Learning Outcomes-based Curriculum Framework (LOCF) for Post-graduate Programme

Name of the Programme:

M.Sc. Applied Statistics and Data Analytics

(Syllabus effective from 2020 Admission onwards)

UNIVERSITY OF KERALA

Syllabus for M.Sc. Applied Statistics and Data Analytics

Programme Specific Outcomes (PSO) for M.Sc. Applied Statistics and Data Analytics

PSO 1 Expertise in the field of statistical theory and its applications.

PSO 2 Enables to apply more rigorous high performance data mining tools to enhance the knowledge and apply statistical techniques to the real world problems.

PSO 3 Expertise to take up responsibilities as efficient Statisticians/Statistical Officers/Research Officers/Statistical Analytics.

PSO 4 Expertise on data analysis or use of statistical techniques.

PSO 5 Awareness on recent developments in statistical theory and practice.

PSO 6 Enable to become efficient Statisticians through occupying positions in various areas where expertise on data analysis or use of statistical techniques are essential components.

Structure of the Programme

Sem. No.	Course code	Name of the Course	Number of Credits
I	Core Courses (CC)		
	STAD-CC-511	Statistical Mathematics	4
	STAD-CC-512	Distribution Theory	4
	STAD-CC-513	Theory of Sampling	4
	STAD-CC-514	Statistical Computing for Data Analytics using R and Python	4
II	Core Courses (CC)		
	STAD-CC-521	Machine Learning	4
	STAD-CC-522	Time Series and Statistical Forecasting	4
	STAD-CC-523	Multivariate Analysis	4
	STAD-CC-524	Probability Theory	4
	STAD-CC-525	Practical 1 for Data Analytics	2
III	Core Courses (CC)		
	STAD- CC-531	Stochastic Processes	4
	STAD-CC-532	Estimation	4
	STAD-CC-533	Testing of Hypotheses	4
	Discipline Specific		
	Electives (DE)		4
	STAD-DE-534(i)	Statistical Quality Control and	
		Reliability Modeling	4
	STAD-DE-534(ii)	Official Statistics	4
	STAD-DE-534(iii)	Data Mining and their	
		Applications	4
	STAD-DE-534(iv)	Categorical Data Analysis	4
	STAD-DE-534(v)	Longitudinal Data Analysis	
IV	Core Courses (CC)		
	STAD-CC-541	Big Data Analytics and Artificial Intelligence	4
	STAD-CC-542	Design and Analysis of Experiments	4
	STAD-CC-543	Practical 2 for Data Analytics	2

	<u>Discipline Specific</u>		
	<u>Electives (DE)</u>		4
	STAD-DE-544(i)	Optimization Techniques	4
	STAD-DE-544(ii)	Stochastic Finance	4
	STAD-DE-544(iii)	Bayesian Inference	4
	STAD-DE-544(iv)	Order Statistics	4
	STAD-DE-544(v)	Spatial Data Analytics	
			4
	STAD-DE-545(i)	Regression Analysis and	
		Econometric Methods	4
	STAD-DE-545(ii)	Actuarial Statistics	4
	STAD-DE-545(iii)	Biostatistics	4
	STAD-DE-545(iv)	Survival Analysis	4
	STAD-DE-545(v)	Statistical Genetics	
	<u>Dissertation</u>		4
	STAD-CC-546	Project/Internship	
Generic Courses(GC)			
II	STAD-GC-521	Data Analytics Using R	2

STAD – Representing Department of Applied Statistics .CC, DE and GC – Representing Core Courses, Discipline Specific Electives and Generic Courses. First numerical number 5 representing 5th level of Education for PG. Middle numerical number 1,2,3 and 4 representing Semester 1,2,3 and 4 respectively. Last numerical number 1,2,3 etc. representing Course Numbers.

SEMESTER : I
COURSE CODE : STAD-CC-511
COURSE TITLE : STATISTICAL MATHEMATICS
CREDITS : 4

COURSE OUTCOMES

On completion of the course, the students should be able to:

CO1. Explain classes of open and closed sets of R.
CO2. To understand the concept of semi-ring, ring, field, sigma-ring, sigma-field and monotone class with the help of examples.
CO3. Explain the concept of additive and totally additive set functions with the help of certain examples.
CO4. Identify the properties of measurable functions.
CO5. Have a clear understanding of Metric space - Metric in Rn.
CO6. Explain the concept of Cauchy sequence, completeness, compactness and connectedness with the help of example
CO7. State and prove generalized mean value theorem.
CO8. Describe the properties of Riemann- integral.
CO9. State and prove mean value theorem of integral calculus.
CO10. Identify the applications of multiple integrals.
CO11. To achieve ideas on vector space, subspaces, independence of vectors, basis and dimension.
CO12. State and prove Cayley-Hamilton theorem.
CO13. Establish the relation between algebraic and geometric multiplicity.
CO14. To achieve ideas on quadratic forms and reduction of quadratic forms and gets an ability for solving problems in these areas.
CO15. Define Moore-Penrose g-inverse and derive its properties

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy Level**
Module: 1	M01. Explain and exemplify the concepts of Boral field M02. State and prove Hahn-Jordan decomposition theorem. M03. Articulate the basic concept of measure.	Understand Remember Understand

	M04. Give the examples of measure like length, area and volume.	Apply
Module 2.	M01. To be familiar with the idea of open and closed set. M02. Find the limit point of a sequence. M03. State and prove Hein-Boral theorem.	Understand Evaluate Remember
Module: 3	M01. Find the limit and continuity of a given function. M02. Find the conditional maximum and minimum of a given function. M03. Verify generalized mean value theorem to a given function.	Apply Apply Analysis
Module 4	M01. Write the Riemann integral of the given function. M02. State and prove Fubini's theorem.	Apply Remember
Module 5	M01. Find the rank of a given matrix. M02. Understand the concept of determinants and its properties. M03. Solve the system of linear equations.	Evaluate Understand Apply
Module 6	M01. Determine the Eigen values and Eigen vectors of the given matrix M02. Obtain the diagonal form and triangular form of a given matrix. M03. Write down the spectral decomposition of the given matrix M04. Find the nature of the quadratic form. M05. Articulate the concept of generalized inverse.	Evaluate Apply Evaluate Analysis Understand

COURSE CONTENT

MODULE I : Classes of open and closed sets of R, Classes of sets. Monotone class: semi-ring, ring, sigma- ring, field and sigma -field, sigma-field generated by a class of sets. Borel field. Additive and countably additive set functions, Hahn – Jordan decomposition theorem. Outer measure, measure and counting measure, examples of measure like length, area and volume.

MODULE II: Metric space-Metric in Rn, open set, closed set, limit point of a set, sequence in Rn-Cauchy sequence, completeness, compactness, connectedness, Hein-Boral theorem.

MODULE III: Functions in Rn, limit and continuity, differentiability, extreme values of a function (more than one variable), conditional maximum and minimum, generalized mean value theorem.

MODULE IV: Reimann integrals, properties of Reimann Integrals, mean value theorem of integral calculus, multiple integrals, Fubini's theorem, Change of order of integration, transformation of variables.

MODULE V : Vector space and subspaces, independence of vectors, basis and dimensions. Matrices and determinants, rank of a matrix, null space, and nullity, partitioned matrices, Linear transformations, matrix representation of linear transforms. Solution of system of linear equations.

MODULE VI: Eigen values and eigen vectors, algebraic and geometric multiplicity of eigen values, Cayley-Hamilton theorem, Spectral decomposition of Matrices, canonical forms, diagonal form, triangular form, Jordan form. Quadratic forms, reduction of quadratic forms, generalized inverse, Moore-Penrose inverse, Jacobian of transformation, derivative of a function with respect to a vector, with respect a matrix.

REFERENCES

- Kingman and Taylor, S.J. (1977): Introduction to Measure and Probability, Cambridge University Press, Cambridge.
- Roydon, H. L. (1968) : Real Analysis, Macmillan, New York.
- Rudin, W.(1970) : Real and Complex Analysis, McGraw Hill Book Co., New York.
- Apostol,T.M.(1987): Mathematical Analysis, 2nd edn, Narosa Publishing House, New Delhi.
- Goldberg,R.R.(1970): Methods of Real Analysis, Oxford & IBH publishing Co. (P) Ltd. NewDelhi.
- Graybill,A and Belmont, C.A.(1983): Matrices with Applications in Statistics, II Edition, John Wiley, New York.
- Pringle,R.M. and Rayner,A. A(1971): Generalised Inverse of Matrices with Application to Statistics, Griffin, London.
- Rao,C.R (1973): Linear Statistical Inference and its Applications, Wiley Eastern, New York.

SEMESTER **:** I
COURSE CODE **:** STAD-CC-512
COURSE TITLE **:** DISTRIBUTION THEORY
CREDITS **:** 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. understand the properties of probability density functions and cumulative distribution functions.

CO2. Define expectation, and be introduced to its important linearity property.

CO3. Calculate raw moments and central moments, including their special cases, the mean and variance.

CO4. Calculate the moment generating function, and appreciate its link to moments.

CO5. Realize the difference between discrete and continuous probability distributions.

CO6. Finding conditional and marginal distributions from a bivariate probability distribution.

CO7. Apply transformation of variable technique for finding the distribution of functions of random variables and solve related problems

CO8. Explains different sampling distributions and find expressions for their probability density function.

CO9. Explain the concept of order statistics and solving problems related to it

CO10. Distinguish between partial and multiple correlation and the concept of empirical distribution function.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. Explains the basic concepts of pgf, mgf and characteristic functions M02. Examine various types of discrete and continuous distributions and articulate their properties M30. Examine the various properties of probability distribution. M04. Solving problems related to several distributions	Understand Analysis Evaluate Apply
Module 2	M01. Finding the joint, marginal and conditional pdf of a bivariate distribution M02. Articulate the various concepts of Conditional expectation and conditional variance M03. Explains the concept of pgf, mgf and independence of a bivariate random vector M04. Articulate Multinomial and bivariate normal distributions and their properties in detail	Understand Analysis Remember Apply
Module 3	M01. Apply transformation of variable technique for finding the distribution of functions of random variables and solve related problems M02. Derive the distributions of sum, product and ratios of random variables	Apply Analysis

Module 4	M01. Explain the concept of order statistics M02. Find the joint distribution of two order statistics M03. Calculate Distribution of functions of two order statistics	Understand Evaluate Apply
Module 5	M01. Explains different sampling distributions and M02. Find expressions for probability density function of sampling distribution	Understand Apply
Module 6	M01. Distinguish between partial and multiple correlation coefficient M02. Finding the Distribution of empirical distribution function M03. Explain the concept of generalized standard deviation.	Understand Evaluate Remember

COURSE CONTENT

MODULE I : Basic concepts in distribution theory : p.g.f., m.g.f., and characteristic function. Univariate distributions: Binomial, Negative Binomial, Poisson, Hyper geometric, Geometric, Beta, Gamma, Normal, Log-Normal, Pareto, Weibull, Cauchy. Laplace, Logistic, Log-logistic. Rayleigh and Generalized exponential distributions.

MODULE II : Bivariate distributions: Joint, conditional and marginal distributions. Conditional expectation, conditional variance. Independence, p.g.f. and m.g.f. of bivariate random vector. Multinomial and bivariate normal distributions and their properties. Compound, Truncated and mixture distributions.

MODULE III : Functions of random variables and their distributions using transformation of variable technique: Distributions of sum, product and ratios of random variables.

MODULE IV : Order statistics: Distribution of order statistic, Joint distribution of two order statistics, Distribution of functions of two order statistics.

MODULE V : Sampling distributions: Chi-square, t distribution, and F distributions (both central and non-central) and their applications.

MODULE VI : Empirical distribution; Distribution of empirical distribution function. Correlation and regression; simple partial and multiple correlation coefficients. Generalized standard deviation.

REFERENCES

- Anderson,T.W.(1984) Introduction to Multivariate Statistical Analysis, Macmillan

Publishing Company.
• Fisz,M (1963) Probability Theory and Mathematical Statistics, 3rd Edition, John Wiley.
• Hogg,R.V. and Craig ,A.T.(1989) Introduction to Mathematical Statistics, Macmillan Publishing Company
• Johnson ,N.L. and Kotz,S. (1969) Distributions in Statistics; Discrete distributions. John Wiley and Sons ,New York.
• Johnson ,N.L. . Kotz,S. and Balakrishnan, N (1994)Continuous Univariate Distributions -1 ,2nd Edition John Wiley and Sons ,New York.
• Johnson ,N.L. . and Kotz,S. (1995) Continuous Univariate Distributions -2 ,2nd Edition, John Wiley and Sons ,New York..
• Rohatgi,V.K.(1990) An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd.

SEMESTER **:** I
COURSE CODE **:** STAD-CC-513
COURSE TITLE **:** THEORY OF SAMPLING
CREDITS **:** 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Understand the principles underlying sampling as a means of making inferences about a population.

CO2. Collect data from a smaller part of a large group so that the students can able to learn something about the larger group.

CO3. understand the difference between randomization theory and model based analysis.

CO4. understand the concepts of bias and sampling variability and strategies for reducing these.

CO5. Analyse data from multi-stage surveys.

CO6. have an appreciation of the practical issues arising in sampling studies.

MODULE OUTCOMES

Sl. No:	Outcomes	Taxonomy level

	On completion of each module, students should be able to:	
Module 1	M01. Distinguish between ordered and unordered sampling method M02. Use various procedures of selecting samples from the population M30. Explain method of selecting samples using simple random sampling. M04. Articulate the estimation of population mean, total and proportion and hence the estimation of the standard error. M05. Explain determination of sample size and confidence interval	Understand Analysis Evaluate Apply
Module 2	M01. Explain PPS sampling and articulate to selection procedures using PPS with and without replacement M02. Find methods to estimate population mean, total and variance with respect to PPS sampling M03. Apply various selection procedures for selecting samples using PPS	Evaluate Analysis Remember
Module 3	M01. Explain the purpose of stratification M02. Explain stratified random sampling M03. Examine the various properties of stratified sampling M04. Explain the concept of double sampling M05. Discuss various allocation procedures	Apply Analysis
Module 4	M01. Explain systematic sampling M02. Distinguish between single start linear systematic sampling and multiple start linear systematic sampling M03. Describe the comparison of simple random sample, systematic sample and stratified sample for a population with linear trend	Understand Evaluate Apply
Module 5	M01. Distinguish between ratio and regression estimators M02. Explain various properties of ratio and regression estimators M03. Discuss ratio and regression estimator for stratified population	Understand Apply
Module 6	M01. Explain cluster sampling. M02. Distinguish between clusters of equal and unequal sizes. M03. Describe the two stage cluster sampling of equal size and unequal size.	Understand Evaluate Remember

COURSE CONTENT

MODULE I: Probability sampling. Simple random sampling with replacement and without replacement. Procedures of selection, Estimation of population mean, total, proportion and variance. Estimation of standard errors of these estimators. Confidence intervals. Determination of sample size.

MODULE II: Sampling with varying probabilities: Probability proportional to size (PPS) sampling, Procedure of selecting a PPS sample with and without replacement. Estimation of population mean, total and variance in PPS sampling with replacement. Estimated standard error of estimator of population mean and total. Estimation of population mean in PPS sampling without replacement. Des Raj ordered estimator, Murthy's unordered estimator, Horvitz-Thompson estimator and their estimated standard errors .Yates –Grundy estimator, Midzuno-Sen scheme of sampling, IPPS sampling.

MODULE III: Stratified random sampling: Purpose of stratification, Estimation of population mean and its variance, Various methods of allocation, comparison with unstratified simple random sampling, Estimation of population mean with post-stratification, Effect of using inaccurate stratum size, concept of double sampling and its applications in stratified random sampling when strata sizes are unknown and for Neymann method of allocation, Estimation withstratified PPS sample.

MODULE IV: Systematic sampling; single start linear systematic sampling and multiple start systematic sampling, Circular systematic sampling, selection procedures. Advantages and disadvantages, Estimation of population mean and its variance, Comparison of systematic sampling with simple random sampling, Comparison of systematic sample, SRS and stratified random sample for a population with linear trend.

