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

Name of the Programme:

M.Sc. Statistics with Specialization in Data Analytics

(Syllabus effective from 2020 Admission onwards)

UNIVERSITY OF KERALA

Syllabus for M.Sc. Statistics with Specialization in Data Analytics

Programme Specific Outcomes (PSO) for

M.Sc. Statistics with Specialization in Data Analytics

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

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

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

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

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

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

Structure of the Programme

Sem. No.	Course code	Name of the Course
I	**Core Courses** STSD 211 STSD 212 STSD 213 STSD 214	 Analytic Tools for Statistics Probability Theory Statistical Programming Using R Statistical Programming Using Python
II	**Core Courses** STSD 221 STSD 222 STSD 223 STSD 224	 Distribution Theory Design of Experiments and Theory of Sampling Statistical Modeling Practical I (Using R and Python)
III	**Core Courses** STSD 231 STSD 232 STSD 233 STSD 234	 Machine Learning Statistical Inference Elective Practical II (Using R and Python)
	Electives Courses STSD 233(i) STSD 233(ii) STSD 233(iii) STSD 233(iv) STSD 233(v) STSD 233(vi)	 Order Statistics Multivariate Analysis Data Mining and their Applications Categorical Data Analysis Longitudinal Data Analysis Operations Research
IV	**Core Courses** STSD 241 STSD 242 STSD 243 STSD 244 STSD 245 STSD 246	 Stochastic Processes and Time Series Big Data Analytics and Artificial Intelligence Elective Practical III (Using R and Python) Project/Internship Viva voce

	Electives Courses	
	STSD 243(i)	Reliability Modeling and Statistical Quality Control
	STSD 243(ii)	Official Statistics
	STSD 243(iii)	Stochastic Finance
	STSD 243(iv)	Spatial Data Analytics
	STSD 243(v)	Actuarial Statistics
	STSD 243(vi)	Bayesian Inference

Pattern of Question Papers for the End Semester Examination

Theory Papers:

For each paper the duration of the examination is 3 hours and maximum mark is 75. The question paper will have 3 parts: Part A, Part B and Part C. Part A will consist of 10 short answer questions (two questions from each of the 5 units), each carrying 3 marks and a candidate has to answer any 5 of them. Part B will consist of 6 questions (2 questions each from Unit 1, Unit 2 and Unit 3), each carrying 12 marks and the candidate has to answer 3 questions from this part. Part C will consist of 4 questions (2 questions each from Unit 4 and Unit 5) each carrying 12 marks and the candidate has to answer 2 questions from this part.

Practical Papers:

The practical paper will be conducted using R and Python programme. For each practical paper, a record of work done by the student should be prepared and submitted for internal evaluation. The components of CA mark for the practical paper are Attendance (5 marks), Record book (10 marks) and class test (10 marks). The Board of examiners will prepare the question paper for the practical examination (ESE) covering the papers specified in the syllabus. An external examiner along with an internal examiner, appointed by the University will conduct the practical examination and its evaluation. For each practical paper, the duration of the examination is 3 hours and the maximum mark is 75.

The question paper of Practical I examination will consist of 3 parts: Part A, Part B and Part C which corresponds to the papers STSD 221, STSD 222 and STSD 223. There will be 3 questions in each Part and a candidate has to answer 5 questions without omitting any Part. Each question carries 15 marks.

The question paper of Practical II examination will consist of 3 parts: Part A, Part B and Part C which corresponds to the papers STSD 231, STSD 232 and STSD 233. There will be 3 questions in each Part and a candidate has to answer 5 questions without omitting any Part. Each question carries 15 marks.

The question paper of Practical -III examination will consist of 3 parts: Part A, Part B and Part C which corresponds to the papers STSD 241, STSD 242 and STSD 243,. There will be 3 questions in each Part and a candidate has to answer 5 questions without omitting any Part. Each question carries 15 marks.

Project/Internship

A project work has to be done using primary or secondary data and to be submitted at the end of the fourth semester. For field survey project, a maximum of three students can do the survey jointly on the same topic and the data analysis and reporting can be done either jointly or independently. The project report of 30 to 40 pages should consist of literature review, methodology, data analysis and summary.

SEMESTER : I

COURSE CODE : STSD 211

COURSE TITLE : ANALYTIC TOOLS FOR STATISTICS

Course outcomes

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

CO1. Describe classes of open and closed sets of R.

CO2. Describe the concept of compactness

CO3. Describe Metric space - Metric in R^n.

CO4. Use the concept of Cauchy sequence, completeness, compactness and connectedness to solve the problems

CO5. Explain the concept of Riemann- integral and Describe the properties of Riemann- integral.

CO6. Apply integral calculus in problem solving

CO7. Compute Partial derivatives of functions of several variables

CO8. Compute maxima, minima of functions

CO9. Compute conditional maxima and conditional minima

CO10. Describe vector space, subspaces, independence of vectors, basis and dimension.

CO11. Apply matrices and determinants to solve problems

CO12. Apply linear transformations in statistics

CO13. Solve system of linear equations by various methods

CO14. Apply the concept of Eigen values, Eigen vectors and related results in statistics.

CO15. Apply quadratic forms and reduction of quadratic forms in statistical problem solving.

CO16. Apply g-inverse in statistics.

