DEPARTMENT OF STATISTICS

UNIVERSITY OF KERALA

Outcomes Based Syllabus (OBS)

of

MASTER OF SCIENCE IN APPLIED STATISTICS AND DATA ANALYTICS

(Effective from Academic Year 2024-25)

UNIVERSITY OF KERALA

कर्मणि व्यज्यते प्रज्ञा

NAAC A++ Grade

UNIVERSITY OF KERALA

One of the first **16** Universities in India, the University of Kerala was founded as the University of Travancore in the erstwhile princely state of Travancore by his highness Sri. Chithira Thirunal Maharaja in **1937**. At present, the University has **16** faculties and **44** departments of teaching and research in addition to study centers and other departments. The Manuscript Library of the University is a treasure trove of Indian culture. The library has over 65,000 works in 30,000 copies mainly of palm leaf manuscripts, in addition to paper manuscripts, copper plates, birch bark and textiles, which includes those belonging to Burma, Malaysia, Indonesia, Nepal etc. The Adult Education Department of the University has received UNESCO-NLM award in 2005. The University Library established in 1942 has over 3,00,000 books and gives access to over 9000 on-line journals. The University has published an English translation of the classic Dutch work, “Hortus Malabaricus” a 20-volume treatise on plant wealth of Kerala, in 2003.

This department is the **second oldest post-graduate department of Statistics in India**, established in **1945** based on the recommendation of a committee consisting of eminent members including Prof. P. C. Mahalanobis. The post-graduate Department of Statistics, University of Kerala has played a crucial role in the dissemination of advanced knowledge in Statistics and in the propagation of the studies in Statistics. Apart from imparting training in advanced topics of Statistics to its students, the department has been also catering the Statistical requirements of the Govt. of Kerala. Dr. U. S. Nair was the Founder Professor of this department followed by eminent successors Dr. Aleyamma George, Dr. R. Krishna Pillai, Dr. R. Ramakumar, Prof. P. S. J. Nair, Dr. Jacob Sundara Raja, Dr. R. N. Pillai, Dr. T. S. K. Moothathu, Dr. P. Yageen Thomas, Dr. G. Lesitha, Dr. C. Satheesh Kumar, Dr. E. I. Abdul Sathar and Dr. Manoj Chacko. The Department of Science and Technology (DST), New Delhi has identified this department under its FIST Programme. This department has moulded several reputed statisticians like Prof. K. C. S. Pillai, Prof. A. M. Mathai, Prof. Thomas Mathew and so on. Research papers are published regularly from this department in refereed journals in Statistics and have been extensively cited in the latest Statistical publications.

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

On completion of the programme, students should be able to:	
PSO1	**Apply statistical theory and its applications in various fields**
PSO2	**Describe recent developments in statistical theory and practice**
PSO3	**Apply statistical theory and methodologies to analyze and interpret data encountered in day-to-day life situations**
PSO4	**Demonstrate proficiency in collecting, organizing, and analyzing data using statistical methods and software tools**
PSO5	**Develop skills in building statistical models to describe, analyze, and interpret real-world phenomena**
PSO6	**Utilize appropriate statistical software for data analysis and interpretation**

Structure of the M.Sc. Applied Statistics and Data Analytics Programme

Sem. No.	Course code	Name of the Course	No. of Credits	Marks	
				CA	ESA
I	STAD-CC-511	Statistical Mathematics	4	40	60
I	STAD-CC-512	Distribution Theory	4	40	60
I	STAD-CC-513	Sampling Theory	4	40	60
I	STAD-CC-514	Big Data Analytics and Artificial Intelligence	4	40	60
I	STAD-CC-515	Statistical Computing for Data Analytics using R and Python	4	40	60
II	STAD-CC-521	Design and Analysis of Experiments	4	40	60
II	STAD-CC-522	Multivariate Analysis	4	40	60
II	STAD-CC-523	Probability Theory	4	40	60
II	STAD-CC-524	Theory of Estimation	4	40	60
III	STAD-CC-531	Stochastic Processes	4	40	60
III	STAD-CC-532	Time Series Analysis	4	40	60
III	STAD-CC-533	Testing of Hypotheses	4	40	60
III	STAD-DE-534(i)	Machine Learning	4	40	60
III	STAD-DE-534(ii)	Survival Analysis	4	40	60
III	STAD-DE-534(iii)	Statistical Genetics	4	40	60
IV	STAD-CC-541	Practical for Data Analytics	4	40	60
IV	STAD-DE-542(i)	Statistical Quality Control and Reliability Modeling	4	40	60
IV	STAD-DE-542(ii)	Official Statistics	4	40	60
IV	STAD-DE-542(iii)	Categorical Data Analysis	4	40	60
IV	STAD-DE-542(iv)	Longitudinal Data Analysis	4	40	60
IV	STAD-DE-543(i)	Optimization Techniques	4	40	60
IV	STAD-DE-543(ii)	Stochastic Finance	4	40	60
IV	STAD-DE-543(iii)	Spatial Data Analytics	4	40	60
IV	STAD-CC-544	Dissertation/ Project & Viva	8		100

STAD - Statistics and Data Analytics

CC - Core Course

DE - Discipline Specific Elective

CA Marks (40) distributed as follows

Mid Semester Examination	20 Marks
Assignment	10 Marks
Seminar	10 Marks

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

COURSE OUTCOMES

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

CO1	Have a clear understanding of Metric space - Metric in Rn
CO2	Explain the concept of Cauchy sequence, completeness, compactness and connectedness with the help of examples
CO3	To have a clear knowledge on differentiability of a function
CO4	State and prove generalized mean value theorem
CO5	Describe the properties of Riemann- integral
CO6	State and prove mean value theorem of integral calculus
CO7	Identify the applications of multiple integrals
CO8	To achieve ideas on vector space, subspaces, independence of vectors, basis and dimension
CO9	State and prove Cayley-Hamilton theorem
CO10	Establish the relation between algebraic and geometric multiplicity
CO11	To achieve ideas on quadratic forms and reduction of quadratic forms and gets an ability for solving problems in these areas
CO12	Define Moore-Penrose g-inverse and derive its properties

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. To be familiar with the idea of open and closed set	Understand
	Find the limit point of a sequence	Evaluate
	State and prove Hein-Boral theorem	Remember
Module 2	MO1. Find the limit and continuity of a given function	Apply
	MO2. Find the conditional maximum and minimum of a given function	Apply
	Verify generalized mean value theorem to a given function	Analyse
Module 3	MO1. Write the Riemann integral of the given function	Apply
	MO2. State and prove Fubini's theorem	Remember
Module 4	MO1. Find the rank of a given matrix	Evaluate
	MO2. Understand the concept of determinants and its properties	Understand
	MO3. Solve the system of linear equations	Apply
Module 5	MO1. Determine the Eigen values and Eigen vectors of the given matrix	Evaluate
	MO2. Obtain the diagonal form and triangular form of a given matrix	Apply
	MO3. Write down the spectral decomposition of the given matrix	Evaluate
Module 6	MO1. Find the nature of the quadratic form	Analyse
	MO2. Articulate the concept of generalized inverse	Understand

	MO3. Obtain the g-inverse and Moore- Penrose g-inverse of the given matrix	Apply

COURSE CONTENT

MODULE I: 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 II: 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 III: 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 IV: 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 V: 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.