MODULE V: Estimation of population mean using auxiliary information: Ratio estimator and its properties . First order expression of bias, mean square error and comparison with simple arithmetic mean estimator. Optimum property of ratio estimator. Sampling schemes that provide unbiased ratio estimator. Hartley-Ross unbiased ratio type estimator. Confidence interval of population ratio. Ratio estimator with double sampling. Optimum procedure with cost consideration. Separate and combined ratio estimator for stratified population. Regression estimator and its properties First order expression of bias, mean square error. Comparison with simple arithmetic mean estimator and ratio estimator. Optimum property . First order expression of bias, mean square error and comparison with simple arithmetic mean estimator. Optimum property. Regression estimator with double sampling. Separate and combined regression estimators.

MODULE VI: Cluster sampling with equal and unequal clusters. Estimators of population mean and their estimated standard errors. Relative efficiency of cluster as unit of sampling compared to a single element as unit of sampling. Two stage sampling: Two stage sampling with equal first stage units-estimator of population mean, its standard error and its estimator. Two stage sampling with unequal first stage units-estimators of population mean and their properties. Estimation of standard error of estimators of population mean. Two stage sampling with equal first stage units and sampling is done according to PPS at the first stage and simple random sampling at the second stage. Estimation of the population mean and its variance. Non sampling errors, Various factors, methods of reducing non-response.

REFERENCES

- Cochran , W.G. (1977). Sampling Techniques, Wiley Eastern Ltd.
- Murthy, M. N. (1967). Sampling Theory and Methods, Statistical Publishing Society, Calcutta.
- Singh, D. and Chaudhary, F.S. (1986). Theory and Analysis of Sample Survey Designs, Wiley Eastern Ltd.
- Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). Sampling Theory of Surveys With Applications, Indian Society of Agricultural Statistics, New Delhi.

SEMESTER : I
COURSE CODE : STAD-CC-514
COURSE TITLE : STATISTICAL COMPUTING FOR DATA ANALYTICS USING R AND PYTHON
CREDITS : 4

COURSE OUTCOMES

After completion of this course the students will be able to

C01.Define the basic concepts of R software and R packages
C02.Describe various concepts required for developing the R Language
C03.Build our own new functions in R
C04.Illustrate different R-Graphics facilities
C05.Apply Python Programming Language to develop different statistical concepts
C06.Interpret the Object Oriented Programming concept using Python Programming

MODULE OUTCOMES

Sl. No.	Outcomes On Completion of each module, Students should be able to:	Taxonomy level

Module 1.	M01.Define basic concepts of statistical software R such as Basic operations in R, Mathematical functions used in R, Assign values to variables etc. M02.Demonstrate the important data structures such as arrays, matrix, data frames, Class function etc. M03. Illustrate using help facilities in R M04. Summarize an overview of R packages	Remember Apply Apply **Understand**
Module.2	M01.Design an overview of the R Language such as Expressions, Objects, Symbols, Functions, Special Values, R Syntax- Constants, Numeric vectors, Character vectors, Symbols, Operators, Order of operations, Assignments, Expressions. M03. Demonstrate the use of control Structures- Conditional Statements, Loops, Accessing data Structures, M04. Illustrate the use of R Objects such as Vectors, Lists, Matrices, Arrays, Factors, Data Frames, Formulas, Time series.	Create Apply Apply
Module.3	M01.Describe the use of Functions- The function keyword, Arguments, Return values, Functions as arguments, properties of functions M02. Demonstrate the use of writing functions in R. M03.Perform working with data – Entering data within R, Importing data from external files, Exporting data, Combining data sets, Merging data.	**Understand** Apply Apply
Module 4	M01. Demonstrate the use of R-Graphics: An overview of R graphics, Scatterplots, Bar charts, Histogram, Pie charts. M02. Illustrate the use of Plotting distributions, Plotting time series, Box plots, Stem and lead plot,Q-Q plots. M03. Summarize the use of Graphical parameters, Basic graphic functions. M04. Demonstrate drawing of mathematical functions, Logarithmic functions, Trigonometric functions, polynomial functions.	Apply Apply Understand Apply
Module 5	M01. Define basic concepts of Python Programming Language, Python Data Types: Expressions, Variables and Assignments, Strings, List, Objects and Classes, Python Standard Library, Imperative Programming: Python programs M02. Demonstrate the Execution Control Structures, User-Defined Functions, Python Variables and Assignments, Parameter Passing. Text Files: Strings, Formatted Output, Files, Errors and Exception Handling, Execution and Control Structures: if Statement, for Loop, Two Dimensional Lists, while Loop, More Loop Patterns,	Remember Apply

	Additional Iteration Control Statements, M03. Illustrate Containers and Randomness: Dictionaries, Other Built-in Container Types, Character Encoding and Strings, Module random, Set Data Type.	Apply
Module 6	M01. Describe the use of Object Oriented Programming: Fundamental Concepts, Defining a New Python Class M02. Design User-Defined Classes, Designing New Container Classes, Overloaded Operators, Inheritance, User-Defined Exceptions, Namespaces: M03. Illustrate Encapsulation in Functions, Global versus Local Namespaces, Exception Control Flow, Modules and Namespaces.	Understand Create Apply

COURSE CONTENT

MODULE I: Introduction to the statistical software R, Basic operations in R, Mathematical functions used in R, Assign values to variables, Introduction to data structures - arrays, matrix, data frames. Class function, Getting help, An overview of R packages

MODULE II: An overview of the R Language- Expressions, Objects, Symbols, Functions, Special Values. R Syntax- Constants, Numeric vectors, Character vectors, Symbols, Operators, Order of operations, Assignments, Expressions, Control Structures- Conditional Statements,Loops, Accessing data Structures, R Objects- Vectors, Lists, Matrices, Arrays, Factors, Data Frames, Formulas, Time series.

MODULE III: Functions- The function keyword, Arguments, Return values, Functions as arguments, properties of functions, Writing functions in R. Working with data – Entering data within R, Importing data from external files, Exporting data, Combining data sets, Merging data.

MODULE IV : R-Graphics: Graphics- An overview of R graphics, Scatterplots, Bar charts, Histogram, Pie charts, Plotting distributions, Plotting time series, Box plots, Stem and lead plot,Q-Q plots, Graphical parameters, Basic graphic functions, Drawing- mathematical functions,Logarithmic functions, Trigonometric functions, polynomial functions

MODULE V: Introduction: The Process of Computational Problem Solving, Python Programming Language, Python Data Types: Expressions, Variables and Assignments, Strings, List, Objects and Classes, Python Standard Library, Imperative Programming: Python programs, Execution Control Structures, User-Defined Functions, Python Variables and Assignments, Parameter Passing. Text Files: Strings, Formatted Output, Files, Errors and Exception Handling, Execution and Control Structures: if Statement, for Loop, Two Dimensional Lists, while Loop, More Loop Patterns, Additional Iteration Control Statements, Containers and Randomness: Dictionaries, Other Built-in Container Types, Character Encoding and Strings, Module random,Set Data Type.

MODULE VI: Object Oriented Programming: Fundamental Concepts, Defining a New Python Class, User-Defined Classes, Designing New Container Classes, Overloaded Operators, Inheritance, User-Defined Exceptions, Namespaces: Encapsulation in Functions, Global versus Local Namespaces, E66xception Control Flow, Modules and Namespaces.

REFERENCES

• Everitt, B.S. and Hothorn T. (2010) A Handbook of Statistical Analysis Using R, Second Edition, CRC Press.
• Joseph Adler (2011) R in a Nutshell, Second Edition, Shroff Publishers and Distributors Pvt. Ltd.
Michael J. Crawley (2013) The R book, Second Edition, John Wiley & Sons Ltd.
• Rubinstein, R.Y. (1981) Simulation and Monte Carlo Methods, Wiley.
• Perkovic, L. (2011). Introduction to computing using python: An application development focus. Wiley Publishing.
• McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc."

SEMESTER: II
COURSE CODE: STAD-CC-521
COURSE TITLE: MACHINE LEARNING
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:
CO1. Analyze the data and understand the insights from it.
CO2. A clear understanding of machine learning algorithms and applications.
CO3. Apply various ML techniques as per the requirements
CO4. Calculate the maximum likelihood estimation
CO5. Differentiate bias and variance estimation
CO6. To calculate the number of clusters for clustering algorithms
CO7. The full derivation of forward and backward propagation in neural network

MODULE OUTCOMES

Sl. No:	**Outcomes** **On completion of each module, students should be able to:**	**Taxonomy Level**
Module 1.	M01. Explain about machine learning and applications M02. Thorough understanding about supervised,semi-supervised and unsupervised learning algorithms	Understand Understand
Module 2.	M01: Introduction of estimation methods M02: To get an idea about the inference of the data (population) M03: Use of different types of classifications in data	Understand Understand Apply
Module 3.	M01. Explain about various clustering algorithms M02. Introduction to data mining M03. Apply the concepts of clustering algorithms into different types of data. M04. Apply data reduction techniques	Understand Understand Apply Apply
Module 4.	M01. A better understanding of SVM M02. Apply SVM algorithms for face detection M03. Analysis using data validation techniques M04. Explain about data warehouse M05. Various applications of data warehouse	Understand Apply Evaluate Understand Understand
Module 5.	M01: Good knowledge in both Multivariate analysis and missing data M02: To get the understanding of the relationship between more than two variables through multivariate analysis	Apply Apply
Module 6.	M01: Introduction and importance of dimensionality reduction. M02: Understanding of methods used for the dimensionality reduction.	Understand Apply

COURSE CONTENT

MODULE I: Introduction: Machine Learning, Applications, Supervised Learning: Learning a Class from Examples, Unsupervised Learning, Semi Supervised learning , Dimensions of a Supervised Machine Learning Algorithm, Vapnik -Chervonenkis (VC) Dimension, Probably Approximately Correct (PAC) Learning, Noise, Learning Multiple Classes.

MODULE II: Model Selection and Generalization, Bayesian Decision Theory: Introduction, Classification, Losses and Risks, Discriminant Functions, Utility Theory, Association Rules. Parametric Methods: Introduction, Maximum Likelihood Estimation, Evaluating an Estimator-Bias and Variance, The Bayes' Estimator, Parametric Classification, Regression, Tuning Model Complexity: Bias/Variance Dilemma.

MODULE III: Data Reduction and Classification introduction to Data mining, ,Clustering: Introduction, Mixture Densities, k-Means Clustering, Nearest Neighbour Method, Supervised Learning after Clustering, Hierarchical Clustering, Choosing the Number of Clusters, Principles of Decision Trees, Neural Network and Random Forests

MODULE IV: Support Vector Machine, Naive Bayes: Methods and their implementation using R.. Meaning of Analytics, Components of Data Architecture, Data Warehouse, Column oriented data structure. Parallel vs Distributed Computing. Data Analytics Lifecycle: Discovery, Data validation, Model formulation, Model building, Transduction, Learning to learn, reinforced learning, Training Data, use of regression and classification methods for implementation.

MODULE V: Model Selection Procedures, Multivariate Methods: Multivariate Data, Parameter Estimation, Estimation of Missing Values, Multivariate Normal Distribution, Multivariate Classification, Tuning Complexity, Discrete Features, Multivariate Regression. Mixtures of Latent Variable Models, Expectation-Maximization Algorithm.

MODULE VI: Dimensionality Reduction: Introduction, Subset Selection, Principal Components Analysis, Factor Analysis, Multidimensional Scaling, Linear Discriminant Analysis, Isomap, Locally Linear Embedding

REFERENCES

- Alpaydin, E. (2009). Introduction to machine learning. MIT press.
- Trevor, H., Robert, T., & JH, F. (2009). The elements of statistical learning: data mining, inference, and prediction.
- Bhat, B. R. (1985). Modern Probability Theory: An Introductory Text Book, 2nd Edition, Wiley Eastern.
- Brian Coffo. Statistical Inference for Data Science.
- Tan, T., Steinbach, M. and Kumar, V. (2006): Introduction to Data Mining, Pearson Education.
- Gupta, G.K. (2008): Introduction to Data Mining with case studies, Prentice – Hall of India Pvt. Ltd.
- Daniel T. Larose (2006): Data Mining: Methods and Models, John Wiley and sons. (relevant portions of Chapter 4).

SEMESTER : II
COURSE CODE : STAD-CC-522
COURSE TITLE : TIME SERIES AND STATISTICAL FORECASTING
CREDITS : 4

COURSE OUTCOMES

After completion of this course the students will be able to

C01. Generate model selection techniques using the concept of regression models
C02. List different forecasting techniques using regression models.
C03. Choose an appropriate model for time series data using the concept of linear time series models
C04. Analyze exploratory time series analysis using time series data
C05. Predict the time series data using different stationary time series data
C06. Describe estimation methods for time series

MODULE OUTCOMES

Sl.No.	Outcomes On Completion of each module, Students should be able to:	Taxonomy level
Module 1.	M01.Develop Regression models, estimation and testing of regression coefficients M02.Perform Model selection techniques, problem of correlated errors, auto correlation, M03.Describe Durbin Watson statistic, detection and correction of multicollinearity, indicator variables.	Create Apply Understand
Module.2	M01.Explain the Method of least squares for curve fitting, concept of weighted least squares M02. Illustrate Nonlinear regression and fitting of quadratic, exponential and power curves M03.Forecasting using regression models.	Remember Apply Apply
Module.3	M01. Explain Time series, components of time series, exploratory time series analysis M02. Perform Test for trend and seasonality M03 Demonstrate Exponential and moving average smoothing, Holt-winter smoothing, M04.Forecasting based on smoothing.	Remember Apply Apply Apply
Module 4	M01. Explain the Stationary time series, Autocorrelation, partial auto correlation function M02. Develop Linear stationary models: auto regressive, moving average and mixed processes. M03.Apply Linear non-stationary models- Autoregressive integrated moving average (ARIMA) models.	Remember Create Apply
Module 5	M01 Forecasting using ARMA and ARIMA models: MMSE methods M02. Forecasting using different forms of the ARIMA models	Apply Apply
Module 6	M01. Develop Estimation techniques: Yule Walker methods and	Create

	least squares method M02. Analyze time series data using Correlogram and periodogram	Analyze

COURSE CONTENT

MODULE I: Regression models, estimation and testing of regression coefficients, model selection techniques, problem of correlated errors, auto correlation, Durbin Watson statistic, detection and correction of multicollinearity, indicator variables.

MODULE II: Method of least squares for curve fitting, concept of weighted least squares, nonlinear regression and fitting of quadratic, exponential and power curves, forecasting using regression models.

MODULE III: Time series: components of time series, exploratory time series analysis, test for trend and seasonality, exponential and moving average smoothing, Holt-winter smoothing, forecasting based on smoothing.

MODULE IV: Stationary time series, Autocorrelation, partial auto correlation function, linear stationary models: auto regressive, moving average and mixed processes. Linear non-stationary models- Autoregressive integrated moving average (ARIMA) models.

MODULE V: Forcasting using ARMA and ARIMA models: MMSE methods, using different forms of the ARIMA models

MODULE VI : Estimation: Yule Walker methods and least squares method, correlogram and periodogram analysis.