Module Outcomes

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy Level**
MODULE: 1	MO1. Describe classes of open and closed set MO2. Apply the concept of compactness MO3. Describe Metric space - Metric in R^n. MO4. Use the concept of Cauchy sequence, completeness, compactness and connectedness to solve the problems	Understand Understand Understand Apply
MODULE 2.	MO1. Explain the concept of Riemann- integral MO2. Describe the properties of Riemann- integral. MO3. Solve integration problems MO4. Apply integral calculus in solving statistical problems	Understand Understand Evaluate Apply
MODULE: 3	MO1. Compute Partial derivatives of functions of several variables MO2. Compute maxima, minima of functions MO3. Compute conditional maxima and conditional minima MO4. Apply partial derivatives, maxima and minima in statistics	Evaluate Evaluate Evaluate Apply
MODULE 4	MO1. Describe vector space, subspaces, independence of vectors, basis and dimension. MO2. Solve problems using matrices and determinants MO3. Solve system of linear equations by various methods MO4. Apply linear transformations in statistics MO5. Use methods of solution of system of linear equations in statistics	Understand Evaluate Evaluate Apply Apply
MODULE 5	MO1. Describe Eigen values, Eigen vectors and related results MO2. Apply Eigen values, Eigen vectors and related results in statistics MO3. Apply quadratic forms and reduction of quadratic forms in statistical problem solving. MO4. Apply g-inverse in statistics.	Understand Apply Apply Apply

Course Content

MODULE 1

Euclidean space R^n, open balls, interior point, open sets, limit points, adherent points, closed sets, Bolzano – Weierstrass theorem, Cantor intersection theorem(statement only), compactness in R^n, Heine-Borel theorem(statement only), Metric space (definition and examples). Compact subsets of a metric space, sequence in metric space, convergent sequence, Cauchy sequence(definition and concept), completeness of metric space, limit of real valued functions, continuous functions, continuity and inverse images of open and closed sets, Connected sets, uniform continuity and monotone functions (definition, examples and applications only).

MODULE 2

Definition and existence of Riemann integral(concepts only), Riemann-Stieltjes integral, its reduction to Riemann integral, properties of Riemann-Stieltjes integrals (viz. linearity, product, quotient and modulus of integrals). Fundamental theorem of integral calculus, mean value theorem (statement only), Functions of bounded variation, properties, total variation and additive property, continuous functions of bounded variation.

MODULE 3

Functions of several variables, partial derivatives, maximum and minimum of functions, conditional maxima and minima, Lagranges multiplier method. (problems only).

MODULE 4

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(problems only).

MODULE 5

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

TEXT BOOKS

1. Apostol T. M. (1974): Mathematical Analysis, Narosa Publishing House, New Delhi.
2. Malik, S.C., Arora, S. (2012): Mathematical Analysis, New Age International, New Delhi

3. Biswas, S. (2012). Textbook of Matrix Algebra, Third edition, PHI Learning Pvt Ltd, New Delhi.
4. Graybill,A and Belmont, C.A.(1983): Matrices with Applications in Statistics, II Edition, John Wiley, New York.
5. Mathai, A.M. (1999). Linear Algebra (Part I, II & III), Centre for Mathematical Sciences, Trivandrum.

REFERENCES

1. Pringle, R.M. and Rayner,A. A(1971): Generalized Inverse of Matrices with Application to Statistics, Griffin, London.
2. Rao,C.R (1973): Linear Statistical Inference and its Applications, Wiley Eastern, NewYork.
3. Roydon, H. L. (1968) : Real Analysis, Macmillan, New York.

SEMESTER : I
COURSE CODE : STSD 212
COURSE TITLE : PROBABILITY THEORY

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

CO7. Identify the applications of various moment inequalities

CO8. Explain the concept of convergence and check for the of convergence of a given sequences of random variables.

CO9. Find the expressions for the characteristic function of a random variable and verify its properties

C10. Apply the various laws of large numbers to sequences of random variables.

Module Outcomes

Sl.No	**Outcomes** On completion of each module, students should be able to:	**Taxonomy Level**
MODULE 1	MO1. Explain and exemplify the concepts of probability space MO2. Examine the various properties of a probability measure. MO3. State and explain various results associated with probability measure MO4. Explain the concepts of independence of events MO5. Construct examples of independent events MO6. Construct counter examples for proving/illustrating certain results associated with probability measure MO7. Evaluate the conditional probability and verify its properties MO8. Articulate the Bayes theorem and apply it to calculate apriori probabilities	Understand Analysis Understand Remember Create Create Evaluate Apply
MODULE 2	MO1. Distinguish between various types of random variables and articulate their properties MO2. Describe the various functions of random variables and find the functions for various random variables MO3. Explain and exemplify the concepts of independent of sequence of random variables.	Apply Apply Understand
MODULE 3	MO1. Calculate the mathematical expectation, moments and generating functions of random variables MO2 Articulate the various theorems associated with the characteristic functions and identify their applications.	Evaluate Apply
MODULE 4	MO1. Apply the various moment inequalities. MO2.Articulate and appraise stochastic convergence of sequence of random variables MO3.Construct counter examples for not satisfying certain convergence implications.	Apply Understand Evaluate
MODULE 5	MO1. State various laws of large numbers MO2. Apply the laws to sequences of random variables MO3. Articulate the concepts of central limit theorems. MO4. Apply various central limit theorems	Understand Apply Understand Apply

Course Content

MODULE 1

Sequence of sets, limit supremum, limit infimum and limit of sequence of sets, Monotone sequence of sets. Class of sets- Semi ring, ring, sigma ring (definition and examples only), field and sigma field. Borel sigma field and monotone class. Set functions, additive set functions and sigma additive set functions. Measure and its properties. Measure space., finite measure, sigma finite measure, complete measure, counting measure and signed measure (definition and examples only). Measurable functions and properties (viz. linearity, product, maxima, minima, limit sup, limit inf, and modulus of measurable functions) Sample space and events, probability space. Probability measure, Limit of sequence of events, monotone and continuity properties of probability measure. Independence of sequence of events, conditional probability and Bayes theorem. Borel- Cantelli lemma and Borel zero-one law.

