Machine Learning** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for various statistical and machine learning concepts and methods.
2. To expose students to understand machine learning solutions to regression, classification and clustering problems.

**Learning Outcomes**

Upon completion of this course, the student should be able to Perform end-to-end process of investigating data through a machine learning lens.

| | Topic | Lectures |
|---|---|---|
| I | Data pre-processing vectors, matrices and arrays, loading data, Data handling Handling numerical data and categorical data, Handling text, dates and time, Handling images. Statistical Learning What is statistical learning, assessing model accuracy. | 15 |
| II | Linear Regression, Multiple Linear Regression, Other Considerations in Regression Model, The Marketing Plan, Comparison of Linear Regression with K-Nearest. | 15 |
| III | Classification An overview of classification, why not linear regression, logistic regression, linear discriminant analysis, a comparison of classification methods. | 15 |
| IV | Unsupervised Learning The challenge of unsupervised learning, principal components analysis, clustering methods (density-based methods, hierarchical-based methods, portioning-based methods, grid-based methods), clustering algorithms (k-means, k-nearest neighbours). | 15 |

References

1. "An Introduction to Statistical Learning With Application in R", By Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer Texts in Statistics.
2. "Machine Learning", Mitchell Tom, McGraw Hill,1997.
3. "Pattern classification", 2nd edition, Richard O. Duda, Peter E. Hart, David G. Stork. Wiley, New York, 2001.
4. "Machine Learning: A Probabilistic Perspective", Kevin P. Murphy, MIT Press, 2012
5. "Practical Data Science", Andreas Francois Vermeulen, APress, 2018
6. "Principles of Data Science", Sinan Ozdemir, Packt, 2016.

**S III / P II**

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA302** | **Big Data Technology** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |

**Course Objectives**

1. To provide theoretical foundations for basic and advanced methods of big data technology and tools.
2. To expose students to understand the knowledge of MapReduce, Hadoop and its ecosystem.

**Learning Outcomes**

Upon completion of this course, the student should be able to understand build and maintain reliable, scalable and distributed systems with Apache Hadoop.

| | Topic | Lectures |
|---|---|---|
| I | INTRODUCTION TO BIG DATA<br>Introduction: Distributed file system, Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. INTRODUCTION HADOOP Big Data: Apache Hadoop & Hadoop Ecosystem, Moving Data in and out of Hadoop, Understanding inputs and outputs of MapReduce, Data Serialization. | 15 |
| II | HADOOP ARCHITECTURE<br>Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., Name Node, Secondary Name Node, and Data Node, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers, Cluster Setup, SSH & Hadoop Configuration, HDFS Administering, Monitoring & Maintenance. | 15 |
| III | HADOOP ECOSYSTEM AND YARN<br>Hadoop ecosystem components: Schedulers, Fair and Capacity, Hadoop 2.0 New Features Name Node High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN. | 15 |
| IV | Hive and HiveQL, HBase Hive Architecture and Installation, Comparison with Traditional Database. HiveQL Querying Data, Sorting and Aggregating, Map Reduce Scripts, Joins & Subqueries. HBase concepts Advanced Usage, Schema Design, Advance Indexing, PIG, Zookeeper, how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper. | 15 |

References

1. "Professional Hadoop Solutions", Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Wiley, ISBN: 9788126551071, 2015.
2. "Understanding Big data", Chris Eaton, Dirk deroos et al, McGraw Hill, 2012

3. "HADOOP: The definitive Guide", Tom White, O Reilly 2012. 6 IT2015 SRM(E&T).
4. "Big Data Analytics with R and Hadoop", Vignesh Prajapati, Packet Publishing 2013.
5. "Oracle Big Data Handbook", Tom Plunkett, Brian Macdonald et al, Oracle Press, 2014.
6. "Big Data and Business analytics", Jy Liebowitz, CRC press, 2013.
7. http://www.bigdatauniversity.com/

**S III / P III**

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| RJSPGSTA303 | Data Mining | CC | Lectures | Practical | |
| | | | 4 | - | 4 |

**Course Objectives**

1. To provide theoretical foundations for data mining.
2. To expose students to learn the computational approaches to Modelling and Feature Extraction.