REFERENCES

- Box, G.E.P., Jenkins G.M. and Reinsel, G.C. (2007) Time Series Analysis, Forecasting and Control, Pearson Education.
- Brockwell,P J. and David R. A. (2002). Introduction to time series and forecasting, 2nd edition, Springer.
- Draper, N, R. and Smith H. (1998). Applied Regression Analysis, 3rd edition. John Wiley
- Ezekiel,M (1963). Methods of Correlation and Regression Analysis.
- Kleinbaum, D, G., Kupper, L,L, Muller, K, E and Nizam, A (1998). Applied regression analysis and multivariable methods, Duxbury Press.
- Makridakis, S and Wheelwright, S C. Forecasting methods and applications, John Wiley and Sons
- Kutner, M, H, Nachtsheim, C, J and Neter J, (2004). Applied linear regression models, 4th edition with student CD Mc Graw Hill

SEMESTER **:** II
COURSE CODE **:** STAD-CC-523
COURSE TITLE **:** MULTIVARIATE ANALYSIS
CREDITS **:** 4

COURSE OUTCOME

On completion of the course, students should be able to:
CO1: Describe multivariate normal distribution and its properties
CO2: Find the marginal and conditional distribution of multivariate normal distribution
CO3: Find the distribution of quadratic forms of multivariate normal vectors
CO4: Describe Wishart distribution and its properties
CO5: Obtain the estimators for parameters of multivariate normal distribution
CO6: Describe multiple and partial correlation coefficients
CO7: Define sample multiple and sample partial correlation coefficients for multivariate normal vector
CO8: Test the hypothesis regarding parameters of multivariate normal distribution
CO9: Use Hotelling's T^2 and Mahalanobis D^2 statistics for testing hypothesis
CO10: Perform Multivaraite data analysis

MODULE OUTCOME

MODULE	**Module outcomes** *On completion of each module, students should be able to:*	**Taxonomy Level**
MODULE 1	MO1: Define the multivariate normal density function. MO2: Obtain the characteristic function of multivariate normal density MO3: Find the distribution of linear combination of multivariate normal random vector using characteristic function MO4: Find regression of multivariate normal distribution using conditional distribution	Remember Understand Apply Apply
MODULE 1I	MO1: Characterize quadratic forms of multivariate distribution. MO2: Obtain the distribution of sums and quotients of quadratic forms. MO3: Apply Cocharan's theorem to find distribution of quadratic forms of multivariate normal random vector	Apply Apply Apply
MODULE III	MO1: Obtain the MLEs of mean and variance of multivariate normal distribution MO2 : Find the characteristic function of Wishart distribution MO3 : Show that Whishart distribution possess additive property MO4 : Find the distribution of sample dispersion matrix	Evaluate Apply Understand Remember

MODULE IV	MO1: Find the distribution of sample multiple correlation for multivariate normal distribution MO2: Find the distribution of sample partial correlation coefficients for multivariate normal distribution MO3: Obtain the distribution of test statistic for the test the significance of correlation coefficient and partial correlation coefficient.	Remember Apply Apply
MODULE V	MO1: Test the mean vector of a multivariate normal distribution MO2: Test the equality of means of two or more multivariate normal distributions MO3: Use Hotelling's T^2 and Mahalanobis D^2 statistics in testing hypothesis regarding multivariate normal distributions. MO4: Find the relationship between Hotelling's T^2 and Mahalanobis D^2 statistics	Apply Apply Apply Understand
MODULE VI	MO1: Perform principal component analysis and factor analysis MO2: Classify individuals/items in to one of k multivariate normal populations MO3: Identify canonical variables and quantify canonical correlation	Analysis Analysis Analysis

COURSE OUTCOME

MODULE I: Multivariate normal distribution, properties, characteristic function, standard characteristics, marginal and conditional distributions, distribution of linear combinations of normal variates.

MODULE II: Distribution of quadratic forms in normal variables, distribution of sums and quotient of independent quadratic forms, Cochran's theorem**.**

MODULE III: Samples from multivariate normal distribution, M.L.E. of mean vector and dispersion matrix, distribution of sample mean vector, Wishart distribution: definition, analogy with chi-square distribution, characteristic function, additive property, generalized variance, partitioned Wishart matrix, Distribution of sample dispersion matrix.

MODULE IV : Sampling distribution of correlation matrix and simple correlation coefficient, multiple correlation coefficient, partial correlation coefficient, distribution of the sample multiple correlation and partial correlation under null case, tests of significance.

MODULE V : Tests of hypothesis about mean vector of a multivariate normal distribution, equality of means of two multivariate normal distributions, Hotelling's T^2, Mahalanobi's D^2.

MODULE VI : Classification problem- classifying to one of k multivariate normal populations, Bayes solution, Fisher's discriminant function, principal component analysis; canonical variables and canonical correlations, basics of factor analysis and cluster analysis.

REFERENCES

- Anderson, T.W. (2003) :An Introduction to Multivariate Statistical Analysis, John Wiley, New York.
- Graybill, F. A. (1961): An Introduction to Linear Statistical Model, Vol. 1, McGraw Hill, New York.
- Johnson, R and Wychern (1992): Applied Multivariate Statistical Analysis, Prentice hall, London.
- Kendall, M. G. (1958) :A Course in Multivariate Analysis, Griffin, London.
- Khatri, C.G. and Srivastava (1979) : An Introduction to Multivariate Statistics, North – Holland, New York.
- Muirhead, R.J. (1982): Aspects of Multivariate Statistical Theory, John Wiley & Sons New York.
- Rao, C.R.(1973) : Linear Statistical Inference and its Applications, Wiley Eastern, New York.
- Rohatgi, V.K.(1976): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd, New York.

SEMESTER : II
COURSE CODE : STAD-CC-524
COURSE TITLE : PROBABILITY THEORY
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Identify a probability measure and explain its properties
CO2. Solve problems based on various properties of a probability measure
CO3. Apply the concepts of Bayes theorem and solve related problems
CO4. Distinguish between a discrete and continuous type random variables and illustrate with examples
CO5. Verify the properties of important functions of random variables
CO6. Calculate the expectation and moments of random variables and random vectors
CO7. Identify the applications of various moment inequalities

CO8. Explain the concept of convergence and check for the of convergence of a given sequences of random variables.
CO9. Find the expressions for the characteristic function of a random variable and verify its properties
C10. Apply the various laws of large numbers to sequences of random variables.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy Level**
Module 1.	M01. Explain and exemplify the concepts of probability space M02. Examine the various properties of a probability measure. M03. State and explain various results associated with probability measure M04. Explain the concepts of independence of events M05. Construct examples of independent events M06. Construct counter examples for proving/illustrating certain results associated with probability measure M07. Evaluate the conditional probability and verify its properties M08. Articulate the Bayes theorem and apply it to calculate apriori probabilities	Understand Analysis Understand Remember Create Create Evaluate Apply
Module: 2	M01. Distinguish between various types of random variables and articulate their properties M02. Describe the various functions of random variables and find the functions for various random variables M03. Explain and exemplify the concepts of decomposition of distribution function of a random variable M04. Articulate/exemplify various types of distributions and their important properties	Apply Apply Understand Understand
Module 3	M01. Calculate the mathematical expectation, moments and generating functions of random variables M02. Explain the concepts of random vectors and distribution function of random vectors, and their important properties M03. Articulate the various moment inequalities M04. Apply the various moment inequalities to various distributions	Evaluate Understand Understand Apply
Module 4	M01. Articulate and appraise stochastic convergence of sequence of random variables M02. Apply the concepts of convergence to sequences of random variables M03. Construct counter examples for not satisfying certain convergence implications	Understand Evaluate Create
Module 5	M01. Derive expressions for the characteristic function for various distributions	Evaluate

	M02. Find moments using characteristic function M03. Derive expressions for the probability density function corresponding to a given characteristic function M04. Articulate the various theorems associated with the characteristic function and identify their applications M05. Explain and exemplify the concepts of infinitely divisible distributions	Evaluate Evaluate Apply Understand
Module 6.	M01. State and prove the various laws of large numbers M02. Apply the laws to sequences of random variables M03. Articulate the concepts of stable distributions M04. Explain the concepts of Stable distributions and distribution of class L	Understand Apply Understand Understand

COURSE OUTCOME

MODULE I: Probability space, limit of sequence of events, monotone and continuity properties of probability measure, addition theorem, independence of finite number of events, sequence of events, tail events and tail fields, Borel Cantelli lemma, Borel zero one law. Conditional Probability and Bayes Theorem.

MODULE II: Random variable, its probability distribution and distribution function, properties of distribution function, decomposition of distribution function, discrete and continuous type random variables, discrete & continuous and other types of distributions

MODULE III: Mathematical expectation, moments of random variables, random vectors, independence of random variables and sequence of random variables, Markov, Chebyshev's and Lyapounov inequalities.

MODULE IV: Stochastic convergence of sequence of random variables:- convergence in distribution, convergence in probability, almost sure convergence and convergence in the rth mean, their interrelationships, examples and counter examples, Helly's and Helly-Bray theorems (statements only).

MODULE V: Characteristic function and their elementary properties, uniform continuity and nonnegative definiteness of characteristic function, characteristic function and moments, inversion theorem (proof not required), uniqueness theorem, Fourier inversion theorem, continuity theorem, Bochner-Khintchine theorem of characteristic functions(proof not required), infinite divisibility of distributions(definition and elementary properties).

MODULE VI: Stochastic convergence of series of random variables:-Law of large numbers, weak law of large numbers: Bernoulli, Chebyshev & Khintchine, Kolmogorov inequality, strong law of large numbers: Kolmogorov- iid & non-iid cases (proof not required), central limit theorem: Classical, Demoiver-Laplace, Liapounov and Lindberg-Feller (without proof), Stable distributions: distribution of class L, stable distributions & domain of attraction (definition and examples only).

REFERENCES

- Bhat, B.R. (1985): Modern Probability Theory: An Introductory Text Book, 2nd edition, Wiley Eastern.
- Gnedenoko, B.V. (1969): The Theory of Probability, Mir Publishers, Moscow.
- Laha, R.G. and Rohatgi, V.K. (1979): Probability Theory, John Wiley, New York.
- Loeve, M. (1968): Probability Theory, D.Van Nostrand Co.Inc., Princeton, New Jersey.

SEMESTER : II
COURSE CODE : STAD-CC-525
COURSE TITLE : PRACTICAL- I FOR DATA ANALYTICS
CREDITS : 2

COURSE OUTCOMES

After completion of this course the students will be able to

C01. Evaluate the trend and seasonal fluctuations of a real life time series data using R programming language
C02. Formulate different Index numbers for economic data
C03. Demonstrate different sampling techniques using real life data set
CO4. Outline relative efficiency of sampling procedures
C05. Illustrate different statistical tests using the software SPSS
C06. Interpret different multivariate techniques using SPSS

MODULE OUTCOMES

Sl.No.	Outcomes On Completion of each module, Students should be able to:	Taxonomy level
Module 1.	M01. Evaluate the Trend component of a time series using different methods like Method of Semi-Averages, Method of Curve fitting , Moving average Method Using the R programming M02. Generate the Seasonal component of a time series using different method like method of simple averages, Ratio to trend method, Ratio to moving average method, Link relative method Using the R programming	Evaluate Create
Module.2	M01. Generate different Index numbers using different methods like Simple Aggregate method, Weighted	Create

	aggregate method: Laspeyres, Paasches', Bowley, Marshall-Edgeworth, Fisher index numbers, Computation of cost of living index numbers Using the R programming M02. Calculate different measures of mortality such as CDR, SDR, completing of missing life table, computation of measures of Fertility such as CBR, GFR, SFR, TFR, GRR, NRR Using the R programming	 Apply
Module.3	M01. Design different sampling techniques such as Simple random sample with and without replacement, PPS sample, with and without replacement, Midzuno-Sen sample ,linear systematic sample, circular systematic sample, multiple start systematic sample Using the R programming. M02.Determine different estimates such as Estimate of population mean (total) and estimate of standard error of the estimate of mean(total based upon a simple random sample with or without replacement, linear systematic sample, circular systematic sample, multiple start systematic sample ,stratified random sample, cluster sample, two stage sample). Estimated gain in efficiency due to stratification Using the R programming.	Create Apply
Module 4	M01. Outline relative efficiency of cluster as a unit of sampling compared to an element as unit of sampling, Relative efficiency of two stage sampling and systematic sampling Using the R programming. M02. Evaluate Ratio and regression estimator of population mean based on a simple random sample without replacement Using the R programming.	Remember Evaluate
Module 5	M01. Perform tests of hypothesis such as Compare means- one sample t- test, two sample test, independent sample t test, chi-square test using the software SPSS. M02. Illustrate the concepts of correlation and Regression-simple and multiple using the software SPSS.	Apply Apply
Module 6	M01. Perform tests of hypothesis using multivariate data such as Equality of means of two multivariate normal vectors M02.Analyze the multivariate data using the multivariate techniques such as Canonical correlation, Principal component analysis, Factor analysis, Cluster analysis.	Apply Analyze

COURSE CONTENT

MODULE 1: (Using R programming) Measurement of the Trend: Method of Semi-Averages, Method of Curve fitting , Moving average Method, Measurement of Seasonal Fluctuations: Method of simple averages, Ratio to trend method, Ratio to moving average method, Link relative method.

MODULE 1I: (Using R programming) Calculation of the Index numbers: Simple Aggregate method, Weighted aggregate method: Laspeyres, Paasches', Bowley, Marshall-Edgeworth, Fisher index numbers, Computation of cost of living index numbers Computation of different measures of mortality: CDR, SDR, completing of missing life table, computation of measures of Fertility: CBR, GFR, SFR, TFR, GRR, NRR.

MODULE III: (Using R programming) Selection of a random sample :Simple random sample with and without replacement, PPS sample, with and without replacement, Midzuno-Sen sample ,linear systematic sample, circular systematic sample, multiple start systematic sample. Estimation; Estimate of population mean (total) and estimate of standard error of the estimate of mean(total based upon a simple random sample with or without replacement, linear systematic sample, circular systematic sample, multiple start systematic sample ,stratified random sample, cluster sample, two stage sample. Estimated gain in efficiency due to stratification.

MODULE 1V: (Using R programming) Relative efficiency of cluster as a unit of sampling compared to an element as unit of sampling, Relative efficiency of two stage sampling and systematic sampling. Ratio and regression estimator of population mean based on a simple random sample without replacement.

MODULE V: (Using SPSS) Descriptive Statistics ; Compare means- one sample t- test, two sample test, independent sample t test ; chi-square test ; Correlation; Regression-simple and multiple.

MODULE VI: (Using SPSS) Equality of means of two multivariate normal vectors; Canonical correlation, Principal component analysis, Factor analysis, Cluster analysis.

SEMESTER : III
COURSE CODE : STAD- CC-531
COURSE TITLE : STOCHASTIC PROCESSES
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Describe and exemplify concepts of Stochastic processes, time space and state space,

classification of stochastic processes based on the nature of time space and state space, Classical stochastic processes like processes with stationary independent increments, Markov process, renewal process, martingales, Wiener process, Gaussian process
CO2. Explain Markov chains: Definition, transition probability matrix, n-step transition Probability and Chapman-Kolmogorov equation
CO3. Calculate n-step transition probabilities
CO4. Classify states of a finite Markov chain
CO5. Describe periodicity and ergodicity of chains.
CO6. Describe limiting beha vior of n-step transition probabilities,
CO7.Obtain the stationary distribution of a Markov chain
CO8. Describe random walks & gambler's ruin problem.
CO9. Explain and exemplify continuous time Markov chain, Poisson process, pure birth process, birth and death processes, compound Poisson process, Markov Process with discrete states.
CO10. Distinguish between strict and weak (covariance or wide sense) stationarity,
CO11. Explain and exemplify renewal processes, renewal equation.
CO12. Describe and apply renewal theorem.
CO13. Describe Branching processes, offspring distribution, extinction probabilities.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy Level**
Module 1.	M01. Articulate and exemplify the concepts of Stochastic processes, time space and state space. M02. Construction of examples of Stochastic processes M03. Explain the concepts of particular types of stochastic processes like process with stationary independent increments, Markov process, martingales, Wiener process, Gaussian process etc.	Understand Understand Create Apply
Module: 2	M01. Articulate concepts of Markov chains, transition probability matrix, n-step transition probabilities M02. Calculate n-step transition probabilities M03. Describe and exemplify classification of states in a Markov Chain M04. Calculate the periodicity of a Markov Chain M05. Explain the concepts of recurrence, ergodic chains	Understand Evaluation Understand Evaluation Understand
Module 3	M01. Explain and exemplify concepts of limiting behaviour of n-step transition probabilities. M02. Describe stationary distributions and solve problems M03. Describe various types of random walks M04. Explain a gambler's ruin problem M05. Derivation of the probability of ruin of a gambler's ruin problem	Understand Apply Understand Understand Evaluation
Module 4	M01. Describe and exemplify: Continuous time Markov chains,	Understand

	Poisson process, pure birth process, birth and death processes. MO2. Derive of steady state probabilities/differential difference equations in case of Poisson process, pure birth process, birth and death processes. M03. Describe and exemplify: Compound Poisson process, M04 Derive properties of Poisson process and Compound Poisson process M05. Explain the concept of Markov Process with discrete states. M06. Illustrate these processes with examples	Evaluation Understand Evaluation Understand Apply
Module 5	M01. Distinguish between strict and weak (covariance or wide sense) stationarity M02. Describe and exemplify: renewal processes, renewal equation. M02. Explain the statement and applications of renewal theorem M03. Solve problems based on the applications of renewal theorem	Analyse Understand Apply Apply
Module 6	M01. Describe and exemplify Galton-Watson branching processes M02. Explain concepts of offspring distribution and its implications M03. Interpret the concept of extinction probabilities M04. Compute the probability extinction in case of a particular offspring distribution.	Understand Apply Apply

COURSE CONTENT

MODULE I: Introduction to Stochastic processes, time and state space, classification of stochastic processes, processes with stationary independent increments, Markov process, renewal process, martingales, Wiener process, Gaussian process (definitions and examples).