MODULE 2

Random Variables, discrete and continuous-type random variables, probability distribution and distribution function, properties of distribution function. Functions of a random variable. Independence of sequence of random variables.

MODULE 3

Expectation of random variables and its properties. Probability generating function, moment generating function and cumulant generating function. Characteristic function (c.f.) and their elementary properties, uniform continuity and non-negative definiteness of characteristic function. Uniqueness theorem. Inversion theorem (without proof). Fourier inversion theorem. Convolution theorem. Levy's continuity theorem (without proof) and Bochners theorem (without proof).

MODULE 4

Inequalities-Markov, tChebychev's, Lyapunov (for moments) and Jensen's inequality. Stochastic convergence of sequence of random variables: - convergence in probability, almost sure convergence, convergence in p^{th} mean, weak and complete convergence of distribution functions and their interrelations. Slutsky's theorem and its applications.

MODULE 5

Stochastic series of sequence of random variables: - Law of large numbers, weak law of large numbers due to Bernoulli, Tchebyshev and Khintchine. Strong law of large numbers- Kolmogorov's strong law of large numbers for independent and identically distributed random variables and for independent random variables (both statement and applications only). Central limit theorem: Classical Central limit theorem, De Moivre-Laplace Central limit theorem, Lyapunov Central limit theorem (without proof) and Lindberg-Feller Central limit theorem (without proof). Applications of various central limit theorems.

Text books:

1. Kingman, J.F.C. and Taylor, S.J. (1977). A text book of Introduction to Measure Theory and Probability, 3rd Edn., Cambridge University Press, London .
2. Laha, R.G. and Rohatgi, V.K. (1979). Probability Theory, John Wiley, New York.
3. Rohatgi, V.K. and Saleh, Ehsanes (2014). An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd.
4. Bhat, B.R.(1991). Modern Probability Theory, 2nd Edn., Wiley Eastern Ltd., New Delhi.

References:

1. Cacoullos, T. (1989). Exercise in Probability, Springer-Verlag, New-York.
2. Feller W. (1968) Introduction to Probability Theory and Its Applications Vol. 1 and 2, John Wiley, New York.
3. Loeve, M (1968). Probability Theory Allied East-West Press.
4. Roussas, G.G.(2014). An Introduction to Measure-Theoretic Probability, Academic Press, USA.

SEMESTER : I

COURSE CODE : STSD 213

COURSE TITLE : STATISTICAL PROGRAMMING USING R

Course Outcomes

After completion of this course the students will be able to

C01. Write programs for statistical applications and Data Analysis

C02. Develop R packages for statistical applications

C03. Use built-in packages for Data Analysis

C04. Modify the built-in R packages

Module Outcomes

Sl.No.	Outcomes On Completion of each module, Students should be able to:	Taxonomy level
MODULE 1.	MO1.Download and install R and R packages	Remember and Apply
	MO2. Apply R syntax and R objects	Understand and Apply

MODULE.2	MO1.Apply built –in functions in R MO2: Read data from various packages/databases/spreadsheets MO3.Reshape and manipulate Data MO4. Write own functions in R	Understand and Apply Understand and Apply Understand and Apply Understand Apply and create
MODULE.3	MO1.Apply Graphical functions in R MO2.Write programs to create graphs/plots	Understand and Apply Apply and Create
MODULE4	MO1. Transform Data by filtering ordering aggregating etc. MO2. Handle non-tabular Data MO3.Acquire basic knowledge on bioconductor and Hadoop.	Understand and Apply Understand and Apply Understand
MODULE 5	MO1. Write SQL statements in R MO2. Create dynamic reports	Understand and Apply Understand

Course Content

MODULE 1

Installing R, R user interface, An overview of R packages-exploring R package repositories, **Meaning of**:- expressions, objects, symbols, functions, special values(NA, Inf and –Inf, NaN, NULL), R interpreter. **R Syntax**:- Constants, Numeric vectors, Character vectors, Symbols, Order of operations, Assignments, use of semi colons-parentheses- curly braces, Conditional statements, Loops, indexing by Integer vector, indexing by Logical vector, indexing by Name. **R Objects**:- Object types, vectors, lists, matrices, arrays, missing values, factors, data frames, formulas, time series, shingles, dates and times, connections, attributes, class.