MODULE VI: 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

1. Apostol, T.M. (1987): Mathematical Analysis, 2nd edn, Narosa Publishing House, New Delhi.
2. Goldberg, R.R. (1970): Methods of Real Analysis, Oxford & IBH publishing Co. (P) Ltd. New Delhi.
3. Graybill, A. and Belmont, C.A. (1983): Matrices with Applications in Statistics, II Edition, John Wiley, New York.
4. Pringle, R.M. and Rayner, A. A. (1971): Generalised Inverse of Matrices with Application to Statistics, Griffin, London.
5. 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

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Explains the basic concepts of pgf, mgf and characteristic functions	Understand
	MO2. Examine various types of discrete and continuous distributions and articulate their properties	Analyse
	M30. Examine the various properties of probability distribution	Evaluate
	MO4. Solving problems related to several distributions	Apply
Module 2	MO1. Examine various types of discrete and continuous distributions and articulate their properties	Analyse
	MO2. Examine the various properties of probability distribution	Evaluate
	MO3. Solving problems related to several distributions	Apply
Module 3	MO1. Finding the joint, marginal and conditional pdf of a bivariate distribution	Understand
	MO2. Articulate the various concepts of Conditional expectation and conditional variance	Analyse
	MO3. Explains the concept of pgf, mgf and independence of a bivariate random vector	Remember
	MO4. Articulate Multinomial and bivariate normal distributions and their properties in detail	Apply

Module 4	MO1. Apply transformation of variable technique for finding the distribution of functions of random variables	Apply
	MO2. Derive the distributions of sum, product and ratios of random variables	Analyse
Module 5	MO1. Explain the concept of order statistics	Understand
	MO2. Find the joint distribution of two order statistics	Evaluate
	MO3. Calculate Distribution of functions of two order statistics	Apply
Module 6	MO1. Explains different sampling distributions	Understand
	MO2. Find expressions for probability density function of sampling distribution	Apply
	MO3. Distinguish between partial and multiple correlation coefficient	Apply
	MO4. Finding the Distribution of empirical distribution function	Evaluate

COURSE CONTENT

MODULE I: Basic concepts in distribution theory: p.g.f., m.g.f., and characteristic function. Univariate discrete distributions: Binomial, Negative Binomial, Poisson, Hyper geometric, Geometric.

MODULE II: Univariate continuous distributions: Beta type I, Beta type II, Gamma, Normal, Log-Normal, Pareto, Weibull, Cauchy. Laplace, Logistic, Log-logistic. Rayleigh and generalized exponential distributions.

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

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

MODULE VI: Sampling distributions: Chi-square, t distribution, and F distributions (both central and non-central) and their applications, Empirical distribution; Distribution of empirical distribution function. Correlation and regression; simple partial and multiple correlation coefficients.

REFERENCES

1. Anderson, T.W. (1984) Introduction to Multivariate Statistical Analysis, Macmillan Publishing Company.

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

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

COURSE OUTCOMES

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

CO1	Distinguish between probability sampling and non-probability sampling.
CO2	Understand SRSWR and SRSWOR, estimate the population mean, population total, population proportion and variance.
CO3	Understand the concept of sampling with varying probabilities like Probability proportional to size (PPS) sampling and the procedure of selecting a PPS sample with and without replacement.
CO4	Explain Midzuno-Sen- Lahiri, Murthy's, DesRaj's sampling strategies under varying probability without replacement.
CO5	Familiar with the concept of stratified random sampling and various types of allocations.
CO6	Comparison of systematic sample, SRS and stratified random sample for a population with linear trend.
CO7	Explain the concept of Hartley-Ross unbiased ratio type estimator
CO8	Describe the regression estimator and its properties.
CO9	Understand the cluster sampling with equal and unequal clusters.
CO10	Find the relative efficiency of cluster as unit of sampling compared to a single element as unit of sampling.

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Use various procedures of selecting samples from the population	Apply
	MO2. Explain method of selecting samples using simple random sampling	Evaluate
	MO3. Articulate the estimation of population mean, total and proportion and hence the estimation of the standard error.	Apply
	MO4. Explain determination of sample size and confidence interval	Evaluate
Module 2	MO1. Explain PPS sampling and articulate to selection procedures using PPS with and without replacement	Evaluate
	MO2. Find methods to estimate population mean, total and variance with respect to PPS sampling	Apply
	MO3. Apply various selection procedures for selecting samples using PPS	Apply
Module 3	MO1. Explain the purpose of stratification	Evaluate
	MO2. Explain stratified random sampling	Understand
	MO3. Examine the various properties of stratified sampling	Evaluate
	MO4. Explain the concept of double sampling	Understand

	MO5. Discuss various allocation procedures	Analyse
Module 4	MO1. Explain systematic sampling	Understand
	MO2. Distinguish between single start linear systematic sampling and multiple start linear systematic sampling	Evaluate
	MO3. Describe the comparison of simple random sample, systematic sample and stratified sample for a population with linear trend	Understand
Module 5	MO1. Distinguish between ratio and regression estimators	Understand
	MO2. Explain various properties of ratio and regression estimators	Apply
	MO3. Discuss ratio and regression estimator for stratified population	Analyse
Module 6	MO1. Explain cluster sampling	Understand
	MO2. Distinguish between clusters of equal and unequal sizes	Remember
	MO3. Describe the two-stage cluster sampling of equal size and unequal size	Evaluate

COURSE CONTENT

MODULE I: Census, Sampling, Probability sampling and non-probability sampling. Simple random sampling with replacement and without replacement. Procedures of selection, Estimation of population mean, population total, population 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 Neyman method of allocation, Estimation with stratified 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

1. Cochran, W.G. (1977). Sampling Techniques, Wiley Eastern Ltd.
2. Des Raj (1976). Sampling Theory. McGraw Hill.
3. Murthy, M. N. (1967). Sampling Theory and Methods, Statistical Publishing Society, Calcutta.
4. Singh, D. and Chaudhary, F.S. (1986). Theory and Analysis of Sample Survey Designs, Wiley Eastern Ltd.
5. 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 : **BIG DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE**
CREDITS : **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
CO6	Can implement Distributed computing to process large data

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Introduction to Big Data	Understand
	MO2. Understanding the properties of Big Data	Analyse
	MO3. Explanation of HDFS and Map Reduce	Remember
	MO4. Setup a Hadoop Cluster	Apply
Module 2	MO1. Deep understanding of HDFS architecture	Remember
	MO2. Understanding Hadoop YARN	Understand
	MO3. Describe Hadoop Ecosystems and components	Analyse
Module 3	MO1. Understand distributed computation with map reduce	Understand
	MO2. Explore mapreduce configuration	Analyse
	MO3. Store in nosql database using Hbase	Understand
	MO4. Implement MapReduce	Apply
Module 4	MO1. Understanding the architecture of YARN	Understand
Module 5	MO1. Building machine learning models using mahout	Apply
	MO2. Understand how Machine learning is applied on large scale data	Understand
	MO3. What is NoSQL and when to use NoSQL	Apply
Module 6	MO1. 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

1. Berson, A. and Smith, S.J. (1997): Data Warehousing, Data Mining, and OLAP. McGraw-Hill.
2. Breiman, L. Friedman, J.H. Olshen, R.A. and Stone, C.J. (1984): Classification and Regression Trees. Wadsworth and Brooks/Cole.
3. Han, J. and Kamber, M. (2000): Data Mining; Concepts and Techniques. Morgan Gaufmann.
4. Mitchell, T.M. (1997): Machine Learning. McGraw-Hill.
5. Ripley, B.D. (1996): Pattern Recognition and Neural Networks. Cambridge University Press.