MODULE II: Markov chains: Definition, transition probability matrix, n-step transition probability, Chapman-Kolmogorov equation, calculation of n-step transition probability and its limit, classification of states, periodicity, recurrence, ergodic chains,

MODULE III: limiting behavior of n-step transition probabilities, stationary distributions, random walk & gambler's ruin problem.

MODULE IV: Continuous time Markov chains, Poisson process, pure birth process, birth and death processes, compound Poisson process, Markov Process with discrete states.

MODULE V: Stationary processes, strict and weak (covariance or wide sense) stationarity, renewal processes, renewal equation, statement and applications of renewal theorem.

MODULE VI: Branching process: Galton-Watson branching processes, offspring distribution, extinction probabilities.

REFERENCES

- Bartlett, M.S. (1955): An Introduction to Stochastic Processes (with special reference to application and methods), Cambridge.
- Bhat, U.N. and Miller, G.K. (2002): Elements of Applied Stochastic Processes. 3rd Edn., John Wiley, New York.
- Feller, W. (1968): Introduction to Probability Theory and its Applications, Vols. I & II, John Wiley, New York.
- Karlin, S. (1972): A First Course in Stochastic Processes, Academic Press, New York.
- Parzen, E. (1962): Stochastic Processes, Holden-Day Inc, San Francisco.
- Srinivasan, S.K. and Mehata, K.M. (1976): Stochastic Processes, Tata McGraw-Hill Publishing Company Limited, New Delhi.

SEMESTER : III
COURSE CODE : STAD-CC-532
COURSE TITLE : ESTIMATION
CREDITS : 4

COURSE OUTCOMES

On completion of this course, the students will be able to**:**

C01:List the important properties of estimators of an unknown parameter of a distribution
C02:Derive the UMVUE of a parameter or function of a parameter
C03:Apply the concept of Rao-Blackwell and Lehmann-Scheffe theorems
C04:Able to select the best estimators using different properties
C05: Differentiate between classical and Bayesian inference
C06: Determine the estimators of unknown parameters using methods like MLE, Method of moments etc.
C07: Differentiate between location and scale family of distributions
C08:Outline Bayes estimation of parameters of standard distributions

MODULE OUTCOMES

Sl.No.	Outcomes On Completion of each module, Students should be able to:	Taxonomy level
Module 1.	M01. Derive the important properties of estimators M02.Determine the sufficient statistic using NP lemma M03.Explain minimal sufficiency M04.Define Basu's Theorem M05.Use likelihood equivalence to obtain minimal sufficient statistic	Create Evaluate Understand Remember Apply
Module.2	M01. Develop a characterization theorem for finding UMVUE M02. Apply Rao-Blackwell and Lehmann-Scheffe theorems to find UMVUE M03. Determine UMVUE estimator of any parametric function	Create Apply Apply
Module.3	M01. Derive Fisher information measure M02. Calculate Cramer-Rao inequality, Chapman -Robbin's bound, Bhattacharya bounds. M03. Identify Efficient estimators M04. Evaluate Consistent estimators	Create Apply Remember Evaluate
Module 4	M01. Describe different methods of estimation such as method of moments, MLE, Minimum chi-square, Least square estimation etc. M02. Derive properties of the estimators	Understand Create
Module 5	M01. Formulate location and scale family of distributions M02. Evaluate Location and scale invariant estimators M03. Calculate Pitman estimators of location and scale parameters M04. Discuss BLUE for location and scale distributions.	Create Evaluate Apply Understand
Module 6	M01. Design basic elements of Bayesian Inference M02. Define Bayes Theorem M03.Calulate Bayes estimators of parameters of standard distributions. M04.Differentiate between classical and Bayesian Inference	Create Remember Apply Analyze

COURSE CONTENT

MODULE I: Point estimation, Sufficiency and minimal sufficiency, Neyman-Pearson factorization theorem, Exponential family of distributions, Pitman family, Likelihood equivalence, Unbiased estimation; Completeness, Basu's Theorem.

MODULE II: UMVUE estimators and their characterizations, Methods of finding UMVUE, Rao-Blackwell and Lehmann-Scheffe theorems, UMVUE estimation of parametric function from standard distributions.

MODULE III: Fisher information measure and its properties, Lower bound to the variance of an unbiased estimates, Cramer-Rao inequality. Chapman -Robbin's bound, Bhattacharya bounds,Efficiency, Consistency.

MODULE IV: Methods of estimation: Method of moments, Maximum likelihood estimators and their properties, Minimum chi-square and its modification, Least square estimation.

MODULE V: Location and scale family of distributions, Location and scale invariant estimators, Pitman estimators of location and scale parameters. BLUE for location and scale distributions.

MODULE VI: Basic elements of Bayesian Inference, Loss function, Bayes risk, Prior distribution, Bayes Theorem, Posterior distributions, Bayes estimation of parameters of standard distributions.

REFERENCES

- Hogg, R. V. and Craig, A. T. (1989): Introduction to Mathematical Statistics, Macmillan Publishing Company.
- Kale, B. K. (1999): A First Course on Parametric Inference, Narosa Publishing House.
- Kendall, M.G. and Stuart, A. (1967): The Advanced Theory of Statistics, Vol. II, 2nd Edition, Charles Griffin & Company Ltd., London.
- Lehmann, E. L. (1983): Theory of Point Estimation, John Wiley, New York.
- Mood, A. M., Graybill, F. A. and Boes, D. C.(1972): Introduction to the Theory of Statistics, 3rd Edition, Mc-Graw Hill International, New York.
- Mood, A. M., Graybill, F. A. and Boes, D. C.(1972): Introduction to the Theory of Statistics, 3rd Edition, Mc-Graw Hill International, New York.
- Rajagopalan M. and Dhanavanthan, P. (2012) Statistical Inference, PHI Learning Private Limited, New Delhi.
- Rohatgi, V. K. (1990): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Limited.
- Wasan, M. T. (1970): Parametric Estimation, Mc-Graw Hill, New York.

SEMESTER : III
COURSE CODE **:** STAD-CC-533
COURSE TITLE **:** TESTING OF HYPOTHESES
CREDITS **:** 4

COURSE OUTCOME

On completion of the course, students should be able to:
CO 1: Formulate hypothesis for a given problem
CO2: Find critical region and power of the test
CO3: Find most powerful test for testing simple hypothesis against simple alternative
CO4: State and prove Neyman-Pearson lemma
CO5: Find UMP test for testing composite hypothesis
CO7: Derive likelihood ratio test for testing the hypothesis for normal populations
CO8: Obtain sequential probability ratio test for testing the hypothesis.
CO9: Construct confidence interval for parameters
CO10: Construct UMA and UMAU confidence sets
C011: Perform non-parametric test for a given data

MODULE OUTCOME

MODULE	**Module outcomes** *On completion of each module, students should be able to:*	**Taxonomy Level**
MODULE 1	MO1: Identify simple and composite hypothesis MO2: Find critical region, size and power of the test MO3: Distinguish between random and non-randomized test	Understand Remember Apply
MODULE 1I	MO1: Apply Neyman-Pearson lemma to find most powerful test MO2: Check the unbiasedness of a test MO3: Find UMP and UMPU test	Apply Apply Apply
MODULE III	MO1: Apply likelihood ratio test principle for testing the mean for a normal population. MO2: Apply likelihood ratio test principle for testing the equality of means for two normal population MO3: Test the variance of normal population. MO4: Test the equality of means of two normal populations	Apply Apply Apply Apply
MODULE IV	MO1: Derive SPRT for test the parameters of normal distribution, binomial and Poisson distributions MO2: Find OC function and Average sample Number of a SPRT	Apply Remember
MODULE V	MO1: Construct shortest level confidence interval for parameters MO2: Construct confidence interval using pivot MO3: Use UMP and UMPU test to construct UMA and UMAU confidence sets	Evaluate Evaluate Apply
MODULE VI	MO1: Perform one sample and two sample non-parametric test	Apply

	MO2: Perform Kruskal-Wallis one-way test and Friedman two-way test	Apply

COURSE CONTENT

MODULE I: Basics of Testing of hypothesis, simple hypotheses and composite hypotheses, critical regions and test functions, randomized and nonrandomized tests,

MODULE II: Neyman-Pearson lemma and its applications, most powerful tests, UMP tests, Unbiasedness, UMPU.

MODULE III: Likelihood ratio tests, asymptotic properties, tests concerning normal distribution (one sample and two samples).

MODULE IV: Sequential procedures, SPRT, Wald's identity, OC and ASN functions, applications to Binomial, Poisson and Normal distributions

MODULE V: Confidence sets, shortest confidence intervals, construction of confidence interval using pivots, UMA confidence sets, UMAU confidence sets, relation to tests of hypotheses.

MODULE VI: Nonparametric tests- Test based on Mann-Whiteny U statistics, Kolmogorov-Smirnov one sample and two sample tests, Sign test, Wilcoxon signed rank test, run test. Median test. Kruskal-Wallis one-way analysis of variance by ranks, Friedman two way analysis of variance by ranks. Kendall's rank order correlation coefficient and Kendall's coefficient of concordance as measure of association.

REFERENCES

- Ferguson, T. S. (1967): Mathematical Statistics, Academic Press, New York.
- Kendall, M.C, and Stuart. A, (1967): The Advanced Theory of Statistics, Vol 2, IV Edn., Mc Millan, New York.
- Lehmann, E. L. (1986): Testing of Statistical Hypothesis, John Wiley & Sons.
- Wald, A.(1977): Sequential Analysis, Dover Publications Inc., New York.
- Wetherill, G. B. (1966): Sequential Methods in Statistics, Methuen & Co. Ltd. New Delhi.
- Siegel, S. and Castellan Jr. N. J. (1988): Non-parametric Statistics for the Behavioral Sciences. McGraw Hill, New York.

- Randles, R.H. and Wolfe, D.A. (1979): Introduction to the Theory of Nonparametric statistics, Wiley New York.
- Rohatgi, V.K.(1976): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd, New York.

SEMESTER **:** III
COURSE CODE **:** STAD-DE-534(i)
COURSE TITLE **:** STATISTICAL QUALITY CONTROL AND RELIABILITY MODELING
CREDITS **:** 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Understand the concept of quality control statistical process control.

CO2. Measure the performance of a process.

CO3. Identify assignable causes.

CO4. Define reliability including the different types and how they assessed.

CO5. Ensure the validity and precision of statistical analysis.

CO6. Explains reliability in discrete set up.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. Explain statistical process control M02. Describe various control charts	Understand Analysis
Module 2	M01. Describe different sampling inspection techniques M02. Explains six-sigma concepts M03. Distinguish chain sampling and continuous sampling	Evaluate Analysis Remember
Module 3	M01. Explain basic concepts of reliability M02. Describe reliability of a coherent system M03. Describe several measures of reliability	Apply Analysis Understand

	M04. Calculate measures of reliability based on several distributions	Apply
Module 4	M01. Explain reliability in discrete set up M02. Derive expression of the relation connecting various measures of reliability	Understand Evaluate
Module 5	M01. Explain inference in reliability models M02. Describe estimation of parameters based on censored sampling MO3. Solve problems related to estimation of parameters based on censored sampling	Understand Apply Apply

COURSE CONTENT

MODULE I: Statistical process control, Theory of control charts – Shewart control charts for variables- $\bar{X}$, R,S charts, Attribute control charts - np, p, c and u charts – OC, ARL & process capability of control charts, CUSUM charts, Acceptance sampling for attributes and variables.

MODULE II: Sampling inspection techniques: Single, double and multistage sampling plans and their properties, Chain sampling, Continuous sampling, Taguchi method, Total quality management, ISO standardization, ISO 9001, six sigma concepts.

MODULE III: Basic reliability concepts: Reliability concepts and measures, Components and systems, coherent systems, reliability of coherent systems, cuts and paths, series and parallel system, k-out-of-n systems, Bounds on System Reliability. Failure rate, mean residual life, Mean time to failure in the univariate cases, Exponential, Weibull, Pareto, Inverse Gaussian and Gamma as life distribution models, Characterization of life distribution based on failure rate and mean residual life function.

MODULE IV: Reliability concepts in discrete set up, Notion of ageing based on failure rate and mean residual life, NBU, NBUE, HNBUE classes and their duals, Interrelationships.

MODULE V: Inference in reliability models: Estimation of parameters based on complete and censored samples in exponential, Weibull and Gamma models. Non-parametric estimation of failure rate and reliability function.

REFERENCES

- Barlow, R.E. and Proschan, F. (1985): Statistical Theory of Reliability and Life Testing, Holt, Rinehart and Winston.
- Cox, D.R. and Oakes, D. (1984): Analysis of Survival Data, Chappman Hall.
- Duncan, A. J. (1959): Quality Control and Industrial Statistics (5th edition), Irwin, Homewood I.
- Galambos, J. and Kotz, S. (1978) Characterization of Probability Distributions.
- Klefjo, B. (1982) The HNBUE and HNWUE Classes of Life distributions, Naval Research Logistic Quarterly, 29, 331-344.

• Lawless, J. F. (2003): Statistical Models and Methods for Lifetime Data, John Wiley.
• Montgomery, D.C. (2005): Introduction to Statistical Quality Control, 5th edition, John Wiley.
• Nelson, W. (1982): Applied life data analysis, Wiley.
• Sinha, S. K. (1986) Reliability and Life Testing, Wiley.

SEMESTER : III
COURSE CODE : STAD-DE-534(ii)
COURSE TITLE : OFFICIAL STATISTICS
CREDITS : 4

COURSE OUTCOME

On completion of the course, students should be able to:

CO1. Have a detailed understanding on the working of official statistics system and its various responsibilities.

CO2. Understand population and related issues.

CO3. Explain the importance of various family and health schemes by the government.

CO4. Recall various statistics and index numbers prevailing in our country.

CO5. Explain the procedure of national income estimation.

CO6. Have a general awareness about basic concepts in using statistical techniques.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. A general understanding of the official statistics system that prevails in our country and others. M02. Responsibilities of various central and state divisions under the Indian official statistics system. M03. Dissemination and use of data collected.	Understand Evaluate Apply
Module 2	M01. Learn about population and related issues. M02. Learn various health and family welfare schemes. M03. Understand in detail the Population Census in India.	Understand Analysis Remember
Module 3	M01. Understand various other social statistics.	Understand

	M02. Understand various index numbers used in the official statistics system.	Analysis
Module 4	M01. Learn the economic aspects of growth and development. M02. Estimation of national income through various approaches.	Understand Apply
Module 5	M01. Understand various income inequality measures. M02. Understand various poverty measurement methods.	Understand Apply

.