MODULE 2

Functions - The function keyword, Arguments, Return values, Functions as arguments, Anonymous functions, properties of functions, calling basic functions, math functions, vector and matrix operations, solving linear systems, Calculating Minima and Maxima- Calculus statistical functions, **Reading Data into R-** Reading CSVs,Excel Data, Reading from Databases, Data from other Statistical Tools, R Binary Files, Data included with R, Extract Data from Websites, Entering data within R, Merging data. **Manipulating Data-** Apply Family, aggregate, plyr, data.table,. **Data Reshaping-** cbind, rbind, joins, reshape2. Manipulating Strings – paste, sprintf, extracting Text, Regular Expressions. **Writing functions in R, function documentation**.

MODULE 3

R-Graphics: - An overview of R graphics, Scatterplots, Bar charts, Histogram, Pie charts, Plotting time series, Box plots, Lattice Graphics, Stem and leaf plots, Q-Q plots,Graphical parameters, Basic graphic functions, Drawing- mathematical functions, Logarithmic functions, Trigonometric functions, polynomial functions, dotchart and violinplot examples,.Using the ggplot2 package to visualize data, Applying themes from ggthemes to refine and customize charts and graphs. Building data graphics for dynamic reporting.

MODULE 4

More Complex Data Transformations- Filtering and ordering data, Summaries and aggregates, New variables, Relational data, Joins on Keys, Introduction into fuzzy joins, Transforming wide and long tables. **Introduction to Non-tabular Data Types**- Time-series, Spatial data, Network data. Big Data Problems - Define Big Data citing examples. Data Transformations - Converting Numeric Variables into Factors, Date Operations , String Parsing, Geocoding. Dirty Data Problems- missing values, data imputation, duplicates, forms of data dates, outliers, spelling.**Introduction to Bioconductor** (introduction of Data structures-eSet,AssayData,MIAME etc. and key bioconductor packages). Introduction to **R and Hadoop**.

MODULE 5

Data querying- SQL and R -Writing SQL statements in R , Using the Select, From, Where, Is, Like, Order By, Limit, Max, Min SQL functions. Dynamic Reports and Reproducible Research-Installing a LATEX Program, LATEX Primer, Using knitr with LATEX, Introduction to markdown, First steps with knitr and Markdown in R with pandoc, Chunk options and document formats in rmarkdown and kintr.

TEXTBOOKS

1. Wickham, H. & Grolemund, G. (2018). R for Data Science. O'Reilly: New York. Available for free at http://r4ds.had.co.nz

2. Adler, J. (2010). *R in a nutshell: A desktop quick reference*. " O'Reilly Media, Inc.".

.

REFERENCES

1. Braun, W. J., & Murdoch, D. J. (2016). *A first course in statistical programming with R*. Cambridge University Press.
2. Bloomfield, V. A. (2018). *Using R for numerical analysis in science and engineering*. CRC Press.
3. Everitt, B.S. and Hothorn T. (2010) A Handbook of Statistical Analysis Using R, Second Edition, CRC Press.
4. Fonseca i Casas, P., & Tormos, R. (2018). Using the R language to manage and show statistical information in the cloud. *Technologies*, *6*(4), 113.
5. Michael J. Crawley (2013) The R book, Second Edition, John Wiley & Sons Ltd.
6. Rubinstein, R.Y. (1981) Simulation and Monte Carlo Methods, Wiley.
7. Sosulski, K. (2018). R Fundamentals. Bookdown: New York. Available at: http://becomingvisual.com/rfundamentals (FREE).
8. Dietze, M. (2018). The R package "eseis"–a software toolbox for environmental seismology. Earth Surface Dynamics, 6, 669-686.

Required software

1. R: http://www.r-project.org/ (FREE)

2. RStudio (additional libraries required): http://www.rstudio.com/ (FREE)

Learning resources

1. R Project: http://www.r-project.org/

2. RStudio (additional libraries required): http://www.rstudio.com

3. Quick-R http://www.statmethods.net/

4. Google's R Style Guide: http://google-styleguide.googlecode.com/svn/trunk/ Rguide.xml

SEMESTER : I

COURSE CODE : STSD 214

COURSE TITLE : STATISTICAL PROGRAMMING USING PYTHON

COURSE OUTCOMES

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

CO1. Write programs in Python

CO2. Analyze data using Python

CO3. Use Python for statistical Data Analytics

MODULE OUTCOMES

Sl. No:	Outcomes On completion of each module, students should be able to:	Taxonomy Level
MODULE: 1	MO1. Use correct syntax in Python programs MO2. Use data types and operators in Python programs MO3. Use control flows in Python programs	Understand and Apply Understand and Apply Understand and Apply
MODULE 2.	MO1. Use Data Structures in Python MO2. Use Mutable and Immutable Objects	Understand and Apply Understand and Apply
MODULE: 3	MO1. Use Functions in Python MO2. Use Types of variables in Python MO3. Use Modules in Python MO4. Use Packages in Python	Understand and Apply Understand and Apply Understand and Apply Understand and Apply
MODULE 4	MO1. Working with Databases - SQL MO2. Use SQLite and MYSQL MO3. Create Database MO4. Pull Data from a DB	Understand and Apply Understand and Apply Understand and Apply Understand and Apply
MODULE 5	MO1. Program Data Visualization in Python MO2. Write programs to create diagrams and graphs	Understand and Apply Understand and Apply

COURSE CONTENT

MODULE 1

Part A - Introduction (Not for examination purpose): History of Python, Need of Python Programming, Installing Python, Applications

Part B - Basics of Python Programming, Running Python Scripts, Using the Terminal Command Prompt, IDLE, and Other IDEs, Variables, Assignment, Keywords, Input-Output, Indentation. Types, Operators and Expressions: Types - Integers, Strings, Booleans; Operators- Arithmetic Operators, Comparison (Relational) Operators, Assignment Operators, Logical Operators,

Bitwise Operators, Membership Operators, Identity Operators, Expressions and order of evaluations Control Flow- if, if-elif-else, for, while, break, continue, pass, Exception handling.