SEMESTER : **I**
COURSE CODE : **STAD-CC-515**
COURSE TITLE : **STATISTICAL COMPUTING FOR DATA ANALYTICS USING R AND PYTHON**
CREDITS : **4**

COURSE OUTCOMES

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

CO1	Utilize R and Python programming languages for statistical computing and data analysis
CO2	Apply fundamental programming concepts and syntax specific to R and Python environments
CO3	Perform data manipulation and analysis tasks, including subsetting, filtering, sorting, merging, reshaping, and summarizing data using both R and Python
CO4	Visualize data effectively using base graphics, ggplot2, Matplotlib, and Seaborn in R and Python
CO5	Develop and debug user-defined functions and control structures in R and Python
CO6	Implement advanced programming techniques such as object-oriented programming (OOP) and functional programming in both R and Python
CO7	Conduct basic statistical analyses including hypothesis testing, regression analysis, non-parametric tests, survival analysis, and time series analysis using appropriate libraries in R and Python

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Understand the basics of R programming language and its applications in data analysis	Remember
	MO2. Navigate the R environment and execute basic R syntax	Apply
	MO3. Manipulate R objects and data types, including vectors, matrices, arrays, data frames, and lists	Apply
	MO4. Perform data input/output operations, reading and writing data from/to different file formats	Apply
	MO5. Implement basic data manipulation techniques such as subsetting, filtering, sorting, merging, reshaping, and summarizing data	Apply
	MO6. Develop user-defined functions and control structures in R	Create
Module 2	MO1. Work with factors, dates, and times in R	Understand
	MO2. Create data visualizations using base R graphics and ggplot2	Apply
	MO3. Handle large datasets efficiently using data.table and dplyr packages	Apply
	MO4. Debug and profile R code for optimization	Analyze
	MO5. Understand and implement object-oriented programming principles in R	Understand

	MO6. Apply functional programming techniques using apply functions, lapply, sapply, and mapply	Apply
Module 3	MO1. Perform basic statistical analysis using built-in functions and libraries in R	Apply
	MO2. Conduct hypothesis testing and construct confidence intervals	Apply
	MO3. Implement regression analysis techniques including linear regression, logistic regression, and Poisson regression	Apply
	MO4. Conduct non-parametric tests and analysis of variance	Apply
	MO5. Perform survival analysis using Kaplan-Meier survival curves and Cox proportional hazards regression	Apply
	MO6. Apply time series analysis techniques such as ARIMA models and seasonal decomposition	Apply
Module 4	MO1. Understand the basics of Python programming language and its applications in data analysis	Remember
	MO2. Navigate the Python environment and execute basic Python syntax	Apply
	MO3. Manipulate Python objects and data types, including numbers, strings, lists, tuples, sets, and dictionaries	Apply
	MO4. Implement control structures such as if-else statements and loops	Apply
	MO5. Develop user-defined functions in Python	Create
	MO6. Perform input/output operations, reading and writing data from/to different file formats	Apply
Module 5	MO1. Work with Numpy arrays and Pandas data frames for data manipulation and analysis	Apply
	MO2. Visualize data using Matplotlib and Seaborn libraries in Python	Apply
	MO3. Debug and profile Python code for optimization	Analyze
	MO4. Understand and implement object-oriented programming principles in Python	Understand
	MO5. Introduce functional programming concepts in Python	Understand
Module 6	MO1. Conduct basic statistical analysis using built-in functions and libraries in Python	Apply
	MO2. Perform hypothesis testing and construct confidence intervals in Python	Apply
	MO3. Implement regression analysis techniques including linear regression, logistic regression, and Poisson regression in Python	Apply
	MO4. Conduct non-parametric tests and analysis of variance in Python	Apply
	MO5. Perform survival analysis using Kaplan-Meier survival curves and Cox proportional hazards regression in Python	Apply
	MO6. Apply time series analysis techniques such as ARIMA models and seasonal decomposition in Python	Apply

COURSE CONTENT

MODULE I: Introduction to R Programming: Overview of R and its uses in data analysis, Introduction to the R environment and basic R syntax, Working with R objects and data types (vectors, matrices, arrays, data frames, lists), Data input/output in R (reading/writing data from/to different file formats), Basic data manipulation and analysis in R (subsetting, filtering, sorting, merging, reshaping, summarizing, and visualizing data), Writing user-defined functions and control structures in R.

MODULE II: Advanced R Programming: Working with factors, dates, and times in R, Data visualization in R (using base R graphics and ggplot2), Working with large datasets in R (using data.table and dplyr) Debugging and profiling R code, Object-oriented programming in R (S3 and S4 classes). Introduction to functional programming in R (using apply functions, lapply, sapply, and mapply).

MODULE III: Statistical Analysis with R: Basic statistical analysis in R (using built-in functions and libraries such as stats and MASS), Hypothesis testing and confidence intervals in R, Regression analysis in R (linear regression, logistic regression, Poisson regression), Non-parametric tests and analysis of variance (ANOVA) in R, Survival analysis in R (Kaplan-Meier survival curves, Cox proportional hazards regression), Time series analysis in R (ARIMA models, seasonal decomposition of time series.

MODULE IV: Introduction to Python Programming: Overview of Python and its uses in data analysis, Introduction to the Python environment and basic Python syntax, Working with Python objects and data types (numbers, strings, lists, tuples, sets, dictionaries), Control structures in Python (if-else statements, loops), Writing user-defined functions in Python, Input/output in Python (reading/writing data from/to different file formats).

MODULE V: Advanced Python Programming: Working with Numpy arrays and Pandas data frames, Data manipulation and analysis in Python (subsetting, filtering, sorting, merging, reshaping, summarizing, and visualizing data), Data visualization in Python (using Matplotlib and Seaborn), Debugging and profiling Python code, Object-oriented programming in Python, Introduction to functional programming in Python.

MODULE VI: Statistical Analysis with Python: Basic statistical analysis in Python (using built-in functions and libraries such as Scipy and Statsmodels), Hypothesis testing and confidence intervals in Python, Regression analysis in Python (linear regression, logistic regression, Poisson regression), Non-parametric tests and analysis of variance (ANOVA) in Python, Survival analysis in Python (Kaplan-Meier survival curves, Cox proportional hazards regression), Time series analysis in Python (ARIMA models, seasonal decomposition of time series).

REFERENCES

1. Hadley Wickham and Garrett Grolemund (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, O'Reilly Media, Inc.

2. Bradley C. Boehmke (2016). *Data Wrangling with R*, Springer.
3. Phil Spector (2008). *Data Manipulation with R*, Springer.
4. Hadley Wickham (2009). *ggplot2: Elegant Graphics for Data Analysis*, Springer.
5. Carson Sievert (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*, Taylor and Francis.
6. Winston Chang (2018). *R Graphics Cookbook: Practical Recipes for Visualizing Data,* Second Edition, O'Reilly Media, Inc.
7. Yihui Xie, J. J. Allaire, Garrett Grolemund (2019). *R Markdown: The Definitive Guide*, CRC Press.
8. Hadley Wickham (2021). *Mastering Shiny: Build Interactive Apps, Reports, and Dashboards Powered by R*, O'Reilly Media, Inc.
9. Hadley Wickham (2015). *R Packages: Organize, Test, Document, and Share Your Code*, O'Reilly Media, Inc.
10. Hadley Wickham (2014). *Advanced R*, CRC Press.
11. Eric Matthes (2019). *Python Crash Course: A Hands-On, Project-Based Introduction to Programming*, 2nd edition, No Starch Press.
12. Al Sweigart (2019). *Automate the Boring Stuff with Python: Practical Programming for Total Beginners*, 2nd edition, No Starch Press.
13. Wes Mckinney (2017). *Python for Data: Data Wrangling with Pandas, NumPy, and IPython*, 2nd edition, O'Reilly.
14. Jake VanderPlas (2016). *Python Data Science Handbook: Essential Tools for Working with Data*, O'Reilly.
15. John Paul Mueller and Luca Massaron (2019). *Python for Data Science for Dummies*, 2nd edition, Wiley.
16. Joel Grus (2019). *Data Science from Scratch: First Principles with Python*, 2nd edition, O′Reilly.
17. Fabio Nelli (2018). *Python Data Analytics: With Pandas, NumPy, and Matplotlib*, 2nd edition, Apress.
18. Andreas Muller (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists*, 1st edition, O'Reilly.
19. Max Kuhn and Kjell Johnson (2018). *Applied Predictive Modeling*, 2nd edition, Springer.