**Learning Outcomes**

Upon completion of this course, the student should be able to design algorithms by employing Map Reduce technique for solving Big Data problems.

| | Topic | Lectures |
|---|---|---|
| I | Data Mining—On What Kind of Data?<br>Relational Databases, Data Warehouses, Transactional Databases, Advanced Data and Information Systems and Advanced Applications.<br>Data Mining Functionalities—What Kinds of Patterns Can Be Mined?<br>Concept/Class Description: Characterization and Discrimination, Mining Frequent Patterns, Associations, and Correlations, Classification and Prediction, Cluster Analysis, Outlier Analysis, Evolution Analysis.<br>Classification of Data Mining Systems. | 15 |
| II | Data Pre-processing : Descriptive Data Summarization , Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization and Concept Hierarchy Generation . | 15 |
| III | Data Warehouse and OLAP Technology: What is a Data Warehouse?, A Multidimensional Data Model, Data Warehouse Architecture, Data Warehouse Implementation, From Data Warehousing to Data Mining. | 15 |
| IV | Classification and Prediction: What Is Classification? What Is Prediction?, Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Rule-Based Classification, Support Vector Machines, Lazy Learners (or Learning from Your Neighbours). | 15 |

Reference

1.Jiawei Han University of Illinois at Urbana-Champaign Micheline Kamber : Data Mining: Concepts and Techniques (Second Edition).
2.Ian H.Witten, Eibe Frank, Morgan Kaufman: Data Mining – Practical Machine Learning Tools and Techniques.
3. GalitShmueli, Nitin Patel, Peter Bruce, (2010): Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner , Wiley.
4. David Hand, HeikkiMannila and Padhraic Smyth: "Principles of Data Mining", MIT PRESS.

**S III / P IV**

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTA304** | **Data Warehousing** | **CC** | **Lectures** | **Practical** | |
| | | | **4** | **-** | **4** |
| **Course Objectives** | | | | | |
| **Learning Outcomes**<br>Upon completion of this course, the student should be able to understand Perform ETL process on source data and send it to data warehouse database. | | | | | |

| | Topic | Lectures |
|---|---|---|
| I | Data Warehousing, Business Intelligence, and Dimensional Modeling Primer Different Worlds of Data Capture and Data Analysis, Goals of Data Warehousing and Business Intelligence, Dimensional Modeling Introduction, Kimball's DW/BI Architecture, Alternative DW/BI Architectures, Dimensional Modeling Myths. | 15 |
| II | Kimball Dimensional Modeling Techniques Overview Fundamental Concepts, Basic Fact Table Techniques, Basic Dimension Table Techniques, Integration via Conformed Dimensions, Dealing with Slowly Changing Dimension Attributes, Dealing with Dimension Hierarchies. | 15 |
| III | Retail Sales Four-Step Dimensional Design Process, Retail Case Study, Dimension Table Details, Retail Schema in Action, Retail Schema Extensibility, Fact less Fact Tables, Dimension and Fact Table Keys, Resisting Normalization Urges. Order Management Order Management Bus Matrix, Order Transactions, Invoice Transactions, Accumulating Snapshot for Order Fulfilment Pipeline. | 15 |
| IV | Inventory Value Chain Introduction, Inventory Models, Fact Table Types, Value Chain Integration, Enterprise Data Warehouse Bus Architecture, Conformed Dimensions. Customer Relationship Management Overview, Customer Dimension Attributes, Bridge Tables for Multivalued Dimensions, Complex Customer Behaviour, Customer Data Integration Approaches, Low Latency Reality Check. | 15 |

References

1. "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling", Ralph Kimball Margy Ross, Wiley.
2. "The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data", Ralph Kimball, Joe Caserta.
3. "Building the Data Warehouse", Fourth Edition, W. H. Inmon, Wiley.

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTAPA301** | **Statistical Computing-I** | **CC** | **Lectures** | **Practical** | |
| | | | | **6** | **3** |

**List of Practical's:**
**Practical 1:** Linear and multiple linear regression
**Practical 2:** Logistics regression
**Practical 3:** Linear discriminant analysis & Quadratic discriminant analysis
**Practical 4:** Clustering using K-means and
**Practical 5:** K-nearest neighbours
**Practical 6:** Big Data Technology-I
**Practical 7:** Big Data Technology-II
**Practical 8:** Big Data Technology-III

| Course code | Course name | Group | Teaching scheme | | credits |
|---|---|---|---|---|---|
| **RJSPGSTAPA302** | **Statistical computing-II** | **CC** | **Lecture** | **Practical** | |
| | | | **-** | **6** | **3** |

**List of Practical's:**
**Practical 1 :**Cross-validation, model evaluation and selection
**Practical 2:** Bootstrap
**Practical 3:** Dimensionality reduction using feature extraction.
**Practical 4:** Dimensionality reduction using feature selection.
**Practical 5:** Polynomial Regression
**Practical 6:** Creating predicated probabilities
**Practical 7:** Creating the database using various constraints
**Practical 8:** Using DDL, DML, DCL and TCL statements.
**Practical 9:** Introduction to ER model and Relational Model.
**Practical 10:** Creating Dimension Model for a Data warehouse.

# Semester IV

| Course Code | Course Name | Group | Teaching Scheme (Hrs/Week) | | Credits |
|---|---|---|---|---|---|
| | | | Lectures | Practical | |
| RJSPGSTA4INT | Industrial Internship | | - | 48 | 24 |
| | **Total** | | **-** | **48** | **24** |