MODULE 2

Data Structures: Lists - Operations, Slicing, Methods; Tuples, Sets, Dictionaries, Built-in methods of lists, sets and dictionaries, Mutable and Immutable Objects.

MODULE 3

Functions: Defining Functions, Calling Functions, Passing Arguments, Recursion, Keyword Arguments, Default Arguments, Variable-length arguments, Anonymous Functions, Fruitful Functions (Function Returning Values), Scope of the Variables in a Function - Global and Local Variables. Modules: Built-in Modules, Creating modules, import statement, from. Import statement, name spacing. Python packages: Introduction to PIP, Installing Packages via PIP.

MODULE 4

Files: file Operations, creating a file, reading from a file, writing file. Working with Databases: working with a database, using SQL to query a database, Python and SQLite and MYSQL, Creating Database, Pulling Data from a DB

MODULE 5

Popular Libraries for Data Visualization in Python: Matplotlib, Seaborn, Plotly, Geoplotlib, and Pandas. Data Visualization: Direct Plotting, Line Plot, Bar Plot, Pie Chart, Box Plot, Histogram Plot, Scatter Plot, Seaborn Plotting System , Strip Plot , Box Plot, Swarm Plot, Joint Plot , Matplotlib Plot , Line Plot Bar Chart ,Histogram Plot ,Scatter Plot , Stack Plot and Pie Chart.

Text Books:

1. Chun, W. (2006) Core python programming. Prentice Hall Professional.
2. Embarak, O. (2018). Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems. Apress.
3. Lambert, K. A. (2011). Fundamentals of Python: First Programs. Cengage Learning.

References:

4. Thereja, R. (2019). Python Programming Using Problem Solving Approach. Oxford University Press
5. Kurniawan, A (2019). Python and SQL Server Development. PE press
6. Jackson, C. (2018). Learn Programming in Python with Cody Jackson, Packt Publishing
7. Balagurusamy, E. (2017). Introduction to Computing & Problem Solving using Python, McGraw Hill Education (India) Private Limited
8. Boschetti, A. Massaron, L. (2015) - Python Data Science Essentials - Learn the fundamentals of Data Science with Python. Packt Publishing .

SEMESTER : II
COURSE CODE : STSD 221

COURSE TITLE : DISTRIBUTION THEORY

COURSE OUTCOMES

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

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

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

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

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

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

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

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

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

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

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

MODULE OUTCOMES

Sl. No:	Outcomes On completion of each module, students should be able to:	Taxonomy level
MODULE 1	MO1. Explains the basic concepts of pgf, mgf and characteristic functions MO2. Examine various types of discrete and continuous distributions and articulate their properties MO3. Examine the various properties of probability distribution. MO4. Solving problems related to several distributions	Understand Analysis Evaluate Apply

MODULE 2	MO1. Finding the joint, marginal and conditional pdf of a bivariate distribution MO2. Articulate the various concepts of Conditional expectation and conditional variance MO3. Explains the concept of pgf, mgf and independence of a bivariate random vector MO4. Articulate Multinomial and bivariate normal distributions and their properties in detail	Understand Analysis Remember Apply
MODULE 3	MO1. Apply transformation of variable technique for finding the distribution of functions of random variables and solve related problems MO2. Derive the distributions of sum, product and ratios of random variables	Apply Analysis
MODULE 4	MO1. Explain the concept of order statistics MO2. Find the joint distribution of two order statistics MO3. Calculate Distribution of functions of two order statistics	Understand Evaluate Apply
MODULE 5	MO1. Explains different sampling distributions MO2. Find expressions for probability density function of sampling distribution	Understand Apply

COURSE CONTENT

MODULE 1

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

MODULE 2

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 3

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

MODULE 4

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

MODULE 5

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

Text Books

1. Rohatgi,V.K.(1990) An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern Ltd.
2. Anderson,T.W.(1984) Introduction to Multivariate Statistical Analysis, Macmillan Publishing Company.
3. Fisz,M (1963) Probability Theory and Mathematical Statistics, 3rd Edition, John Wiley.

REFERENCES

1. Hogg,R.V. and Craig ,A.T.(1989) Introduction to Mathematical Statistics, Macmillan Publishing Company
2. Johnson ,N.L. and Kotz,S. (1969) Distributions in Statistics; Discrete distributions. John Wiley and Sons ,New York.
3. Johnson ,N.L. . Kotz,S. and Balakrishnan, N (1994)Continuous Univariate Distributions - 1 ,2nd Edition John Wiley and Sons ,New York.
4. Johnson ,N.L. . and Kotz,S. (1995) Continuous Univariate Distributions -2 ,2nd Edition, John Wiley and Sons ,New York..