SEMESTER : II
COURSE CODE : STAD-CC-521
COURSE TITLE : DESIGN AND ANALYSIS OF EXPERIMENTS
CREDITS : 4

COURSE OUTCOMES

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

CO1	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 OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1: Identify estimability of a linear parametric function	Understand
	MO2: Apply Gauss-Markov theorem for finding BLUE of a parametric function	Apply
	MO3: Formulate the model for one way and two-way classification	Understand
	MO4: Perform one way and two-way analysis of varaince	Analyze
Module 2	MO1: Identify connected, balance and orthogonal designs	Understand
	MO2: Design and analysis of CRD, RBD, LSD and GLSD	Analyze
Module 3	MO1: Design and analyse of 22, 3n and pn factorial experiments	Analyze
	MO2: Apply principle of total and partial confounding in factorial experiments	Apply
Module 4	MO1: Analyze BIBD with inter and intra block informations	Apply
	MO2: Analyse PBIBD with only two associates classes	Apply
Module 5	MO1: Analyses Split-plot and split-split plot designs	Apply
	MO2: Formulate the model for one way and two-way analysis of covariance	Apply
	MO3: Perform one way and two-way analysis of covariance	Apply
Module 6	MO1: Perform Missing plot analysis in RBD and LSD	Analyze
	MO3: Perform mixed-up plot analysis in RBD and LSD	Analyze
	MO2: Estimate missing yields in split plot and BIBD	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 2n, 3n and pn 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

1. Aloke Day (1986). Theory of Block Designs, Wiley Eastern, New Delhi.
2. Chakrabarti, M.C. (1962). Mathematics of Design and Analysis of Experiments, Asia Publishing House, Bombay.
3. Das, M. N. and Giri, N. (1979). Design and Analysis of Experiments. Wiley Eastern Limited, New Delhi.
4. John, P.W.M. (1971). Statistical Design and Analysis of Experiments, Macmillan.
5. Joshi, D. D. (1987). Linear Estimation and Design of Experiments, Wiley Eastern, Wiley Eastern Limited, New Delhi.
6. Montgomery, C.D. (1976). Design and Analysis of Experiments, John Wiley, New York.

SEMESTER : II
COURSE CODE : STAD-CC-522
COURSE TITLE : MULTIVARIATE ANALYSIS
CREDITS : 4

COURSE OUTCOMES

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 OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1: Define the multivariate normal density function.	Understand
	MO2: Obtain the characteristic function of multivariate normal density	Understand
	MO3: Find the distribution of linear combination of multivariate normal random vector using characteristic function	Apply
	MO4: Find regression of multivariate normal distribution using conditional distribution	Apply
Module 2	MO1: Characterize quadratic forms of multivariate distribution	Apply
	MO2: Obtain the distribution of sums and quotients of quadratic forms	Apply
	MO3: Apply Cocharan's theorem to find distribution of quadratic forms of multivariate normal random vector	Apply
Module 3	MO1: Obtain the MLEs of mean and variance of multivariate normal distribution	Evaluate
	MO2: Find the characteristic function of Wishart distribution	Apply
	MO3: Show that Wishart distribution possess additive property	Understand
	MO4: Find the distribution of sample dispersion matrix	Apply
Module 4	MO1: Find the distribution of sample multiple correlation for multivariate normal distribution	Understand
	MO2: Find the distribution of sample partial correlation coefficients for multivariate normal distribution	Apply
	MO3: Obtain the distribution of test statistic for the test the significance of correlation coefficient and partial correlation coefficient	Apply

Module 5	MO1: Test the mean vector of a multivariate normal distribution	Apply
	MO2: Test the equality of means of two or more multivariate normal distributions	Apply
	MO3: Use Hotelling's T^2 and Mahalanobis D^2 statistics in testing hypothesis regarding multivariate normal distributions	Apply
	MO4: Find the relationship between Hotelling's T^2 and Mahalanobis D^2 statistics	Understand
Module 6	MO1: Perform principal component analysis and factor analysis	Analyze
	MO2: Classify individuals/items in to one of k multivariate normal populations	Analyze
	MO3: Identify canonical variables and quantify canonical correlation	Analyze

COURSE CONTENT

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

1. Anderson, T.W. (2003): An Introduction to Multivariate Statistical Analysis, John Wiley, New York.
2. Graybill, F. A. (1961): An Introduction to Linear Statistical Model, Vol. 1, McGraw Hill, New York.
3. Johnson, R and Wychern (1992): Applied Multivariate Statistical Analysis, Prentice Hall, London.

4. Kendall, M. G. (1958): A Course in Multivariate Analysis, Griffin, London.
5. Khatri, C.G. and Srivastava (1979): An Introduction to Multivariate Statistics, North – Holland, New York.
6. Muirhead, R.J. (1982): Aspects of Multivariate Statistical Theory, John Wiley & Sons New York.
7. Rao, C.R. (1973): Linear Statistical Inference and its Applications, Wiley Eastern, New York.
8. Rohatgi, V.K. (1976): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd, New York.

SEMESTER : II
COURSE CODE : STAD-CC-523
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
CO10	Apply the various laws of large numbers to sequences of random variables.

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Explain and exemplify the concepts of probability space	Understand
	MO2. Examine the various properties of a probability measure.	Analyse
	MO3. State and explain various results associated with probability measure	Understand
	MO4. Explain the concepts of independence of events	Remember
	MO5. Construct examples of independent events	Create
	MO6. Construct counter examples for proving/illustrating certain results associated with probability measure	Create
	MO7. Evaluate the conditional probability and verify its properties	Evaluate
	MO8. Articulate the Bayes theorem and apply it to calculate apriori probabilities	Apply
Module 2	MO1. Distinguish between various types of random variables and articulate their properties	Apply
	MO2. Describe the various functions of random variables and find the functions for various random variables	Apply
	MO3. Explain and exemplify the concepts of decomposition of distribution function of a random variable	Understand
	MO4. Articulate/exemplify various types of distributions and their important properties	Understand
Module 3	MO1. Calculate the mathematical expectation, moments and generating functions of random variables	Evaluate

	MO2. Explain the concepts of random vectors and distribution function of random vectors, and their important properties	Understand
	MO3. Articulate the various moment inequalities	Understand
	MO4. Apply the various moment inequalities to various distributions	Apply
Module 4	MO1. Articulate and appraise stochastic convergence of sequence of random variables	Understand
	MO2. Apply the concepts of convergence to sequences of random variables	Evaluate
	MO3. Construct counter examples for not satisfying certain convergence implications	Create
Module 5	MO1. Derive expressions for the characteristic function for various distributions	Evaluate
	MO2. Find moments using characteristic function	Evaluate
	MO3. Derive expressions for the probability density function corresponding to a given characteristic function	Evaluate
	MO4. Articulate the various theorems associated with the characteristic function and identify their applications	Apply
	MO5. Explain and exemplify the concepts of infinitely divisible distributions	Understand
Module 6	MO1. State and prove the various laws of large numbers	Understand
	MO2. Apply the laws to sequences of random variables	Apply
	MO3. Articulate the concepts of stable distributions	Understand
	MO4. Explain the concepts of Stable distributions and distribution of class L	Understand

COURSE CONTENT

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

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

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

COURSE OUTCOMES

After completion of this course the students will be able to

C01	Understand fundamental concepts of statistical estimation and inference
C02	Apply various estimation methods to real-world problems
C03	Analyze the properties and performance of different estimators
C04	Evaluate the efficiency and consistency of estimation procedures
C05	Demonstrate proficiency in Bayesian inference techniques

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy level
Module 1	MO1. Define point estimation and sufficiency	Remember
	MO2. Apply Neyman-Pearson factorization theorem	Understand
	MO3. Identify exponential and Pitman families of distributions	Remember
	MO4. Understand Likelihood equivalence and unbiased estimation	Understand
Module 2	MO1. Explain UMVUE estimators and their characterizations	Understand
	MO2. Apply Rao-Blackwell and Lehmann-Scheffe theorems	Understand
	MO3. Find UMVUE using different methods	Understand
	MO4. Estimate parametric functions from standard distributions using UMVUE	Analyze
Module 3	MO1. Define Fisher information measure and its properties	Remember
	MO2. Apply Cramer-Rao inequality and Chapman-Robbins bound	Understand
	MO3. Understand Bhattacharya bounds, efficiency, and consistency	Understand
	MO4. Determine lower bounds on variance of unbiased estimates	Analyze
Module 4	MO1. Apply method of moments and maximum likelihood estimators	Understand
	MO2. Analyze properties of maximum likelihood estimators	Analyze
	MO3. Understand minimum chi-square and least square estimation	Understand
	MO4. Apply estimation methods to real-world scenarios	Apply
Module 5	MO1. Identify location and scale families of distributions	Remember
	MO2. Apply location and scale invariant estimators	Understand
	MO3. Understand Pitman estimators for location and scale parameters	Understand
	MO4. Implement BLUE for location and scale distributions	Apply