COURSE CONTENT

MODULE I: Introduction to Indian and International Statistical systems. Role, function and activities of Central and State Statistical organizations. Organization of large-scale sample surveys. Role of Nation Sample Survey Organization. General and special data dissemination systems

MODULE II: Population growth in developed and developing countries, Evaluation of performance of family welfare programmers, projections of labor force and man power. Scope and Content of population census of India.

MODULE III: Statistics related to Industries, foreign trade, balance payment, cost of living, inflation, educational and other social statistics.

MODULE IV: Economic Development: Growth in per capita income and distributive justice indices of development, human development index. National income estimation-Product approach, income approach and expenditure approach.

MODULE V: Measuring inequality in incomes: Gini Coefficient, Theil's measure; Poverty measurement: Different issues, measure of incidence and intensity; Combined Measure: indices due to Kakwani, Sen etc.

REFERENCES

- Basic Statistics Relating to Indian Economy (CSO) 1990
- Guide to official Statistics (CSO)1999
- Statistical System in India (CSO)1995
- Principles and accommodation of National Population Census, UNEDCO
- Panse, V.G.: Estimation of Crop Yields (FAO)
- Family Welfare Year Book. Annual Publication of D/O Family Welfare
- Monthly Statistics of Foreign Trade in India, DGCIS, Calcutta and other Govt.
- Publications.
- CSO (1989) a: National Accounts Statistics- Source and Methods
- Keyfitz, N (1997): Applied Mathematical Demography- Springer Verla.
- Sen, A (1997): Poverty and Inequality
- UNESCO: Principles for Vital Statistics systems, Series M-12
- CSO (1989)b: Statistical Systems in India

SEMESTER : III
COURSE CODE : STAD-DE-534(iii)
COURSE TITLE : DATA MINING AND THEIR APPLICATIONS
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Able to learn the concept of data base technology which has led to the need for data mining and its applications

CO2. Examine the types of data to be mined and present a general classification of task to integrate data mining system.

CO3. Evaluate and select appropriate data mining algorithms and apply, interpret and report the output appropriately.

CO4. Apply statistical methods for any given raw data.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. Explain classification methods for data M02. Establish decision trees	Understand Analysis
Module 2	M01. Distinguish clustering from statistical and data mining M02. Explain vector quantization	Understand Analysis
Module 3	M01. Explain dimension reduction M02. Analyze unsupervised learning from univariate and multivariate data	Apply Analysis
Module 4	M01. Explain supervised learning from moderate to high dimensional input spaces M02. Analyze regression trees	Understand Evaluate
Module 5	M01. Explain simple relation data bases M02. Analyze online analytical data processing M03. Establish the applications to electronic commerce	Understand Apply

COURSE CONTENT

MODULE I: Data mining- History-Definitions-Data Mining Functionalities- Classification of Data mining System- Major Issues in Data mining-Data warehouse and OLAP Technology-Multidimensional Data Model-Data warehouse Architecture- Data Warehouse Implementation.

MODULE II: Data Preprocessing-Data Cleaning- Data Integration and Transformation- Data Reduction-Discretization and concept of Hierarchy Generation- Concept Description characterization and comparison. Association Rule Mining- Mining Single Dimensional – Multilevel Association Rules-mining to correlation analysis-classification and prediction

MODULE III: Overview on outliers – nature of Outliers - Outliers in Univariate Data – Outliers in Multivariate Data - Cluster Analysis, Cluster Vs Classification - impact of Outliers on clustering - clustering problems - Clustering Approaches.

MODULE IV: Data-outliers in regression analysis and Time series - Regression and collinearity: Tools for handling multi- collinearity, methods based on singular value decomposition – Robust Regression- ridge regression. Properties of ridge estimator. Addtive outlier – Multiplicative outlier and innovational outlier.

MODULE V: Stationary time Series, Auto correlation and Partial auto correlation function, Correlogram analysis, Estimation of ARIMA model parameters, forecasting with Box – Jenkins model, Residual analysis and diagnostic checking.

REFERENCES

- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C (2004). Time Series Analysis- Forecasting and Control, Pearson Education, Singapore.
- Daniel T. Larose, (2006): Data Mining: Methods and Models, Wiley-Interscience, New Jersey.
- Draper, N.R, and H. Smith, Applied regression analysis,(2nd Ed) John Wiley and sons, New York.
- Hawkins, D.M, (1980): Identification of Outliers, Chapman and Hall, London.
- Jiawei Han, Micheline Kamber, (2006): Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, second edition, San Francisco.
- Krzysztof J.Cios, Wiltold Pedrycz, Roman W.Swiniarski, Lukasz A.Kurgan, (2007): Data Mining: A Knowledge Discovery Approach, Springer Science +Business Media, New York.
- Montgomery, D.C. and Johnson, L.A. (1977) Forecasting and Time Series Analysis, McGraw Hill, New York.
- Paolo Giudici, (2005): Applied Data Mining: Statistical Methods for Business and Industry, John Wiley & Sons Ltd, England.
- Peter J.Rousseeuw and Annick M.Lorey, (1987): Robust Regression and Outlier Detection, John Wiley & Sons, United States.
- Vic Barnett and Toby Lewis, (1978): Outliers in Statistical Data, John Wiley & sons.

SEMESTER : III
COURSE CODE : STAD-DE-534(iv)
COURSE TITLE : CATEGORICAL DATA ANALYSIS
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Analyze categorical data and make conclusion from it.

CO2. Learn the applications of statistical techniques for fitting log-linear models, logit models, estimation theory for parametric models etc.

CO3. Model Binary and multinomial response variables.

CO4. Explain logistic regression and its analysis.

C05. Test for independence in two –way and three- way contingency tables.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. How to model binary response variables. M02. Explain logit model.	Understand Analysis
Module 2	M01. Model multinomial response variables. M02. Analyse repeated categorical response data.	Understand Analysis
Module 3	M01. Analyse two way and three way contingency tables. M02. Model data with repeated measurements. M03. Explain logistic regression.	Understand Analysis Remember
Module 4	M01. Tests for independence and homogeneity of proportions for contingency tables. M02. Explain Fishers exact test. M03. Understand Simpson's Paradox.	Understand Analysis Understand
Module 5	M01. Explain Polytomous logit models. M02. How to model for multiway tables.	Understand Analysis

COURSE CONTENT

MODULE I: Models for Binary Response Variables, Log Linear Models, Fitting Log linear and Logic Models-Building and applying Log Linear Models, Log- Linear- Logit Models for Ordinal Variables.

MODULE II: Multinomial Response Models - Models for Matched Pairs- Analyzing Repeated Categorical Response Data - Asymptotic Theory for Parametric Models - Estimation Theory for Parametric Models.

MODULE III: Classical treatments of 2 and 3-way contingency tables, measures of association and nonparametric methods - Generalized linear models - Logistic regression for binary – multinomial and ordinal data - Log-linear models - Poisson regression- Modelling repeated measurements - generalized estimating equations.

MODULE IV: Introduction to contingency tables: $2 \times 2$and $r \times c$ tables - tests for independence and homogeneity of proportions - Fishers exact test - Odds ratio and Logit, other measures of association - Introduction to 3-way tables – full independence and conditional independence - collapsing and Simpsons paradox.

MODULE V: Polytomous logit models for ordinal and nominal response - Log-linear models (and graphical models) for multi-way tables - Causality, repeated measures, generalized least squares - mixed models, latent-class models, missing data, and algebraic statistics approach.

REFERENCES

- Agresti, Alan (1996). An Introduction to Categorical Data Analysis, Wiley.Breiman, L.
- Friedman, J.H. Olshen, R.A. and Stone, C.J. (1984): Classification and Regression Trees.
- Wadsworth and Brooks/Cole.
- Bergsma, W., Croon, M.A. and Hagenaars, J.A. (2009). Marginal Models: For
- Dependent, Clustered, and Longitudinal Categorical Data. Springer.
- Bishop, Y.M., Fienberg, S.E. and Holland, P.W. (1975). Discrete Multivariate Analysis:
- Theory and Practice, MIT Press.
- Fienberg, S.E. (1980). The Analysis of Cross-Classified Categorical Data. MIT Press.
- Wasserman, L. (2004). All of Statistics: A Concise Course in Statistical Inference.
- Springer.
- Whittaker, J. (1990). Graphical Models in Applied Multivariate Statistics. Wiley.

SEMESTER : III

COURSE CODE : STAD-DE-534(v)
COURSE TITLE : LONGITUDINAL DATA ANALYSIS
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Understand various approaches for analyzing longitudinal survey data, including methods for handling complex surveys, weights and non-response.

CO2. Identify the important issues when analyzing longitudinal survey data.

MODULE OUTCOMES

Sl. No:	**Outcomes** On the completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. Explain and articulate the background of linear, count and binary regression analysis. M02. Exemplify the various concepts on Linear, log-linear and logistic regression models for continuous, count and binary data. M03. Explain the different methods of estimation and comment on the asymptotic properties of these estimators for regression effects.	Understand Remember Understand
Module 2	M01. Exemplify the concept of longitudinal/panel data. M02. Illustrate the different types of longitudinal data with examples.	Remember Apply
Module 3	M01. Explain how to fit a binary dynamic logits (BDL) model to the longitudinal stationary/non-stationary binary data. M02. Distinguish between binary dynamic logits (BDL) and binary auto-correlation models. M03. Comment of the concept of correlation structure under the BDL model. M04. Exemplify the concept of GQL and likelihood estimation. Also, illustrate and estimate the model using	Apply Evaluate Analysis

	asthma data.	Apply
Module 4	M01. Explain the concept of stationary multinomial dynamic logits (MDL) model and illustrate the model using Three Miles Island Stress Level (TMISL) data.	Understand
Module 5	M01. Describe the analysis of longitudinal count and binary data in mixed effects model setup. M02. Explain the AR(1) type mixed effect models for count and binary data. M03. Exemplify the method of GQL estimation for regression and random effects variance parameters and illustrate it with suitable data.	Remember Evaluate Analysis

COURSE CONTENT

MODULE I: Background of linear, count and binary (exponential family responses) regression analysis in independent setup. Linear, log-linear and logistic regression models for continuous, count and binary data, respectively. Estimation methods: Ordinary least square (OLS), Method of moments (MM), Quasi-likelihood (QL). Asymptotic properties (consistency and relative efficiency) of these estimators for regression effects.

MODULE II: Back ground of longitudinal/panel data. Illustrations of longitudinal data: A rat data example for longitudinal linear data, A health care utilization (HCU) data example for longitudinal count data, An asthma data example for longitudinal binary data, Stationary (time independent) and non-stationary (time dependent) fixed covariates.

MODULE III: Fitting a binary dynamic logits (BDL) model to the longitudinal stationary or non-stationary binary data, Difference between BDL and binary auto-correlation models. Correlation structure under the BDL model. The GQL and likelihood estimation. Illustration of the model and estimation with an asthma data.

MODULE IV: Extension of the stationary BDL model to the stationary multinomial dynamic logits (MDL) model for longitudinal categorical data analysis. Illustration with the Three Miles Island Stress Level (TMISL) data.

MODULE V: Analysis of longitudinal count and binary data in mixed effects model setup. Random effects variance affecting the longitudinal correlations. The AR(1) type mixed effect models for count and binary data. Fitting a BDML (binary dynamic mixed logits) model to the longitudinal binary data. The GQL estimation for the regression and random effects variance parameters. Illustrations with the HCU data for longitudinal count data, and with SLID (Survey of Labor Income Dynamics) data for longitudinal binary data.

REFERENCES

- Longitudinal Categorical Data Analysis (2014) by B. C. Sutradhar, Springer.
- Applying Generalized Linear Models (1997) by J. K. Lindsey, Springer.
- Generalized Linear Models (1989) by P. McCullagh and J. A. Nelder, Chapman and Hall.
- Analysis of Longitudinal Data (2002) by P. Diggle, P. Heagerty, K-Y Liang, and S. L. Zeger, 2nd edition, Oxford University Press, Oxford.

SEMESTER: IV
COURSE CODE: STAD-CC-541
COURSE TITLE: BIG DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE
CREDIT: 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Understand when a data becomes Big Data
CO2. Choose appropriate technology for processing Big Data problems
CO3. store and scale large fast growing of Data
CO4. Build architecture for solving complex data
CO5. Get good understanding of Hadoop and HDFS Ecosystem
CO3. Can implement Distributed computing to process large data

MODULE OUTCOMES

Sl. No:	**Outcomes** **On completion of each module, students should be able to:**	**Taxonomy Level**
Module 1.	M01. Introduction to Big Data M02. Understanding the properties of Big Data M03. Explanation of HDFS and Map Reduce M04. Setup a Hadoop Cluster	Understand Analysis Remember Apply
Module 2.	M01. Deep understanding of HDFS architecture M02. Understanding Hadoop YARN, M03. Describe Hadoop Ecosystems and components	Remember Understand Analysis
Module 3.	M01. Understand distributed computation with map reduce	Understand

	M02. Explore mapreduce configuration M03. Store in nosql database using Hbase M04. Implement MapReduce	Analysis Understand Apply
Module 4.	M01. Understanding the architecture of YARN	Understand
Module 5.	M01. Building machine learning models using mahout M02. Understand how Machine learning is applied on large scale data M03. What is Nosql and when to use Nosql	Apply Understand Apply
Module 6.	M01. Applications of Big Data in different Domains	Apply

COURSE CONTENT

MODULE I: Big Data – Introduction, Structuring Big Data, Elements of Big data, Big data analytics, Big data applications. Big Data in business context.

MODULE II: Technologies for handling big data – Distributed and Parallel computing for Big Data, Data Models, Computing Models, Introducing Hadoop – HDFS and MapReduce.

MODULE III: Understanding Analytics and Big data – Comparison of Reporting and Analysis, Types of Analytics, Analytical approaches. Hadoop EcoSystem, Hadoop Distributed file system, HDFS architecture, MapReduce, Hadoop YARN, Introducing HBase, Hive and Pig

MODULE IV: MapReduce framework, Techniques to Optimize MapReduce, Uses of MapReduce, Role of HBase in Big data processing, Processing Data with MapReduce – Framework, Developing simple MapReduce Application.

MODULE V: MapReduce execution and Implementing MapReduce Programs, YARN Architecture – Limitations of MapReduce, Advantages of YARN, Working of YARN, YARN Schedulers, Configurations, Commands, Containers

MODULE VI: Introduction to Mahout – Machine Learning, Clustering, Classification, Mahout Algorithms, Environment for Mahout. Introduction to NoSQL.

REFERENCES

• Berson, A. and Smith, S.J. (1997): Data Warehousing, Data Mining, and OLAP. McGraw-Hill.
• Breiman, L. Friedman, J.H. Olshen, R.A. and Stone, C.J. (1984): Classification and Regression Trees. Wadsworth and Brooks/Cole.
• Han, J. and Kamber, M. (2000): Data Mining; Concepts and Techniques. Morgan Gaufmann.
• Mitchell, T.M. (1997): Machine Learning. McGraw-Hill.

• Ripley, B.D. (1996): Pattern Recognition and Neural Networks. Cambridge University Press.