SEMESTER : II

COURSE CODE : STSD 222

COURSE TITLE : DESIGN OF EXPERIMENTS AND THEORY OF SAMPLING

Objectives:

One of the aims of this course is to introduce to the students the basic concepts of survey sampling theory including brief introduction to questionnaire design, methods of sample selection, estimation, sampling variance, standard error of estimation in finite population, development of sampling theory for use in sample survey problems and sources of errors in surveys. Practical examples will be used to illustrate the principles and methods. Another aim is to let the students acquainted with the basic concepts of field experimentation. They will get a clear idea of conducting experiments under various situations where the availability of resources is constrained.

Course Outcomes

CO1: Apply the basic principles while designing a statistical experiment

CO2: Design an experiment suitable to a situation.

CO3: Perform the analysis of data coming out of experiments conducted adopting available designs.

CO4: Apply, analyze and interpret real-life survey reports from public agencies.

CO5: Apply concepts and techniques in sampling methods.

CO6: Assess the appropriateness of sampling plans with special reference to survey goals, sampling frames availability, and resource constraints.

CO7: Estimate population parameters for sampling plans.

CO8: Conduct sample surveys within the context of socially acceptable professional and ethical practices.

MODULE	**Module outcomes** *On completion of each module, students should be able to:*	**Taxonomy Level**
MODULE 1	MO1: Identify connected, balanced and orthogonal designs MO2: Design and analyze CRD, RBD, LSD and GLSD MO3:Perform Missing plot analysis in RBD and LSD MO4: Perform analysis of Covariance technique	Understand Apply Analysis Analysis
MODULE 2	MO1: Design and analyze 2^n, 3^n and p^n factorial experiments MO2: Apply the principle of total and partial confounding in factorial experiments	Analysis Analysis
MODULE 3	MO1 : Analyze BIBD with inter and intra block information MO2: Analyze PBIBD with only two associates classes MO3: Understand the concepts of Nested design MO4: Understand the concepts of split plot and strip plot designs	Analysis Analysis Understand Understand
MODULE 4	MO1: Apply various sampling designs such as simple random sampling, stratified random sampling and systematic sampling in real-life surveys MO2: Estimate various parameters and their precisions in various sampling designs along with their relative merits and demerits MO3: Estimate sample size formulae based on given margin of error and required power	Apply Analysis Apply

MODULE 5	MO1: Apply of auxiliary information for getting improved estimates of study variables via ratio and regression methods	Apply
	MO2: Apply probability proportional to size sampling methods	Apply
	MO3: Apply advanced survey methods such as multi-stage and multi-phase sampling together with the concept double sampling	Apply

Prerequisites and general awareness about Indian Statistical System (Not to be Examined)

Concept of population and sample, need for sampling, complete enumeration versus sampling,basic concepts in sampling, principle steps in sampling, variability control in sample surveys, sampling and non-sampling errors, treatment of non-sampling errors, non-response, sampling bias and non-sampling bias, measurement and control of errors, methodologies in sample surveys(questionnaires, sampling design and methods followed in field investigation) by NSSO, CSO. Office of Registrar General: historical development, main functions and publications. ANOVA- one way and two way classifications.

MODULE 1

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. Missing plot analysis in RBD and LSD, Analysis of covariance technique in standard designs.

Module 2

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

MODULE 3

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, Nested designs, Basic concepts (only) of split plot and strip plot designs.

MODULE 4

Simple randomsampling withand without replacement, estimation of population mean, population proportionsand their standard errors, sample size estimation and confidence intervals. Stratified random sampling, proportional and optimum allocation,comparison with simple random sampling for fixed sample size. Covariance and Variance Function. Systematic sampling(when population size (N) is an integer multiple of sampling size (n)). Estimation of population mean and standard error of this estimate, comparison with simple random sampling.

MODULE 5

Ratio, product and regression methods of estimation, estimation of population mean, evaluation of bias and variance to the first order of approximation, comparison with simplerandom sampling.Sampling with probability proportional to size (with and without replacement method), Des

Rajand Das estimators for n=2, Horvitz-Thomson's estimator. Equal size cluster sampling, estimators of population mean and total and their standard errors,comparison of cluster sampling with SRS, concept of multistage sampling and its application, two-stage sampling with equal number ofsecond stage units, estimation of population mean and total.

Text Books

1. Chakrabarti,M.C.(1962) : Mathematics of Design and Analysis of Experiments , Asia Publishing House, Bombay.
2. Das, M. N. and Giri, N. (1979). Design and Analysis of Experiments. Wiley Eastern Limited, New Delhi.
3. Montgomery,C.D.(1976): Design and Analysis of Experiments, John Wiley, New York.
4. Cochran, W.G. (1997): Sampling Techniques. John Wiley and Sons, New York,ISBN-13: 978-0471162384
5. Mukhopadhyay, P.(2008): Theory and Methods of Survey Sampling. Prentice Hall of India, ISBN : 9788120336766

References

1. Sukhatme, P. V., Sukhatme, B. V. and Ashok, C. (1970): Theory of Sample Surveys with Applications. Asia Publishing House, Delhi. ISBN-13 : 978-0210225196
2. Rao, P.S.R.S. (2000): Sampling Methodologies with Applications. Chapman and Hall/CRC. ISBN-13 : 978-1584882145
3. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods. Printice Hall of India ISBN-13 : 978-0137435760
4. Aloke Day (1986). Theoryof Block Designs .Wiley Eastern, New Delhi.
5. John, P.W.M. (1971). Statistical Design and Analysis of Experiments, Macmillan.
6. Joshi, D. D. (1987): Linear Estimation and Design of Experiments, Wiley Eastern, Wiley Eastern Limited, New Delhi.
7. http://mospi.nic.in/

SEMESTER : II
COURSE CODE : 223
COURSE TITLE : STATISTICAL MODELING

COURSE OUTCOMES

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

CO1: Describe simple and multiple linear regression models and its properties

CO2: Apply principle of least square method to estimate the parameters in simple and multiple

linear regression models.