Module 6	MO1. Describe basic elements of Bayesian inference	Remember
	MO2. Analyze loss functions and Bayes risk	Understand
	MO3. Understand prior and posterior distributions	Understand
	MO4. Apply Bayes theorem for parameter estimation	Apply

COURSE CONTENT

MODULE I: Point Estimation and Sufficiency: 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: 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 and Cramer-Rao Inequality: 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 Estimation: 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: Bayesian Inference: Basic elements of Bayesian Inference, Loss function, Bayes risk, Prior distribution, Bayes Theorem, Posterior distributions, Bayes estimation of parameters of standard distributions.

REFERENCES

1. Hogg, R. V. and Craig, A. T. (1989): Introduction to Mathematical Statistics, Macmillan Publishing Company.
2. Kale, B. K. (1999): A First Course on Parametric Inference, Narosa Publishing House.
3. Kendall, M.G. and Stuart, A. (1967): The Advanced Theory of Statistics, Vol. II, 2nd Edition, Charles Griffin & Company Ltd., London.
4. Lehmann, E. L. (1983): Theory of Point Estimation, John Wiley, New York.
5. 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.
6. Rajagopalan M. and Dhanavanthan, P. (2012) Statistical Inference, PHI Learning Private Limited, New Delhi.
7. Rohatgi, V. K. (1990): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Limited.
8. Wasan, M. T. (1970): Parametric Estimation, Mc-Graw Hill, New York.

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 behaviour 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

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Articulate and exemplify the concepts of Stochastic processes, time space and state space.	Understand
	MO2. Construction of examples of Stochastic processes	Create
	MO3. Explain the concepts of particular types of stochastic processes like process with stationary independent increments, Markov process, martingales, Wiener process, Gaussian process etc.	Apply
Module 2	MO1. Articulate concepts of Markov chains, transition probability matrix, n-step transition probabilities	Understand
	MO2. Calculate n-step transition probabilities	Evaluation
	MO3. Describe and exemplify classification of states in a Markov Chain	Understand
	MO4. Calculate the periodicity of a Markov Chain	Evaluation
	MO5. Explain the concepts of recurrence, ergodic chains	Understand

Module 3	MO1. Explain and exemplify concepts of limiting behaviour of n- step transition probabilities.	Understand
	MO2. Describe stationary distributions and solve problems	Apply
	MO3. Describe various types of random walks	Understand
	MO4. Explain a gambler's ruin problem	Understand
	MO5. Derivation of the probability of ruin of a gambler's ruin problem	Evaluation
Module 4	MO1. Describe and exemplify: Continuous time Markov chains, Poisson process, pure birth process, birth and death processes.	Understand
	MO2. Derive of steady state probabilities/differential difference equations in case of Poisson process, pure birth process, birth and death processes.	Evaluation
	MO3. Describe and exemplify: Compound Poisson process	Understand
	MO4 Derive properties of Poisson process and Compound Poisson process	Evaluation
	MO5. Explain the concept of Markov Process with discrete states.	Understand
	MO6. Illustrate these processes with examples	Apply
Module 5	MO1. Distinguish between strict and weak (covariance or wide sense) stationarity	Analyse
	MO2. Describe and exemplify: renewal processes, renewal equation.	Understand
	MO3. Explain the statement and applications of renewal theorem	Apply
	MO4. Solve problems based on the applications of renewal theorem	Apply
Module 6	MO1. Describe and exemplify Galton-Watson branching processes	Understand
	MO2. Explain concepts of offspring distribution and its implications	Apply
	MO3. Interpret the concept of extinction probabilities	Apply
	MO4. Compute the probability extinction in case of a particular offspring distribution	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 behaviour 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

1. Bartlett, M.S. (1955): An Introduction to Stochastic Processes (with special reference to application and methods), Cambridge.
2. Bhat, U.N. and Miller, G.K. (2002): Elements of Applied Stochastic Processes. 3rd Edn., John Wiley, New York.
3. Feller, W. (1968): Introduction to Probability Theory and its Applications, Vols. I & II, John Wiley, New York.
4. Karlin, S. (1972): A First Course in Stochastic Processes, Academic Press, New York.
5. Parzen, E. (1962): Stochastic Processes, Holden-Day Inc, San Francisco.
6. 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 : TIME SERIES ANALYSIS
CREDITS : 4

COURSE OUTCOMES

After completion of this course the students will be able to

C01	Analyse exploratory time series analysis using time series data
C02	Choose an appropriate model for time series data using the concept of linear time series models
C03	Predict the time series data using different stationary time series data
C04	Describe estimation methods for time series
C05	Forecasting using different forms of the ARIMA models
C06	MMSE Forecasting using ARMA and ARIMA models

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy level
Module 1	MO1. Analyse exploratory time series analysis using time series data	Analyse
	MO2. Stationary time series and autocorrelation functions	Understand
Module 2	MO1. Smoothing methods for time series	Analyse
	MO2. Forecasting based on smoothing	Apply
Module 3	MO1. Linear stationary processes	Analyse
	MO2. Analyse time series data using different stationary time series data	Analyse
Module 4	MO1. Estimation using Yule - Walker methods	Apply
	MO2. Maximum likelihood and least squares estimation for ARMA Processes	Analyse
Module 5	MO1. Linear non-stationary models	Analyse
	MO2. Forecasting using ARMA and ARIMA models	Analyse
	MO3. MMSE forecasts using different forms of the ARIMA models	Analyse
Module 6	MO1. Spectral analysis of stationary process	Understand
	MO2. Correlogram and periodogram analysis	Analyse
	MO3. Non-linear time Series models	Understand

COURSE CONTENT

MODULE I: Time series as a discrete parameter stochastic process, Classical regression in time series context. Characteristics of time series, exploratory time series analysis, test for trend and seasonality. Stationary time series, Autocorrelation (ACF) and cross correlations, partial auto correlation function, Estimation of autocorrelations.

MODULE II: Holt-winter additive and multiplicative models. Smoothing methods for time series: Adaptive smoothing, Kalman smoother, Moving average smoothing, Single exponential smoothing, Double exponential smoothing, Holt-winter smoothing. Forecasting based on smoothing.

MODULE III: Detailed study of the stationary process: Autoregressive, Moving Average, and Autoregressive Moving Average Models. Choice of AR MA periods. Wold representation of linear stationary processes.

MODULE IV: Estimation using Yule - Walker methods for AR, MA and ARMA Processes. Maximum likelihood and least squares estimation for ARMA Processes.

MODULE V: Linear non-stationary models: Autoregressive Integrated Moving Average (ARIMA) models, Integrated moving average (IMA) models. Forecasting using ARMA and ARIMA models. MMSE forecasts using different forms of the ARIMA models.

MODULE VI: Spectral analysis of weakly stationary process. Spectral analysis: Time series in frequency domain, spectral density and discrete Fourier transforms, estimation of spectral density. Correlogram and periodogram analysis. Seasonal ARIMA and GARIMA models. Introduction to non-linear time Series: ARCH and GARCH models. Auto Regressive Distributed Lag (ARDL) time series model.