COURSE CODE : STAD-CC-542
COURSE TITLE : DESIGN AND ANALYSIS OF EXPERIMENTS
CREDITS : 4

COURSE OUTCOME

On completion of the course, students should be able to:
CO 1: Identify estimability of a linear parametric function
CO2: Apply Gauss-Markov theorem for finding BLUE of a parametric function
CO3: Identify connected, balance and orthogonal designs
CO4: Perform one way and two way analysis of variances
CO5: Design and analyse CRD, RBD, LSD and GLSD
CO6: Design and analyse factorial experiments
CO7: Apply principle of total and partial confounding in factorial experiments
CO8: construct incomplete block design
CO9: Analyze BIBD and PBIBD
CO10: Perform analysis of covariance in one way and two way classification.
CO11: Perform missing plot and mixed-up plots analysis in RBD and LSD
CO12: Estimate missing yields in split plot and BIBD

MODULE OUTCOME

MODULE	Module outcomes *On completion of each module, students should be able to:*	Taxonomy Level
MODULE 1	MO1: Identify estimability of a linear parametric function MO2: Apply Gauss-Markov theorem for finding BLUE of a parametric function. MO3: Formulate the model for one way and two way classification MO4: Perform one way and two analysis of varaince	Understand Apply Understand Analysis
MODULE 1I	MO1: Identify connected, balance and orthogonal designs MO2: Design and analysis of CRD, RBD, LSD and GLSD	Understand Analysis
MODULE III	MO1: Design and analyse of 2^2, 3^n and p^n factorial experiments MO2: Apply principle of total and partial confounding in	Analysis Apply

	factorial experiments	
MODULE IV	MO1 : Analyze BIBD with inter and intra block informations MO2: Analyse PBIBD with only two associates classes	Apply Apply
MODULE V	MO1:AnalysesSplit-plot and split-split plot designs MO2: Formulate the model for one way and two way analysis of covariance MO3: Perform one way and two way analysis of covariance	Apply Apply Apply
MODULE VI	MO1:Perform Missing plot analysis in RBD and LSD MO3: Perform mixed-up plot analysis in RBD and LSD MO2: Estimate missing yields in split plot and BIBD	Analysis Analysis Evaluate

COURSE CONTENT

MODULE I: General linear models, estimability of linear parametric functions, Gauss-Markov theorem, tests of linear hypothesis, one-way classification, two-way classification with equal and unequal number of observations per cell.

MODULE II: Randomization, replication and local control, Block designs- information matrix of block designs, criteria for connectedness, balance and orthogonality. Standard designs: CRD, RBD, LSD, GLSD. Efficiency of designs and comparison.

MODULE III: Factorial designs: Statistical analysis of symmetrical factorial designs. Total and partial confounding in 2^n, 3^n and p^n experiments. Concepts of fractional replication.

MODULE IV: Incomplete block design, BIBD, analysis with recovery of inter block information and intra block information, PBIBD and analysis of PBIBD with only two associates classes.

MODULE V: Split-plot and split-split plot designs. Strip-plot design. Analysis of covariance technique in standard designs.

MODULE VI: Missing and mixed plot analysis in RBD, LSD. Estimation of missing yields in split plot design and BIBD.

REFERENCES

- Aloke Day (1986). Theoryof Block Designs .Wiley Eastern, New Delhi.
- Chakrabarti,M.C.(1962) : Mathematics of Design and Analysis of Experiments , Asia Publishing House, Bombay.
- Das, M. N. and Giri, N. (1979). Design and Analysis of Experiments. Wiley Eastern Limited, New Delhi.
- John, P.W.M. (1971). Statistical Design and Analysis of Experiments, Macmillan.
- Joshi, D. D. (1987): Linear Estimation and Design of Experiments, Wiley Eastern, Wiley

Eastern Limited, New Delhi.

- Montgomery,C.D.(1976): Design and Analysis of Experiments, John Wiley, New York.

SEMESTER : IV
COURSE CODE : STAD-CC-543
COURSE TITLE : PRATICAL 2 FOR DATA ANALYTICS
CREDITS : 2

COURSE OUTCOMES

After completion of this course the students will be able to

C01.Formulate and evaluate different estimators for unknown parameters using the real life data set using R programming language
C02.Choose an appropriate UMP test for a real life data set using R programming language
C03.Classify the different control charts for industrial data set using R programming language
C04.Perform the appropriate nonparametric tests for a data set using R programming language
C05.Summarize different design of experiments for a statistical experimental data using R programming language
C06. Recognize different nonparametric tests such as ANOVA, RBD etc. For statistical experimental data using R programming language.

MODULE OUTCOMES

Sl.No.	Outcomes On Completion of each module, Students should be able to:	Taxonomy level
Module 1.	M01. Justify different estimation techniques – Maximum likelihood estimation, Uniformly minimum variance unbiased estimate, method of moments, method of minimum chi-square using the R software.	Evaluate
Module.2	M02. Develop different Testing of hypothesis procedures such as Most powerful test, Uniformly most powerful test, Uniformly most powerful unbiased test, Locally most powerful test, Likelihood ratio tests, Sequential probability ratio test using the R software.	Create
Module.3	M01. Illustrate different control charts such as Shewart control charts for variables- X , R, S charts,Attribute control charts - np, p, c and u charts using the R software. M02. Formulate estimation of parameters based on complete and censored samples in exponential, Weibull and Gamma models using the R software. M03. Perform Nonparametric estimation of failure rate and	Analyze Create Apply

	reliability function using the R software.	
Module 4	M01.Demonstrate different nonparametric tests such as Sign test, Wilcoxon signed rank test, run test, Median test. Kendall's rank order correlation coefficient and Kendall's coefficient of concordance as measure of association. Test based on Mann-Whitney Statistics. Friedman two way analysis of variance by ranks using the R software	Apply
Module 5	MO1. Formulate different experimental techniques such as LSD**,** ANCOVA, Factorial experiments. Missing and Mixed-up plot in RBD using the R software	Create
Module 6	M01. Demonstrate different tests procedures such as One way ANOVA, Two way ANOVA- multiple but equal, multiple but unequal, RBD, Kruskal-Wallis one-way analysis of variance by ranks, Kolmogorov- Smirnov one sample and two sample tests using SPSS software	Apply

COURSE CONTENT

MODULE I: (Using R) Estimation – Maximum likelihood estimation, Uniformly minimum variance unbiased estimate, method of moments, method of minimum chi-square.

MODULE II: (Using R) Testing of hypothesis – Most powerful test, Uniformly most powerful test, Uniformly most powerful unbiased test, Locally most powerful test, Likelihood ratio tests,Sequential probability ratio test.

MODULE III: (Using R) Control charts – Shewart control charts for variables- X, R, S charts,Attribute control charts - np, p, c and u charts. Reliability- series and parallel system, k-out-of-n systems, Survival function, Hazard function, Mean residual life function in the univariate cases,Exponential, Weibull, Pareto and Gamma as life distribution models, Estimation of parameters based on complete and censored samples in exponential, Weibull and Gamma models, Nonparametric estimation of failure rate and reliability function.

MODULE IV: (Using R) Nonparametric tests – Sign test, Wilcoxon signed rank test, run test.,Median test. Kendall's rank order correlation coefficient and Kendall's coefficient of concordance as measure of association. Test based on Mann-Whitney Statistics. Friedman two way analysis of variance by ranks.

MODULE V: (Using R) LSD**,** ANCOVA, Factorial experiments. Missing and Mixed-up plot in RBD.

MODULE VI: (Using SPSS) One way ANOVA, Two way ANOVA- multiple but equal, multiple but unequal, RBD, Kruskal-Wallis one-way analysis of variance by ranks, Kolmogorov-Smirnov one sample and two sample tests.

SEMESTER : IV
COURSE CODE : STAD-DE-544(i)
COURSE TITLE : OPTIMIZATION TECHNIQUES
CREDITS : 4

COURSE OUTCOMES

On completion of the course, the students should be able to:

CO1. Describe Simplex method to solve the linear programming problem.
CO2. Explain the steps in solving a linear programming problem by two-phase method.
CO3. Explain the concept of duality in linear programming problem.
CO4. Give the outline of dual simplex method.
CO5. Describe the computational procedure of optimality test in a transportation table.
CO6. Explain the Hungarian method to solve the Assignment problem.
CO7. To understand basic structure of quadratic programming problem.
CO8. Give an account of different types of inventory models and inventory cost.
CO9. Derive an EOQ formula for different rate of demand in different cycles.
CO10. Formulate and solve the purchase inventory problem with one price break.
CO11. Derive the steady state solution of M/M/1 queue model.
CO12. Obtain expected number of units in the M/G/1 queueing system under steady state .
CO13. Derive an expression of the average annual cost of an item over a period of n years.
CO14. Describe Bellmen's principle of optimality.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy Level**
Module 1.	M01. Explain the concepts of linear programming problem. M02. Solve the linear programming problem by using Simplex method. M03. Computational steps of Big-m-Method M04. Write the dual of the given linear programming problem.	Understand Apply Remember Evaluate
Module: 2	M01. Find the initial basic feasible solution to the given Transportation problem. M02. To determine the optimum assignment problem. M03. State and prove Kuhn-Tucker necessary and sufficient	Evaluate Evaluate Remember

	conditions in a non-linear programming problem. M04. Describe briefly the Beal's method for solving quadratic programming problem.	Remember
Module 3	M01. Explain with examples the probabilistic models in inventory. M02. Derive an expression for EOQ for Harri's or Wilson model. M03. Discuss the purchase inventory model with three price break.	Understand Understand Remember
Module 4	M01. State some of the important distributions of arrival interval and service time. M02. To evaluate expected value of queue length in various models. M03. Obtain the differential difference for M/M/1 queue model.	Remember Evaluate Understand
Module 5	M01. Describe the problem of replacement of items whose maintains cost increase with time. M02. Solve Dynamic programming problem by using Linear programming problem.	Understand Apply

COURSE CONTENT

MODULE I: Linear Programming: Convex sets and associated theorems, Graphical method, Computational aspects of simplex method, Duality problems of linear programming, Degeneracy and its solution, Two phase simplex method.

MODULE II: Transportation problems: Finding initial basic feasible solution, Optimality test, Degeneracy, Assignment problem, Hungarian method, Non-linear programming, Kuhn-Tucker theory for convex programming problem, Quadratic programming problem and its solution.

MODULE III: Inventory models: Deterministic models, Single item static models, models with price breaks and storage limitation, probabilistic models with single period and multi periods.

MODULE IV: Queuing theory: Basic structure, Role of the Poisson and exponential distributions. M|M|1, M|M|C, M|Ek|1 and M|G|1 queues and their properties. Waiting time distributions, Steady state solution.

MODULE V: Replacement problem: Replacement of items that deteriorate with time and money value change (i) change with time (ii) does not change with time, Individual replacement policy, Group replacement policy, Sequencing problem, Dynamic programming problem, Recursive approach.

REFERENCES

- Gass, S.I. (1969): Linear Programming Problem, Mc Graw Hill.
- Gross, D. and Harris, C.M. (1974): Fundamental of Queuing Theory, John Wiley.

• Hillier, F.S. and Leiberman, G.J. (1962): Introduction to Operations Research, Holden Day.
• Kanti Swarup, Manmohan and Gupta, M.M. (1985): Operations Research, Sultan Chand & Sons.
• Mittal, K.V. (1990): Optimization Methods.
• Ravindran, A, Philips, D.T. and Soleberg, J.J. (1997): Operations Research – Principles and Practise.
• Saaty, T.L. (1961): Elements of Queuing Theory with Applications, Mc Graw Hill.
• Taha, H.A. (1997): Operations Research, Mc.Millian.

SEMESTER **:** IV
COURSE CODE **:** STAD-DE-544(ii)
COURSE TITLE **:** STOCHASTIC FINANCE
CREDITS **:** 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Stochastic modelling is a form of financial model that is sued to help make investment decisions.

CO2. Explains the basic concepts of financial markets

CO3. In the financial service sector, planners, analysts, and portfolio managers use stochastic modelling to manage their assets and liabilities.

CO4. To understand the concept of stochastic modelling, it helps to compare it to its opposite, deterministic modelling.

CO5. Stochastic investment models attempt to forecast the variations of prices, returns on assets, and asset classes-such as bonds and stocks-over time. The monte carlo simulation is one example of a stochastic model.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. Explains the basic concepts of financial markets M02. Explain the concepts of forward contracts, future contracts, optionscall and put options, European option and American options, hedgers, speculators,	Understand Analysis

	arbitrageurs, M30. Define the interest rates, compounding, present value analysis, risk free interest rates M04. Solving problems related to interest rate, compounding, present value analysis, risk free interest rates	Evaluate Apply
Module 2	M01. Explains the concepts of Returns, gross returns and log returns. Portfolio theory- trading off expected return and risk, one risky asset and one risk free asset. Two risky assets, estimate expected return, optimal mix of portfolio CAPMM, capital market line, betas and security market line.	Understand
Module 3	M01. Explains the Binomial model- single and multiperiod binomial model, martingale measure. Modelling returns, lognormal model, random walk model, geometric Brownian motion process. Ito lemma (without proof). M02. state Arbitrage theorem and The Black-Scholes formula and also describe the Properties of the Black-Scholes option cost.	Apply Analysis
Module 4	M01. Estimating the volatility parameter. M02. Describe the pricing American options, pricing of European options using Monte Carlo and pricing an American option using finite difference methods.	Understand Evaluate
Module 5	M01. Describe the special features of financial series, linear time series models: AR(1), AR(p), ARMA(p,q) processes, and find the first and second order moments. M02. Examine the ARCH(1), ARCH(p), GARCH(p,q) models and their estimation. M03.Distinguish between ARMA and GARCH processes	Understand Apply Analysis

COURSE CONTENT

MODULE I: Basic concepts of financial markets, forward contracts, future contracts, optionscall and put options, European option and American options, hedgers, speculators, arbitrageurs, interest rates, compounding, present value analysis, risk free interest rates.

MODULE II: Returns, gross returns and log returns. Portfolio theory- trading off expected return and risk, one risky asset and one risk free asset. Two risky assets, estimate expected return, optimal mix of portfolio CAPMM, capital market line, betas and security market line.

MODULE III: Options, pricing via arbitrage, low of one price, risk neutral valuation. Binomial model- single and multiperiod binomial model, martingale measure. Modelling returns, lognormal model, random walk model, geometric Brownian motion process. Ito lemma (without proof). Arbitrage theorem. The Black-Scholes formula. Properties of the Black-Scholes option cost, the Delta Hedging arbitrage strategy, some derivatives, their interpretations and applications

MODULE IV: Volatility and estimating the volatility parameter, implied volatility, pricing American options, pricing of European options using Monte Carlo and pricing an American option using finite difference methods. Call options on dividend paying securities. Pricing American put options, modelling the prices by adding jumps to the Geometric Brownian motion. Valuing investments by expected utility. Modelling security market: self-financing portfolio and no arbitrage, price process models, division rule, product rule

MODULE V: Financial Time-Series-Special features of financial series, linear time series models: AR(1), AR(p), ARMA(p,q) processes, the first and second order moments, estimation and forecasting methods, models for conditional heteroscedasticity: ARCH(1), ARCH(p), GARCH(p,q) models and their estimation. Comparison of ARMA and GARCH processes

REFERENCES

- David Ruppert(2004). Statistics and Finance – An Introduction, Springer International edition
- Sheldon M Ross(2003). An Elementary Introduction to Mathematical Finance, Cambridge University Press
- Christian Gouriesous and Joann Jasiak(2005). Financial Econometrics, New Age Interational (P) Ltd
- John C Hull(2008). Options, Futures and other Derivatives, Pearson Education, India
- Masaaki Kijima (2003). Stochastic Process with Applications to Finance, Chapman and Hall
- Ruey S Tsay (2005). Analysis of Time Series, Third Ed., John Wiley and Sons.
- Cuthbertston K and Nitzsche D (2001). Financial Engineering- Derivatives and Risk Management , John Wiley and Sons

SEMESTER : IV
COURSE CODE : STAD-DE-544(iii)
COURSE TITLE : BAYESIAN INFERENCE
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Use relative frequencies to estimate probabilities.

CO2. Calculate conditional probabilities

CO3. Calculate posterior probabilities using Bayes' theorem.