CO3: Identify multicollinearity problem and its consequences.

CO4: Describe Aitken generalized least square method of estimation.

CO5: Understand Residual Analysis and residual plots.

CO6: Explain Generalized Linear models and inference on models with binary response.

Module Outcomes

Sl.No	Outcomes On completion of each module, students should be able to	Taxonomy level
MODULE 1	MO1. Explain simple linear regression model MO2. Describe least square estimators MO3. Articulate to inference regarding regression parameters MO4. Meaning of coefficient of determination.	Understand Remember Apply Analysis
MODULE 2	MO1. Explain multiple linear regression models MO2. Explain inference regarding multiple regression parameters MO3. Inference Generalised Linear Regression model,	Analysis Understand Apply
MODULE 3	MO1. Different methods of scaling residuals MO2. Residual plots MO3: Detecton of Outliers.	Understand Analysis Understand
MODULE 4	MO1. Explain Polynomial regression models MO2. Explain Indicator variables and its usage. MO3. Describe model building strategy MO4: Concept of Stepwise Regressiom. MO5: Problem of multicollinearity	Understand Understand Apply Apply Understand
MODULE 5	MO1. Explain Generalized Linear models MO2. Explain models with binary resonse. MO3. Describe link functions MO4: Inference on Poisson regression MO5: Log linear models for categorical data.	Understand Understand Apply Apply Analysis

Course Content

MODULE 1

Identification of Variables, statistical Models, Principles of statistical modeling, Model formulation. Regression models. Fitting of models, Principle of Least squares, Inference on simple linear regression models. Properties of least square estimators. Significance test and confidence intervals, prediction problems. Coefficient of determination.

MODULE 2

Multiple linear regression models, least square estimation, Properties of least square estimators, hypothesis testing on regression parameters, ANOVA, confidence estimation, prediction of new observations, Generalized Linear Regression model, The Aithken estimator, Heteroscedastic disturbances, Autocorrelation.

MODULE 3

Residual analysis- Methods of scaling residuals, Residual plots, Partial residual plots, PRESS Statistic. Detection and treatments of outliers.

MODULE 4

Polynomial regression, estimation and inference on structural parameters, Indicator variables, uses of Indicator variables, variable selection and model building strategy, All possible regressions, Stepwise regression, Problem of multicollinearity- meaning, sources and consequences.

MODULE 5

Generalized Linear models, contingency tables, binary response variables- logit models, Log linear models, Logistic regression, Estimation and testing the models, link functions, Poisson regression and its inference.

Logit models for categorical data, Goodness of fit, log linear models for categorical variables (two way contingency table)

Dose response models- quartile response, probit models and median lethal dose.

Text Books:

1. Montgomery, D.C. , Peck, E.A. and Vining, G.G. (2003). Introduction to Linear Regression Analysis, John Wiley & Sons, Asia
2. Rao, C, R. and Tutenburg, H. (1995). Linear Models , Springer Series in Statistics, New York
3. Dobson, A.J. (2002) An Introduction to Generalized Linear models, Second edition, CRC Press.

References:

1. Mc Cullagh, P. and Nelder, J.A. (1989). Generalized Linear Models, Chapman and Hall.
2. Neter, J. and Wasserman, D.W. (1983). Applied Linear Statistical Models, Richard, D. Irwin, Inc., Illinois.
3. Rao, C.R. (1973). Linear Statistical Inference and its Applications, Wiley, New York.
4. Draper, N.R. and Smith, R. (2003). Applied Regression Analysis, John Wiley and Sons inc., New York
5. Seber, G.A.F. (1977). Linear Regression Analysis, John Wiley and Sons, New York

SEMESTER: II
COURSE CODE: STSD 224

COURSE TITLE: PRACTICAL I (USING R AND PYTHON)

(Problems covering papers STSD 221, STSD 222 and STSD 223)

SEMESTER: III
COURSE CODE: STSD 231

COURSE TITLE: MACHINE LEARNING

COURSE OUTCOMES

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

CO1. Analyze the data and understand the insights from it.

CO2. A clear understanding of machine learning algorithms and applications.

CO3. Apply various ML techniques as per the requirements

CO4. Calculate the maximum likelihood estimation CO5.

Differentiate bias and variance estimation

CO6. To calculate the number of clusters for clustering algorithms

CO7. The full derivation of forward and backward propagation in neural network

MODULE OUTCOMES

Sl. No:	**Outcomes** **On completion of each module, students should be able to:**	**Taxonomy Level**
MODULE 1.	MO1. Explain about machine learning and applications MO2. Thorough understanding about supervised, semi-supervised and unsupervised learning algorithms	Understand Understand
MODULE 2.	MO1: Introduction of estimation methods MO2: To get an idea about the inference of the data (population) MO3: Use of different types of classifications in data	Understand Understand Apply

MODULE 3.	MO1. Explain about various clustering algorithms MO2. Introduction to data mining MO3. Apply the concepts of clustering algorithms into different types of data. MO4. Apply data reduction techniques	Understand Understand Apply Apply
MODULE 4.	MO1. A better understanding of SVM MO2.Apply SVM algorithms for face detection MO3. Analysis using data validation techniques MO4. Explain about data warehouse MO5. Various applications of data warehouse	Understand Apply Evaluate Understand Understand
MODULE 5.	MO1: Good knowledge in both Multivariate analysis and missing data MO2: To get the understanding of the relationship between more than two variables through multivariate analysis	Apply Apply

COURSE CONTENT

MODULE 1

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 2

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 3

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 4

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 5

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.