REFERENCES

1. Box, G.E.P., Jenkins G.M. and Reinsel, G.C. (2007) Time Series Analysis: Forecasting and Control, Pearson Education.
2. Brockwell, P J. and David R. A. (2002). Introduction to Time Series and Forecasting, 2nd edition, Springer.
3. Draper, N. R. and Smith H. (1998). Applied Regression Analysis, 3rd edition. John Wiley
4. Ezekiel, M. (1963). Methods of Correlation and Regression Analysis.
5. Kleinbaum, D. G., Kupper, L. L., Muller, K. E. and Nizam, A. (1998). Applied Regression Analysis and Multivariable Methods, Duxbury Press.
6. Makridakis, S. and Wheelwright, S. C. (1983). Forecasting: Methods and Applications, 2nd edition, John Wiley and Sons.
7. Kutner, M. H., Nachtsheim, C. J. and Neter, J. (2004). Applied Linear Regression Models, 4th edition with student CD, Mc Graw Hill.
8. Montgomery, D. C., Jennings, C. L. and Kulahci, M. (2015). Introduction to Time Series Analysis and Forecasting, 2nd edition, John Wiley.

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

COURSE OUTCOMES

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

CO1	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
CO6	Derive likelihood ratio test for testing the hypothesis for normal populations
CO7	Obtain sequential probability ratio test for testing the hypothesis.
CO8	Construct confidence interval for parameters
CO9	Construct UMA and UMAU confidence sets
CO10	Perform non-parametric test for a given data
CO11	Formulate hypothesis for a given problem

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1: Identify simple and composite hypothesis	Understand
	MO2: Find critical region, size and power of the test	Evaluate
	MO3: Distinguish between random and non-randomized test	Apply
Module 2	MO1: Apply Neyman-Pearson lemma to find most powerful test	Apply
	MO2: Check the unbiasedness of a test	Apply
	MO3: Find UMP and UMPU test	Apply
Module 3	MO1: Apply likelihood ratio test principle for testing the mean for a normal population.	Apply
	MO2: Apply likelihood ratio test principle for testing the equality of means for two normal population	Apply
	MO3: Test the variance of normal population.	Apply
	MO4: Test the equality of means of two normal populations	Apply
Module 4	MO1: Derive SPRT for test the parameters of normal distribution, binomial and Poisson distributions	Apply
	MO2: Find OC function and Average sample Number of a SPRT	Evaluate
Module 5	MO1: Construct shortest level confidence interval for parameters	Evaluate
	MO2: Construct confidence interval using pivot	Evaluate
	MO3: Use UMP and UMPU test to construct UMA and UMAU confidence sets	Apply
Module 6	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, Neyman-Pearson lemma and its applications, most powerful tests, uniformly most powerful (UMP) tests.

MODULE II: Unbiasedness, uniformly most powerful unbiased (UMPU) tests, similar test, test with Neyman structure. locally most powerful (LMP) tests, locally most powerful unbiased (LMPU) tests.

MODULE III: Likelihood ratio tests, asymptotic properties, tests concerning mean and variance of normal distribution (one sample and two samples), test concerning means of k normal populations.

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

1. Ferguson, T. S. (1967): Mathematical Statistics, Academic Press, New York.
2. Kendall, M.C, and Stuart. A, (1967): The Advanced Theory of Statistics, Vol 2, IV Edn., Mc Millan, New York.
3. Lehmann, E. L. (1986): Testing of Statistical Hypothesis, John Wiley & Sons.
4. Wald, A. (1977): Sequential Analysis, Dover Publications Inc., New York.
5. Wetherill, G. B. (1966): Sequential Methods in Statistics, Methuen & Co. Ltd. New Delhi.
6. Siegel, S. and Castellan Jr. N. J. (1988): Non-parametric Statistics for the Behavioral Sciences. McGraw Hill, New York.
7. Randles, R.H. and Wolfe, D.A. (1979): Introduction to the Theory of Nonparametric statistics, Wiley New York.
8. 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 : 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

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Explain about machine learning and applications	Understand
	MO2. Thorough understanding about supervised,semi-supervised and unsupervised learning algorithms	Understand
Module 2	MO1: Introduction of estimation methods	Understand
	MO2: To get an idea about the inference of the data (population)	Understand
	MO3: Use of different types of classifications in data	Apply
Module 3	MO1. Explain about various clustering algorithms	Understand
	MO2. Introduction to data mining	Understand
	MO3. Apply the concepts of clustering algorithms into different types of data.	Apply
	MO4. Apply data reduction techniques	Apply
Module 4	MO1. A better understanding of SVM	Understand
	MO2. Apply SVM algorithms for face detection	Apply
	MO3. Analysis using data validation techniques	Evaluate
	MO4. Explain about data warehouse	Understand
	MO5. Various applications of data warehouse	Understand
Module 5	MO1: Good knowledge in both Multivariate analysis and missing data	Apply
	MO2: To get the understanding of the relationship between more than two variables through multivariate analysis	Apply
Module 6	MO1: Introduction and importance of dimensionality reduction.	Understand
	MO2: Understanding of methods used for the dimensionality reduction.	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

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

SEMESTER : III
COURSE CODE : STAD-DE-534(ii)
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
CO7	To model the data using lifetime distributions
CO8	To deal with associated methods of inference
CO9	Explain regression models used in survival analysis
CO10	Explain parametric inference and non-parametric inference for the regression models

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Define Survival function, Hazard function, Mean residual life function and median life.	Understand
	MO2. Familiar with parametric models for survival data such as exponential, Weibull, gamma and log-normal distributions.	Understand
Module 2	MO1. Understand Time-to-Event analysis and censoring in Time-to-Event analysis.	Understand
	MO2. Understand left and right censoring in TTE analysis	Understand
	MO3. To estimate the survival function.	Analyse
	MO4. Define Hazard ratio.	Remember
	MO5. Relate Hazard ratio, Risk ratio and Odds ratio.	Apply
	MO6. Discuss relative survival estimation.	Apply
Module 3	MO1. Analyze non-parametric methods for comparing survival distributions.	Analyse
	MO2. Find confidence interval for Hazard ratio.	Apply
Module 4	MO1. Describe Parametric lifetime distributions.	Understand
	MO2. Explain linear failure rate distribution.	Evaluate
	MO3. State and explain accelerated failure time model.	Remember
	MO4. Define Cox-Snell residuals.	Understand
Module 5	MO1. State and explain Cox's proportional hazard regression model.	Remember
	MO2. Derive rank test for the regression coefficient.	Understand
	MO3. Describe parametric and non- parametric inference for this model.	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 model, test with non-proportional hazards, parametric inference and nonparametric inference for this model.

REFERENCES

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

SEMESTER : III
COURSE CODE : STAD-DE-534(iii)
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

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Explains Mendelian ratios and their testing	Understand
	MO2. Explains Hardy-Weinberg law	Understand
	MO3. Discuss Estimation and testing of heterosis	Analyse
Module 2	MO1.Describe Average effect of a gene, breeding value	Understand
	MO2. Explains estimation of phenotypic, genotypic and environmental correlations. Selection, individual and pedigree selection	Remember
Module 3	MO1.Explains Design of plant breeding trials	Understand
	MO2. Discuss compact family block designs, square and rectangular lattice designs,	Remember
	MO3. Establish the multivariate analysis of variants, cluster analysis.	Apply
Module 4	MO1. Explain path analysis.	Understand
	MO2. Distinguish discriminant analysis for selection indices	Understand
	MO3. Explain partial diallel and line x tester crosses	Evaluate
Module 5	MO1. Explain Analysis of genotype	Understand
	MO2. Discuss measurement of stability and adaptability- linear and nonlinear approaches	Apply
	MO3. Analysis of micro array data	Analyse

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 nonlinear approaches, combining ability and specific combining ability in diallel process, analysis of micro array data.

REFERENCES

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

SEMESTER : IV
COURSE CODE : STAD-CC-541
COURSE TITLE : PRACTICAL FOR DATA ANALYTICS
CREDITS : 4

COURSE OUTCOMES

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

CO1	Evaluate the trend and seasonal fluctuations of a real-life time series data using R programming language
CO2	Interpret different multivariate techniques
CO3	Classify the different control charts for industrial data set using R programming language
CO4	Summarize different design of experiments for a statistical experimental data using R programming language
CO5	Recognize different nonparametric tests such as ANOVA, RBD etc. For statistical experimental data using R programming language.