CO4. Calculate simple likelihood functions

CO5. Describe the role of the posterior distribution, the likelihood function and the posterior distribution in Bayesian inference about a parameter.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. Explain prior distribution M02. Interpret Bayes theorem and articulate to find posterior distribution.	Understand Analysis
Module 2	M01. Find conjugate family of prior for a model M02. Choose appropriate member of conjugate prior for a family MO3. Explain non-informative, improper and invariant priors MO4. Define Jeffrey's invariant prior	Understand Apply Understand Remember
Module 3	MO1. Explain different types of loss function. MO2. Evaluate the estimate in terms of posterior risk	Apply Analysis
Module 4	MO1. Explain Bayesian interval estimation MO2. Explain highest posterior density regions MO3. Interpret confidence coefficient of an interval and its comparison with the interpretation of the confidence coefficient for a classical confidence interval	Understand Evaluate Apply
Module 5	M01. Explain testing of hypothesis in Bayesian analysis M02. Distinguish prior and posterior odds. M03. Establish Lindley's Paradox for testing a point hypothesis for normal mean against the two-sided alternative hypothesis.	Understand Apply Analysis

COURSE CONTENT

MODULE I: Subjective interpretation of probability in terms of fair odds. Evaluation of (i) subjective probability of an event using a subjectively unbiased coin (ii) subjective prior distribution of a parameter. Bayes theorem and computation of the posterior distribution.

MODULE II: Natural Conjugate family of priors for a model. Hyper parameters of a prior from conjugate family. Conjugate families for (i) exponential family models, (ii) models admitting sufficient statistics of fixed dimension. Enlarging the natural conjugate family by (i) enlarging hyper parameter space (ii) mixtures from conjugate family, choosing an appropriate member of conjugate prior family. Non-informative, improper and invariant priors. Jeffrey's invariant prior.

MODULE III: Bayesian point estimation: as a prediction problem from posterior distribution. Bayes estimators for (i) absolute error loss (ii) squared error loss (iii) 0 - 1 loss. Generalization to convex loss functions. Evaluation of the estimate in terms of the posterior risk.

MODULE IV: 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.

MODULE V: 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. Specification of the Bayes tests in the above cases. Discussion of Lindley's paradox for testing a point hypothesis for normal mean against the two-sided alternative hypothesis.

REFERENCES

- Berger, J. O. (1980): Statistical Decision Theory and Bayesian Analysis, Springer Verlag.
- Bernando, J. M. and Smith, A. F. M. (1994): Bayesian Theory, Jolm Wiley and Sons.
- DeGroot, M. H. (1970): Optimal Statistical Decisions, McGraw Hill.
- Gemerman, D. (1997): Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference, Chapman Hall.
- Leonard, T. and Hsu, J. S. J. (1999): Bayesian Methods, Cambridge University Press.
- Robert, C. P. (1994): The Bayesian Choice: A decision Theoretic Motivation, Springer.

SEMESTER: IV
COURSE CODE: STAD-DE-544(iv)
COURSE TITLE: ORDER STATISTICS
CREDITS: 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Understand the basic properties of Order statistics.

CO2. Explain Probability mass function of order statistics arising from discrete and distributions.

CO3. Examine order statics of various types of discrete and continuous distributions

CO4.Explains the properties and relations of moments of Oder statistics.

CO5. Realize the difference between discrete and continuous probability distributions.

CO6. Explain the estimation and prediction under Order statistics.

CO7. Explain the concept of order statistics and solving problems related to it

CO8. Understanding concept of Record value and Concomitant Order statistics and their applications

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. Explains the basic concepts of distribution of single order statistic, joint distribution of two or more order statistics. M02. Verify the Markov property of Order statistics. M30. Examine the various properties of Order statistics. M04. . Explain the concept of order statistics	Understand Remember Understand Understand
Module 2	M01. Explain the moments of order statistic and its properties. M02. Derive the recurrence relations on the single and product moments of order statistics. M03. Describe Order statistics from symmetric population.	Understand Remember Understand
Module 3	M01. Examine order statics of various types of discrete and continuous distributions and articulate their properties	Apply
Module 4	M01. Explain the different estimators used in order statistics M02.Examine prediction of order statistics.	Evaluate Apply

	M03. Find the confidence interval using sample quantile.	Understand
Module 5	M01. Explains the concept of Record value and its application. M02. Explains the concept of Concomitant Order statistics and application	Understand Understand

COURSE CONTENT

MODULE I: Basic distribution theory: Distribution of single order statistic, joint distribution of two or more order statistics, conditional distributions and Markov chain property. Distribution of median, range and mid-ranges, Probability mass function of order statistics arising from discrete distributions.

MODULE II: Moments of order statistics, Identities on the moments of order statistics, recurrence relations on the single and product moments of order statistics. Discussion of the above relation for symmetric population.

MODULE III: Order statistics from specific population such as Bernoulli and three point Discrete uniform distribution, exponential distribution, uniform, power function, normal and logistic distributions.

MODULE IV: Order statistics in statistical inference: Order statistics and sufficiency, linear estimations of location and scale parameters, Gupta's simplified linear estimator, prediction of order statistics, confidence intervals using sample quantile.

MODULE V: Record values: Definition and distribution theory of record values, prediction of future records and applications. Concomitants of order statistics: basic distribution theory and illustrations using bivariate samples arising from Morgenstern Family of Distributions. Applications of concomitants of order statistics.

REFERENCES

- Arnold, B. C. and Balakrishnan, N. (1989) : Relations, Bounds and Approximations for order statistics, Lecture notes in Statistics No. 53, Springer- Verlag, New York.
- Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992) : A first course in Order Statistics, John Wiley, New York.
- David, H. A. and Nagaraja, H. N. (2003): Order statistics, 3rd edition, John Wiley, New York.
- Sarhan, A. E. and Greenberg, B. G. (1962): Contributions to Order Statistics, John Wiley, New York.

SEMESTER : IV
COURSE CODE : STAD-DE-544(v)
COURSE TITLE : SPATIAL DATA ANALYTICS
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Analyze spatial temporal data.

CO2. Explain domain specific data analysis and its applications.

CO3. Handle heterogeneous data.

CO4. Distinguish between Quadrant and nearest neighbor methods.

CO5. Understand the applications of exploratory spatial data analysis (ESDA).

CO6. Interpret scatter plots.

CO7. Construct Estimation Criteria.

CO8. Explain the multivariate variant of Ordinary Kriging Operation (CoKriging).

CO9. Explain the needs of complexity measures and complexity modeling.

CO10. Find the applications of point process models.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. Explain the theoretical and practical meanings of spatial patterns and relationships. M02. Explain the use of spatial data analysis in geographic domain (Geospatial data analysis). M03.Explain the spatial aspects of computational, statistical, and visual analytical methodologies for big data analytics.	Understand Evaluate Remember
Module 2	M01. Understand various applications of domain specific data analysis. M02. Solve simple problems related to simulation. M03. Handle heterogeneous data.	Understand Apply Apply

	M04. Explain various aspects of predictive analytics and its applications. M05. Explain the needs of complexity measures and complexity modeling.	Understand Understand
Module 3	M01. Explain the concept of spatial clustering. M02. Distinguish between quadrant and nearest neighbor methods. M03. Find the difference between K-function and L-function. M04. Find the applications of point process models.	Remember Remember Analysis Analysis
Module 4	M01. Explain the applications of exploratory spatial data analysis (ESDA) M02. Understand and plot experimental variogram. M03. Interpret scatter plots. M04. Understand stationarity and isotropy with examples.	Understand Understand Apply Understand
Module 5	M01. Explain deterministic estimation. M02. Articulate Estimation Criteria. M03. Evaluate the uses of Geostatistical Estimation. M04.Understand the multivariate variant of Ordinary Kriging Operation (CoKriging).	Understand Apply Evaluate Understand

COURSE CONTENT

MODULE I: Spatial-temporal data, Theories and models of spatial patterns and relationships; Computational, statistical, and visual analytical methodologies for big data analytics, knowledge discovery, and decision support in geographic domains

MODULE II: Domain-specific data analytics and applications, Simulation, benchmark data generation, complexity modeling, predictive analytics; Big data collection, Organizing and management methodologies for heterogeneous data.

MODULE III: Locations of events versus counts of events, Spatial Clustering, Quadrant and nearest neighbor methods, K-functions and L-functions, Point process models, Estimation and Inference

MODULE IV: Exploratory Spatial Data Analysis, Spatial Continuity Analysis, Experimental Variogram, hScatterplot, Variogram versus Univariate Statistics, Stationarity and isotropy, Higher Dimensions & Statistical Anisotropy, Exploring Anisotropy, Spatial Continuity Analysis, Spatial dependence, Spatial Auto correlation

MODULE V: Deterministic Estimation - Global Estimation, Local Estimation, Estimation Criteria, Geostatistical (Probabilistic) Estimation- Ordinary Kriging, Simple Kriging, Indicator Kriging, Cokriging.

REFERENCES

- Li, Deren. Spatial Data Mining: Theory and Application. S.l.: Springer, 2016.
- Berry, Brian J. L., and Duane Francis Marble. Spatial Analysis; a Reader in Statistical Geography. Englewood Cliffs, NJ: Prentice-Hall, 1968.
- Bivand, Roger S, Edzer J Pebesma, and Virgilio Gómez-Rubio. Applied Spatial Data Analysis With R. New York: Springer, 2008.
- Lloyd, Christopher D. Spatial Data Analysis: An Introduction for GIS Users. Oxford: Oxford UP, 2010.
- Illian, Janine, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan. Statistical Analysis and Modelling of Spatial Point Patterns. Chichester, England: John Wiley, 2008.

SEMESTER : IV
COURSE CODE : STAD-DE-545(i)
COURSE TITLE : REGRESSION ANALYSIS AND ECONOMETRIC METHODS
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1: Describe simple and multiple linear regression models and its properties
CO2: Apply principle of least square method to estimate the parameters in simple and multiple linear regression models.
CO3: Identify multicollinearity problem, its consequences.
CO4: Discuss the problem of estimation of parameters when multicollinearity occurs.
CO5: Explain the Farrar –Glauber test for multicollinearity.
CO6: Describe Aitken generalized least square method of estimation.
CO7: Identify heteroscedastic disturbance.
CO8: Explain Gold field and Quandt test and Glesjer test.
CO9: Identify auto correlation and its consequences.
CO10: Explain the consequence of the presence of error in variables while estimating the parameters of a structural equation.
CO11: Discuss ILS, 2SLS, least variance ratio, full-information maximum likelihood and 3SLS method of estimation.
CO12: Perform identification problem in econometrics.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	MO1. Explain simple linear regression model MO2. Describe least square estimators MO3. Articulate to inference regarding regression parameters MO4. Explain ANOVA	Understand Remember Apply Analysis
Module 2	MO1. Explain multiple linear regression models MO2. Explain inference regarding multiple regression parameters MO3. Explain Polynomial Regression models	Analysis Understand Understand
Module 3	MO1. Establish the scope of econometrics MO2. Explain general linear regression model MO3. Explain multicollinearity MO4. State and prove Farrar-Glauber test for multicollinearity.	Apply Understand Understand Remember
Module 4	MO1. Explain Heteroscedasticity MO2. Explain test for homogeneity of variances MO3. Explain auto-correlation	Understand Apply Remember
Module 5	MO1. Explain estimation in simultaneous equation model MO2. Establish 2 SLS estimation and 3 SLS estimation MO3. Discuss identification problems	Understand Apply Understand

COURSE CONTENT

MODULE I: Simple linear regression models, Assumptions of the linear stochastic regression model, Least square estimators, Properties of the least square estimates, Inference on regression parameters, Analysis of variance, Prediction.

MODULE II: Multiple linear regression models, Estimation of the model parameters, Testing in multiple regression, Confidence intervals in multiple regression, Prediction of new observations. Polynomial Regression models.

MODULE III: Definition and scope of econometrics, Methodology of econometric analysis, General linear regression model – Linear restrictions, significance test and confidence intervals, Multicollinearity problem - meaning, consequences, detection, estimation of parameters, Farrar-Glauber test for multicollinearity.

MODULE IV: Generalized least squares (GLS) method of estimation (Aitken), Heteroscedastic disturbances, Pure and mixed estimation, Test for homogeneity of variances, Gold field and Quandt test, Glesjer test, Auto correlation-meaning, sources and consequences, Tests for autocorrelation – Durbin Watson, Von-Neumann , Errors in variables, Dummy variables, Lagged variables, Linear regression with stochastic regressors, Instrumental variable estimation.

MODULE V: Estimation in simultaneous equation model, Recursive systems, Structural and reduced forms, Indirect least squares (ILS), 2 SLS estimation , 3 SLS estimation, Identification problem, Restrictions on structural parameters-rank and order conditions, Restrictions on variances and covariances, Full-Information maximum likelihood method.

REFERENCES

- Apte, P.G. (1990): Text book of Econometrics, Tata Mc Graw Hill.
- Gujarati, D (1979): Basic Econometrics, McGraw Hill.
- Johnston, J. (1984): Econometric Models, Third edition, McGraw Hill.
- Koutsoyiannis, A (1979): Theory of Econometrics, Macmillian Press.
- Montgomery, D.C., Peck, E.A. and Vining, G.G. (2007): Introduction to Linear Regression Analysis, John Wiley, India.
- Theil, H. (1982): Introduction to the Theory and Practice of Econometrics, John Wiley.
- Wetherill, G.B. (1986): Regression Analysis with Application, Chapman Hall.

SEMESTER : IV
COURSE CODE : STAD-DE-545(ii)
COURSE TITLE : ACTUARIAL STATISTICS
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Develop a greater understanding of statistical principles and their application in actuarial statistics.

CO2. Describe the core areas of actuarial practice and relate to those areas actuarial principles, theories and models.

CO3. Describe estimation procedures for lifetime distributions.

CO4 Explain the concept of survival models.

CO5. Understand the application of knowledge of the life insurance environment.

CO6. Describe Net premiums and its various types.

CO7. Expand their applied knowledge in various specialized areas of actuarial studies and statistics.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	M01. Explains the utility theory and insurance. M02. Explain survival function and application. M30. Examine the properties of force of mortality. M04 Define Life tables and its relation with survival function, examples.	Understand Understand Apply Remember
Module 2	M01.Explain Multiple life functions and its properties. M02. Articulate the insurance and annuity benefits through multiple life functions evaluation for special mortality laws. M03. Explains the Multiple decrement tables. M04.Describe net single premiums and their numerical evaluations.	Understand Analysis Understand Remember
Module 3	M01.Define Distribution of aggregate claims . M02. Derive the compound Poisson distribution and explain its applications. M03. Explain Principles of compound interest and its attributes.	Understand Remember Understand
Module 4	M01. Explain the Life insurance and its types. M02 : Describe Insurance payable at the moment of death and at the end of the year of death-level benefit insurance M03. Explain the Life annuities and its types.	Understand Analysis Understand
Module 5	M01. Explain Net premiums and its importance M02. Distinguish between Continuous and discrete premiums M03. Accumulation type benefits.	Understand Analysis Apply

COURSE OUTCOME

MODULE I: Utility theory, insurance and utility theory, models for individual claims and theirnsums, survival function, curtate future lifetime, force of mortality. Life tables and its relation with survival function, examples, assumptions for fractional ages, some analytical laws of mortality, select and ultimate tables.

MODULE II: Multiple life functions, joint life and last survivor status, insurance and annuity benefits through multiple life functions evaluation for special mortality laws. Multiple decrement tables, central rates of multiples decrement, net single premiums and their numerical evaluations.

MODULE III: Distribution of aggregate claims, compound Poisson distribution and its applications. Principles of compound interest: Nominal and effective rates of interest and discount, force of interest and discount, compound interest, accumulation factor, continuous compounding.

MODULE IV: Life insurance: Insurance payable at the moment of death and at the end of the year of death-level benefit insurance, endowment insurance, differed insurance and varying benefit insurance, recursions, commutation functions. Life annuities: Single payment, continuous life annuities, discrete life annuities, life annuities with monthly payments, commutation functions, varying annuities, recursions, complete annuities immediate and apportion able annuities-due.

MODULE V: Net premiums: Continuous and discrete premiums, true monthly payment premiums, apportion able premiums, commutation functions, accumulation type benefits. payment premiums, apportion able premiums, commutation functions, accumulation type benefits.