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. 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.	Apply
	MO2. Illustrate the concepts of correlation and Regression	Apply
Module 2	MO1. 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	Evaluate
	MO2. 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	Create
	MO1. Perform tests of hypothesis using multivariate data such as Equality of means of two multivariate normal vectors	Apply
	MO2.Analyze the multivariate data using the multivariate techniques such as Canonical correlation, Principal component analysis, Factor analysis, Cluster analysis.	Analyze
Module 3	MO1. 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 4	MO1. 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.	Analyze

	MO1. Formulate different experimental techniques such as LSD, ANCOVA, Factorial experiments. Missing and Mixed-up plot in RBD using the R software	Create

COURSE CONTENT

MODULE I: Descriptive Statistics: Compare means- one sample t- test, two sample test, independent sample t test; chi-square test; Correlation; Regression-simple and multiple. Fitting of distributions.

MODULE II: 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 III: Equality of means of two multivariate normal vectors; Canonical correlation, Principal component analysis, Factor analysis, Cluster analysis. Mann-Whitney U-test, Kruskal-Wallis test.

MODULE IV: Control charts - Shewart control charts for variables- X, R, S charts, Attribute control charts - np, p, c and u charts. CRD, RBD, LSD, Factorial experiments.

SEMESTER : **IV**
COURSE CODE : **STAD-DE-542(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	CUSUM charts and EWMA mean charts
CO5	Define reliability including the different types and how they assessed
CO6	Ensure the validity and precision of statistical analysis
CO7	Explains reliability in discrete set up

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy level
Module 1	MO1. Explain statistical process control	Understand
	MO2. Describe various control charts	Analyse
Module 2	MO1. Describe different sampling inspection techniques	Evaluate
	MO2. Explains six-sigma concepts	Analyse
Module 3	MO1. Distinguish chain sampling and continuous sampling	Remember
	MO2. CUSUM charts and EWMA mean charts	Understand
Module 4	MO1. Explain basic concepts of reliability	Apply
	MO2. Describe reliability of a coherent system	Analyse
	MO3. Describe several measures of reliability	Understand
	MO4. Calculate measures of reliability based on several distributions	Apply
Module 5	MO1. Explain reliability in discrete set up	Understand
	MO2. Derive expression of the relation connecting various measures of reliability	Evaluate
Module 6	MO1. Explain inference in reliability models	Understand
	MO2. Describe estimation of parameters based on censored sampling	Apply
	MO3. Solve problems related to estimation of parameters based on censored sampling	Apply

COURSE CONTENT

MODULE I: Quality and Quality assurance, Methods of Quality assurance, Introduction to TQM. Statistical process control, Theory of control charts - Shewart control charts for variables: $\bar{X}$, R, σ charts, Attribute control charts - np, p, c and u charts - OC, ARL & process capability of control charts. Acceptance sampling for attributes and variables.

MODULE II: Sampling inspection techniques: Single, double and multistage sampling plans and their properties. Sequential sampling plans. Taguchi method, Experimental design in Taguchi Methods, ISO standardization, ISO 9001, six sigma concepts.

MODULE III: Chain sampling, Continuous sampling plan. Process capability studies, CUSUM charts, EWMA mean charts, Statistical process control, Modelling and quality programming, Economic design of control charts.

MODULE IV: 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 V: Reliability concepts in discrete set up, Notions of ageing - IFR, IFRA, DFR, DFRA, NBU, NBUE, NWU, NWUE, IMRL, DMRL, HNBUE, HNWUE and their mutual implications. TTT transforms and characterization of ageing classes.

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. Reliability estimation using MLE: exponential, Weibull and gamma distributions. Bayesian reliability estimation of exponential and Weibull models.

REFERENCES

1. Barlow, R.E. and Proschan, F. (1985). Statistical Theory of Reliability and Life Testing, Holt, Rinehart and Winston.
2. Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data, Chapman Hall.
3. Duncan, A. J. (1959). Quality Control and Industrial Statistics (5th edition), Irwin, Homewood I.
4. Galambos, J. and Kotz, S. (1978). Characterizations of Probability Distributions.: A Unified Approach with an Emphasis on Exponential and Related Models, Springer-Verlag Berlin and Heidelberg GmbH & Co.
5. Klefjo, B. (1982). The HNBUE and HNWUE Classes of Life distributions, Naval Research Logistic Quarterly, 29, 331-344.
6. Lawless, J. F. (2003). Statistical Models and Methods for Lifetime Data, John Wiley.
7. Montgomery, D.C. (2005). Introduction to Statistical Quality Control, 5th edition, John Wiley.
8. Nelson, W. (1982). Applied life data analysis, Wiley.
9. Sinha, S. K. (1986). Reliability and Life Testing, Wiley.

SEMESTER : IV
COURSE CODE : STAD-DE-542(ii)
COURSE TITLE : OFFICIAL STATISTICS
CREDITS : 4

COURSE OUTCOMES

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

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. A general understanding of the official statistics system that prevails in our country and others.	Understand
	MO2. Responsibilities of various central and state divisions under the Indian official statistics system.	Evaluate
	MO3. Dissemination and use of data collected.	Apply
Module 2	MO1. Learn about population and related issues.	Understand
	MO2. Learn various health and family welfare schemes.	Analyse
	MO3. Understand in detail the Population Census in India.	Remember
Module 3	MO1. Understand various other social statistics.	Understand
	MO2. Understand various index numbers used in the official statistics system.	Analysis
Module 4	MO1. Learn the economic aspects of growth and development.	Understand
	MO2. Estimation of national income through various approaches.	Apply
Module 5	MO1. Understand various income inequality measures.	Understand
	MO2. Understand various poverty measurement methods.	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

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

SEMESTER : IV
COURSE CODE : STAD-DE-542(iii)
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.
CO5	Test for independence in two -way and three- way contingency tables.

MODULE OUTCOMES

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. How to model binary response variables.	Understand
	MO2. Explain logit model.	Analyse
Module 2	MO1. Model multinomial response variables.	Understand
	MO2. Analyse repeated categorical response data.	Analyse
Module 3	MO1. Analyse two way and three-way contingency tables.	Understand
	MO2. Model data with repeated measurements.	Analyse
	MO3. Explain logistic regression.	Remember
Module 4	MO1. Tests for independence and homogeneity of proportions for contingency tables.	Understand
	MO2. Explain Fishers exact test.	Analyse
	MO3. Understand Simpson's Paradox.	Understand
Module 5	MO1. Explain Polytomous logit models.	Understand
	MO2. How to model for multiway tables.	Analyse

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 x 2and r x 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

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

SEMESTER : IV
COURSE CODE : STAD-DE-542(iv)
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

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Explain and articulate the background of linear, count and binary regression analysis.	Understand
	MO2. Exemplify the various concepts on Linear, log- linear and logistic regression models for continuous, count and binary data.	Remember
	MO3. Explain the different methods of estimation and comment on the asymptotic properties of these estimators for regression effects.	Understand
Module 2	MO1. Exemplify the concept of longitudinal/panel data.	Remember
	MO2. Illustrate the different types of longitudinal data with examples.	Apply
Module 3	MO1. Explain how to fit a binary dynamic logits (BDL) model to the longitudinal stationary/non-stationary binary data.	Apply
	MO2. Distinguish between binary dynamic logits (BDL) and binary auto-correlation models.	Evaluate
	MO3. Comment of the concept of correlation structure under the BDL model.	Analyse
	MO4. Exemplify the concept of GQL and likelihood estimation. Also, illustrate and estimate the model using asthma data.	Apply
Module 4	MO1. 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	MO1. Describe the analysis of longitudinal count and binary data in mixed effects model setup.	Remember
	MO2. Explain the AR(1) type mixed effect models for count and binary data.	Evaluate
	MO3. Exemplify the method of GQL estimation for regression and random effects variance parameters and illustrate it with suitable data.	Analyse