REFERENCES

• Beard, R.E., Penlikainen, T. and Pesonnen, E (1984): Risk Theory: The Stochastic Basis of Insurance, 3rd Edition, Chapman and Hall, Londan.
• Bowers, N.L., Gerber, H.U., Hickman, J.E., Jones, D.A. and Nesbitt, C.J. (1997): ' Actuarial Mathematics', Society of Actuarias, Ithaca, Illiois, U.S.A., second Edition.
• Neill, A. (1977): Life Contingencies, Heineman.

SEMESTER : IV
COURSE CODE : STAD-DE-545(iii)
COURSE TITLE : BIOSTATISTICS
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Understand the principal concepts about biostatistics.

CO2. Recognize the definition and the relation with real life applications.

CO3. Interpret data via various existing distributions.

CO4. Predict statistical decision through hypothesis testing.

MODULE OUTCOMES

Sl. No:	Outcomes On completion of each module, students should be able to:	Taxonomy level
Module 1	M01. Distinguish various type of survival distributions and their survival time. M02. Familiar with different test procedures and comparison of two survival distribution.	Understand Analysis
Module 2	M01. Explain and exemplify the concept of censoring methods. M02. Evaluate the biological examples with different type of censoring.	Understand Evaluate
Module 3	M01. Articulate the concept of risk theory. M02. Explain indices for measurement of probability of death under competing risks and inter-relation. M03. Evaluate various methods of estimation under competing risk theory. M04. Distinguish independent and dependent risks.	Apply Understand Analysis Understand
Module 4	M01. Explain basic biological concepts in genetics. M02. Describe Mendel's law. M03. Explain mutation and genetic drift. M04. Discuss and estimate linkage in heredity.	Understand Evaluate Understand Evaluate
Module 5	M01. Explain different Phases of clinical trials. M02. Describe designs for comparative trials. M03. Express sample size in a fixed sample design.	Understand Understand Apply

COURSE OUTCOME

MODULE I: Functions of survival time, survival distributions and their applications viz. Exponential, Gamma, Weibull, Rayleigh, Lognormal, death density function for a distribution having bathtub shape hazard function. Tests of goodness of fit for survival distributions (WE test for exponential distribution, W -test for lognormal distribution, Chi-square test for uncensored observations). Parametric methods for comparing two survival distributions viz. L.R test, Cox's F-test.

MODULE II: Type I, Type II and progressive or random censoring with biological examples, Estimation of mean survival time and variance of the estimator for type I and type II censored data with numerical examples. Non-parametric methods for estimating survival function and variance of the estimator viz. Actuarial and Kaplan -Meier methods.

MODULE III: Competing risk theory, Indices for measurement of probability of death under competing risks and their inter-relations. Estimation of probabilities of death under competing risks by maximum likelihood and modified minimum Chi-square methods. Theory of independent and dependent risks. Bivariate normal dependent risk model. Conditional death density functions. Stochastic epidemic models: Simple and general epidemic models (by use of random variable technique).

MODULE IV: Basic biological concepts in genetics, Mendel's law, Hardy- Weinberg equilibrium, random mating, distribution of allele frequency (dominant/co-dominant cases), Approach to equilibrium for X-linked genes, natural selection, mutation, genetic drift, equilibrium when both natural selection and mutation are operative, detection and estimation of linkage in heredity.

MODULE V: Planning and design of clinical trials, Phase I, II, and III trials. Consideration in planning a clinical trial, designs for comparative trials. Sample size determination in fixed sample designs.

REFERENCES

- Biswas, S. (1995): Applied Stochastic Processes. A Biostatistical and Population Oriented Approach, Wiley Eastern Ltd.
- Cox, D.R and Oakes, D. (1984): Analysis of Survival Data, Chapman and Hall.
- Elandt, RC. and Johnson (1975): Probability Models and Statistical Methods in Genetics, John Wiley & Sons.
- Ewens, W. J. (1979): Mathematics of Population Genetics, Springer Verlag.
- Ewens, W. J. and Grant, G.R (2001): Statistical methods in Bioinformatics: An Introduction, Springer.
- Friedman, L.M., Furburg, C. and DeMets, D.L. (1998): Fundamentals of Clinical Trials, Springer Verlag.
- Gross, A. J. and Clark, V.A. (1975): Survival Distribution; Reliability Applications in Biomedical Sciences, John Wiley & Sons.

SEMESTER **:** IV
COURSE CODE **:** STAD-DE-545(iv)
COURSE TITLE **:** SURVIVAL ANALYSIS
CREDITS **:** 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Focus on applications of techniques to biology and medicine.

CO2. Understand the basic parameters used in modeling survival data and some common parametric models.

CO3. Familiar with time-to-event data. Analyze censored or truncated life time data.

CO4. Various categories of censoring and each with different likelihood function which will be the basis of inference.

CO5. Construct estimators of the survival function and the cumulative hazard rate.

CO6. Understand non parametric tests to test the better survival distribution between different survival distributions.

CO6. To model the data using lifetime distributions.

CO7. To deal with associated methods of inference.

CO8. Explain regression models used in survival analysis.

CO9. Explain parametric inference and non parametric inference for the regresson models.

MODULE OUTCOMES

Sl. No:	Outcomes On completion of each module, students should be able to:	Taxonomy level
Module 1	M01. Define Survival function, Hazard function, Mean residual life function and median life. M02. Familiar with parametric models for survival data such as exponential, weibull, gamma and log-normal distributions.	Understand Understand
Module 2	M01. Understand Time-to-Event analysis and censoring in Time-to-Event analysis. M02. Understand left and right censoring in TTE analysis M03. To estimate the survival function. M04. Define Hazard ratio. MO5. Relate Hazard ratio, Risk ratio and Odds ratio. MO6. Discuss relative survival estimation.	Understand Understand Analysis Remember Apply Apply
Module 3	M01. Analyze Non-Parametric methods for comparing survival distributions. MO2. Find confidence interval for Hazard ratio.	Analysis Apply
Module 4	M01. Describe Parametric lifetime distributions. M02. Explain linear failure rate distribution. M03. State and explain accelerated failure time model. MO4. Define Cox-Snell residuals.	Understand Evaluate Remember Understand
Module 5	M01. State and explain Cox's proportional hazard regression model. MO2. Derive rank test for the regression coefficient. MO3. Describe parametric and non- parametric inference for this model.	Remember Understand Apply

COURSE CONTENT

MODULE I: Basic quantities and models-Survival function, hazard function, mean residual life function and median life, Common parametric models for survival data, description of the above characteristic models- Weibull, Gamma, Log-Normal and Exponential.

MODULE II: Concept of time, order and random censoring,- right and left, likelihood in these cases, survival function- actuarial estimator, Kaplan Meier estimator, graphical display for survival, median survival time and confidence interval for median survival, hazard ratio, relation between hazard ratio, relative risk ratio and odds ratio, relative survival estimation.

MODULE III: Non-Parametric methods for comparing survival distributions-log rank test, confidence interval for hazard ratio, stratified log rank, Peto's test, Gehan test, Mantel Haenzel test.

MODULE IV: Lifetime distributions-parametric(exponential, gamma, weibull, logistic), linear failure rate, parametric inference(point estimation, confidence intervals, scores, likelihood ratio test), accelerated failure time model, Cox-Snell residuals.

MODULE V: Identification of prognostic factors related to survival time, Cox's proportional hazards regression model with one and several covariates, rank test for the regression coefficients, adequacy assessment of the proportional hazards model, time dependent extension of the Cox's m9odel, test with non-proportional hazards, parametric inference and nonparametric inference for this model.

REFERENCES

- Klein, J.P and Moesch Burger, M.L(2003). Survival Analysis, Springer
- Elandt, Johnson and Johnson (1998). Survival models and data analysis, John Wiley and Sons
- Miller, R.G(2000). Survivcal Analysis, Second Ed. , John Wiley and Sons
- Machin D, Cheung, Y.B and Parmer, M.K.B (2006). Survival Analysis- A Practical Approach, John Wiley and Sons
- Fisher, L.D and Belle G V(1993). Biostatistics- A Methodolgy for Health Sciences, , John Wiley and Sons
- Pressat R and Atherton A(1972). Demographic Analysis
- Preston S.H, Heuveline P and Guillot M. Demography- Measuring and Modelling Population Processes
- Lawless, J.F(2002). Statistical Models and Methods for Lifetime data, Second Ed, , John Wiley
- Kalbfleisch, J.D and Prentice, R.L(2002). Statistical Analysis of Failure Time Data, Second Ed, , John Wiley

SEMESTER **:** IV
COURSE CODE **:** STAD-DE-545(v)
COURSE TITLE **:** STATISTICAL GENETICS
CREDITS **:** 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. Stochastic modelling is a form of financial model that is sued to help make investment decisions.

CO2. Explains the basic concepts of financial markets

CO3. In the financial service sector, planners, analysts, and portfolio managers use stochastic modelling to manage their assets and liabilities.

CO4. To understand the concept of stochastic modelling, it helps to compare it to its opposite, deterministic modelling.

CO5. Stochastic investment models attempt to forecast the variations of prices, returns on assets, and asset classes-such as bonds and stocks-over time. The monte carlo simulation is one example of a stochastic model.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
Module 1	MO1. Explains Mendelian ratios and their testing MO2. Explains Hardy-Weinberg law MO3. Discuss Estimation and testing of heterosis	Understand Understand Analysis
Module 2	MO1.Describe Average effect of a gene, breeding value MO2. Explains estimation of phenotypic, genotypic and environmental correlations. Selection , individual and pedigree selection	Understand Remember
Module 3	MO1.Explains Design of plant breeding trials MO2. Discuss compact family block designs, square and rectangular lattice designs, MO3. Establish the multivariate analysis of variants, cluster analysis.	Understand Remember Apply
Module 4	M01. Explain path analysis.	Understand

	M02. Distinguish discriminant analysis for selection indices MO3. Explain partial diallel and line x tester crosses	Evaluate Understand
Module 5	M01. Explain Analysis of genotype M02. Discuss measurement of stability and adaptability- linear and non linear approaches M03. Analysis of micro array data	Understand Apply Analysis

COURSE CONTENT

MODULE I: Mendelian ratios and their testing, linkage, detection and estimation, equilibrium population and Hardy-Weinberg law, inbreeding, inbreeding coefficient and coefficient of parentage, estimation and testing of heterosis, affect of migration, mutation and selection, sex linked genes.

MODULE II: Average effect of a gene, breeding value, estimation of genetic components of variation, heritability and repeatability, estimation of phenotypic, genotypic and environmental correlations. Selection , individual and pedigree selection

MODULE III: Design of plant breeding trials-compact family block designs, square and rectangular lattice designs, multivariate analysis of variants, cluster analysis, similarity algorithms and clustering methods foe quantitative data.

MODULE IV: Path analysis, discriminant analysis for selection indices, analysis of designs for diallel, partial diallel and line x tester crosses

MODULE V: Analysis of genotype- environment interactions, ,measurement of stability and adaptability- linear and non linear approaches, combining ability and specific combining ability in diallel process, analysis of micro array data

REFERENCES

- Dabholka, A.R(1992). Elements of Biometrical Genetics. Concept Publishing Company, New Delhi
- Falconer, D. S(1985).Introduction to Quantitative Genetics, English Language Books Society, Longman
- Jain, J.P(1991). Statistical techniques in Quantitative Genetics, Hindustan Publishing Company, New Delhi
- Kempthorne, O(1957). An introduction to Genetic Statistic, The Lowa State University Press London
- Li, C.C(1982). Population Genetics, Chicago Press
- Mather, K and Jinks, J.I(1977). Introduction to biometrical genetics, Chapman and Hall
- Narien P(1990). Statistical Genetics, Wiley Eastern, New Delhi.

- Percus J.K (2001). Mathematics of Genome Analysis, Cambridge University Press
- Singh, R.K and Caudhary, B.D (1985). Biometrical Methods in Quantitative Genetic Analysis, Kalyani Publishers, Ludhiana
- Stansfield, W.D (1991). Theory and Problems of Genetics, Schaum's Outline Series
- Mather K(1938). The Measurement of Linkage in Heridity, Chemical Publishing Company, New York

SEMESTER : IV
COURSE CODE : STAD-D-546
COURSE TITLE : PROJECT/ INTERNSHIP
CREDITS : 4

COURSE OUTCOMES

On completion of the course, students should be able to:

CO1. study and examine data thoroughly and comprehensively towards attaining precision both in detail and in depth.

CO2. Conducts research and analysis crucial to substantiate the report.

CO3. Create the problems in statistical study involving planning, sampling and applications of the methods they have studied.

SEMESTER : II
COURSE CODE : STAD-GC-421
COURSE TITLE : DATA ANALYTICS USING R
CREDITS : 2

COURSE OUTCOMES

After completion of this course the students will be able to

C01.Define the basic concepts of R software and R packages
C02.Describe various concepts required for developing the R Language
C03.Build our own new functions in R
C04.Illustrate different R-Graphics facilities
C05.Perform programming of different statistical methods and procedures

MODULE OUTCOMES

Sl. No.	Outcomes On Completion of each module, Students should be able to:	Taxonomy level
Module 1.	M01.Define basic concepts of statistical software R such as Basic operations in R, Mathematical functions used in R, Assign values to variables etc. M02.Demonstrate the important data structures such as arrays, matrix, data frames, Class function etc.	Remember Apply
Module.2	M01.Design an overview of the R Language such as Expressions, Objects, Symbols, Functions, Special Values, R Syntax- Constants, Numeric vectors, Character vectors, Symbols, Operators, Order of operations, Assignments, Expressions. M02. Demonstrate the use of control Structures- Conditional Statements, Loops, Accessing data Structures,	Create Apply
Module.3	M01.Describe the use of Functions- The function keyword, Arguments, Return values, Functions as arguments, properties of functions M02. Demonstrate the use of writing functions in R. M03.Perform working with data – Entering data within R, Importing data from external files, Exporting data	**Understand** Apply Apply
Module 4	M01. Demonstrate the use of R-Graphics: An overview of R graphics, Scatterplots, Bar charts, Histogram, Pie charts. M02. Illustrate the use of Plotting distributions, Plotting time series, Box plots, Stem and lead plot,Q-Q plots. M03. Summarize the use of Graphical parameters, Basic graphic functions.	Apply Apply Understand
Module 5	M01. Develop programming of basic statistical methods and procedures such as Descriptive Statistics, Graphical representation of data, Measures of Central tendency, Measures of dispersion, Measures of skewenss and kurtosis M02. Analyze statistical inference using R- Plots to check normality, Hypothesis testing – Parametric and non-parametric tests concerning means, Testing proportions of one sample and two samples. One way ANOVA and its nonparametric version. Goodness of fit tests, M03. Illustrate the concept of simple correlation and regression.	Create Evaluate Analyze

COURSE CONTENT

MODULE I: Introduction to the statistical software R, Basic operations in R, Mathematical functions used in R, Introduction to data structures - arrays, matrix and data frames.

MODULE II: An overview of the R Language- Expressions, Objects, Functions, Control Structures- Conditional Statements, Loops, Accessing data Structures.

MODULE III: Functions- The function keyword, Writing functions in R. Working with data – Entering data within R, Importing data from external files, Exporting data.

MODULE IV: Graphics- An overview of R graphics, Basic graphic functions, Scatterplots, Bar charts, Histogram, Pie charts, Plotting distributions, Box plots, Stem and lead plot, Q-Q plots, Graphical parameters.

MODULE V: Descriptive Statistics, Graphical representation of data, Measures of Central tendency, Measures of dispersion, Measures of skewenss and kurtosis. Statistical inference using R- Plots to check normality, Hypothesis testing – Parametric and non-parametric tests concerning means, Testing proportions of one sample and two samples, One way ANOVA and its nonparametric version. Goodness of fit tests, Simple correlation and regression.

REFERENCES

- Everitt, B.S. and Hothorn T. (2010). A Handbook of Statistical Analysis Using R, Second Edition, CRC Press.
- Joseph Adler (2011) R in a Nutshell, Second Edition, Shroff Publishers and Distributors Pvt. Ltd.
- Michael J. Crawley (2013) The R book, Second Edition, John Wiley & Sons Ltd.