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

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

SEMESTER : IV
COURSE CODE : STAD-DE-543(i)
COURSE TITLE : OPTIMIZATION TECHNIQUES
CREDITS : 4

COURSE OUTCOMES

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

CO1	Describe Simplex method to solve the linear programming problem.
CO2	Explain the steps for 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

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Explain the concepts of linear programming problem.	Understand
	MO2. Solve the linear programming problem by using Simplex method.	Apply
	MO3. Computational steps of Big M Method	Remember
Module 2	MO1. Write the dual of the given linear programming problem.	Evaluate
	MO2. Find the initial basic feasible solution to the given Transportation problem.	Evaluate
	MO3. To determine the optimum assignment problem.	Remember
Module 3	MO1. To be familiar the concept of general linear programming problem	Understand
	MO2. State and prove Kuhn-Tucker necessary and sufficient conditions in a non-linear programming problem.	Remember
	MO3. To solve the non-linear programming problem.	Apply
Module 4	MO1. Explain with examples the probabilistic models in inventory.	Understand
	MO2. Derive an expression for EOQ for Harri's or Wilson model.	Understand

	MO3. Discuss the purchase inventory model with three price breaks.	Remember
Module 5	MO1. State some of the important distributions of arrival interval and service time.	Remember
	MO2. To evaluate expected value of queue length in various models.	Evaluate
	MO3. Obtain the differential difference for M/M/1 queue model.	Understand
Module 6	MO1. Describe the problem of replacement of items whose maintains cost increase with time.	Understand
	MO2. Solve dynamic programming problem by using Linear programming problem.	Apply
	MO3. Describe dynamic programming problem	Remember

COURSE CONTENT

MODULE I: Linear Programming: convex sets and associated theorems, graphical method, definition of linear programming problem, properties of a solution to the linear programming problem, generating extreme-point solutions, simplex computational procedure, artificial variables technique - big M method and two-phase method. Degeneracy in linear programming problem.

MODULE II: Duality problems of linear programming, duality theorems, duality and simplex method. Transportation problems: - general transportation problem, Finding initial basic feasible solution, test for optimality, degeneracy in transportation problem, unbalanced transportation problem. Assignment problem: - mathematical formulation of the problem, the assignment method (Hungarian method).

MODULE III: Non-linear programming problem (NLPP):- general non-linear programming problem, Constrained optimization with equality constraints - necessary conditions for a generalized NLPP, sufficient conditions for a general NLPP with one constraint, sufficient conditions for a general problem with m(<n) constraints, constrained optimization with inequality constraints - Kuhn-Tucker conditions for general NLPP with m(<n) constraints. Quadratic programming problem, convex programming problems.

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

MODULE V: 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 VI: 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. Dynamic programming problem, recursive approach.

REFERENCES

1. Gass, S.I. (1969): Linear Programming Problem, Mc Graw Hill.
2. Gross, D. and Harris, C.M. (1974): Fundamental of Queuing Theory, John Wiley.
3. Hillier, F.S. and Leiberman, G.J. (1962): Introduction to Operations Research, Holden Day.
4. Kanti Swarup, Manmohan and Gupta, M.M. (1985): Operations Research, Sultan Chand & Sons.
5. Mittal, K.V. (1990): Optimization Methods.
6. Ravindran, A, Philips, D.T. and Soleberg, J.J. (1997): Operations Research - Principles and Practise.
7. Saaty, T.L. (1961): Elements of Queuing Theory with Applications, Mc Graw Hill.
8. Taha, H.A. (1997): Operations Research, McMillian.

SEMESTER : IV
COURSE CODE : STAD-DE-543(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

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Explains the basic concepts of financial markets	Understand
	MO2. Explain the concepts of forward contracts, future contracts, options call and put options, European option and American options, hedgers, speculators, arbitrageurs	Analyse
	M30. Define the interest rates, compounding, present value analysis, risk free interest rates	Evaluate
	MO4. Solving problems related to interest rate, compounding, present value analysis, risk free interest rates	Apply
Module 2	MO1. 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	MO1. 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).	Apply
	MO2. state Arbitrage theorem and The Black-Scholes formula and also describe the Properties of the Black-Scholes option cost.	Analyse
Module 4	MO1. Estimating the volatility parameter.	Understand
	MO2. Describe the pricing American options, pricing of European options using Monte Carlo and pricing an American option using finite difference methods.	Evaluate

Module 5	MO1. 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.	Understand
	MO2. Examine the ARCH(1), ARCH(p), GARCH(p,q) models and their estimation.	Apply
	MO3.Distinguish between ARMA and GARCH processes	Analyse

COURSE CONTENT

MODULE I: Basic concepts of financial markets, forward contracts, future contracts, options call 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

1. David Ruppert (2004). Statistics and Finance - An Introduction, Springer International edition
2. Sheldon M Ross (2003). An Elementary Introduction to Mathematical Finance, Cambridge University Press
3. Christian Gouriesous and Joann Jasiak (2005). Financial Econometrics, New Age Interational (P) Ltd
4. John C Hull (2008). Options, Futures and other Derivatives, Pearson Education, India

5. Masaaki Kijima (2003). Stochastic Process with Applications to Finance, Chapman and Hall
6. Ruey S Tsay (2005). Analysis of Time Series, Third Ed., John Wiley and Sons.
7. Cuthbertston K and Nitzsche D (2001). Financial Engineering- Derivatives and Risk Management, John Wiley and Sons.

SEMESTER : **IV**
COURSE CODE : **STAD-DE-543(iii)**
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

	On completion of each module, students should be able to:	Taxonomy
Module 1	MO1. Explain the theoretical and practical meanings of spatial patterns and relationships.	Understand
	MO2. Explain the use of spatial data analysis in geographic domain (Geospatial data analysis).	Evaluate
	MO3.Explain the spatial aspects of computational, statistical, and visual analytical methodologies for big data analytics.	Remember
Module 2	MO1. Understand various applications of domain specific data analysis.	Understand
	MO2. Solve simple problems related to simulation.	Apply
	MO3. Handle heterogeneous data.	Apply
	MO4. Explain various aspects of predictive analytics and its applications.	Understand
	MO5. Explain the needs of complexity measures and complexity modeling.	Understand
Module 3	MO1. Explain the concept of spatial clustering.	Remember
	MO2. Distinguish between quadrant and nearest neighbor methods.	Remember
	MO3. Find the difference between K-function and L-function.	Analyse
	MO4. Find the applications of point process models.	Analyse
Module 4	MO1. Explain the applications of exploratory spatial data analysis (ESDA)	Understand
	MO2. Understand and plot experimental variogram.	Understand
	MO3. Interpret scatter plots.	Apply

	MO4. Understand stationarity and isotropy with examples.	Understand
Module 5	MO1. Explain deterministic estimation.	Understand
	MO2. Articulate Estimation Criteria.	Apply
	MO3. Evaluate the uses of Geostatistical Estimation.	Evaluate
	MO4.Understand the multivariate variant of Ordinary Kriging Operation (CoKriging).	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, Scatterplot, 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

1. Li, Deren. Spatial Data Mining: Theory and Application. S.l.: Springer, 2016.
2. Berry, Brian J. L., and Duane Francis Marble. Spatial Analysis; a Reader in Statistical Geography. Englewood Cliffs, NJ: Prentice-Hall, 1968.
3. Bivand, Roger S, Edzer J Pebesma, and Virgilio Gomez-Rubio. Applied Spatial Data Analysis with R. New York: Springer, 2008.
4. Lloyd, Christopher D. Spatial Data Analysis: An Introduction for GIS Users. Oxford: Oxford UP, 2010.
5. 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-CC-544
COURSE TITLE : DISSERTATION/ PROJECT & VIVA
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.