Text Books:

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. Gupta, G.K. (2008): Introduction to Data Mining with case studies, Prentice – Hall of India Pvt. Ltd.

REFERENCES

1. Bhat, B. R. (1985). Modern Probability Theory: An Introductory Text Book, 2nd Edition, Wiley Eastern.
2. Brian Coffo. Statistical Inference for Data Science.
3. Tan, T., Steinbach, M. and Kumar, V. (2006): Introduction to Data Mining, Pearson Education.
4. Daniel T. Larose (2006): Data Mining: Methods and Models, John Wiley and sons. (relevant portions of Chapter 4).

SEMESTER: III
COURSE CODE: STSD 232

COURSE TITLE: STATISTICAL INFERENCE

Course Outcomes

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

CO 1: List the ideal properties of point estimators of an unknown parameter of a distribution and select the best estimators using different properties.

CO 2: Derive the UMVUE of a parameter or function of a parameter.

CO 3: Determine estimators of unknown parameters using methods like MLE, Method of moments etc.

CO 4: Construct confidence intervals for an unknown parameter.

CO 5: Formulate hypothesis for a given problem.

CO 6: Find most powerful test for testing simple hypothesis against simple alternatives.

CO 7: Find UMP test for testing composite hypothesis.

CO 8: Derive likelihood ratio test for testing the hypothesis for normal populations.

CO 9: Obtain sequential probability ratio test for testing the hypothesis.

CO 10: Perform a suitable non-parametric test for a given data.

MODULE OUTCOMES

Sl.No.	Outcomes On Completion of each module, Students should be able to:	Taxonomy level
MODULE 1	MO1 Know whether an estimator is unbiased . MO2 Know whether an estimator is consistent. MO3 Identify Efficient estimators. MO4 Determine the sufficient statistic using factorization theorem. MO5 Construct UMA confidence intervals for an unknown parameter.	Evaluate Evaluate Evaluate Create Apply
MODULE 2	MO1 Explain minimal sufficiency. MO2 Apply Rao-Blackwell and Lehmann-Scheffe theorems to find UMVUE. MO3 Calculate Cramer-Rao inequality and Chapman -Robbin's bound. MO4 Use likelihood equivalence to obtain minimal sufficient statistic. MO5 Apply different methods of estimation such as method of moments, MLE, Minimum chi-square, Modified minimum chi-square etc.	Understand Apply Apply Evaluate Understand Apply Apply
MODULE 3	MO1 Identify simple and composite hypothesis. MO2 Find critical region, size and power of the test. MO3 Distinguish between randomized and non-randomized test. MO4 Apply Neyman-Pearson lemma to find most powerful test. MO5 Find UMP and UMPU test.	Understand Evaluate Identify Apply Apply
MODULE 4	MO1 Apply likelihood ratio test principle for testing the mean for a normal population. MO2 Apply likelihood ratio test principle for testing the equality of means for two normal population. MO3 Obtain sequential probability ratio test for testing the hypothesis. MO4 Apply ANOVA for testing the equality of means of several normal populations.	Apply Apply Apply Apply

MODULE 5	MO1 Perform one sample non-parametric test. MO2 Perform two sample non-parametric test. MO3 Perform Kruskal-Wallis one-way test . MO4 Perform Friedman two-way test.	Apply Apply Apply Apply

COURSE CONTENT

MODULE 1

Point estimation: unbiasedness, consistency (strong, weak and squared error), asymptotically unbiased estimator, bias and mean square error, BLUE , Efficient estimator. Sufficient condition for weak consistency. Invariance property of consistent estimator, Sufficiency and completeness. Complete sufficient statistic. Factorization criterion for sufficiency (without proof). Interval estimation: UMA confidence sets. Methods of finding confidence intervals using pivot.

MODULE 2

Construction of minimal sufficient statistic (example). UMVU estimators. Rao-Blackwell theorem (proof not required), Lehmann-Scheffe theorem (proof not required) and its applications. Ancilliary statistic and Basu's theorem (statement and examples). Fisher's information measure- (i) in a random variable (ii) in a random sample (iii) in a statistic (concept only), Cramer- Rao inequality (proof not required), Chapman-Robin bound (statement only).Methods of estimation- maximum likelihood, method of moments, method of minimum Chi-square and modified minimum Chi-square.

MODULE 3

Fundamental concepts of hypothesis testing: Hypotheses- Null and alternative, simple and composite, problem of testing of hypothesis, critical region, two kinds of errors, level of significance, test function, randomized and non- randomized tests, size and power function of a test, Test of a simple hypothesis: Neyman-Pearson Lemma (proof not required), Generalization of Neyman-Pearson Lemma (Statement and applications only), MP and UMP tests. Test of a composite hypothesis: Family of distributions with monotone likelihood ratio, UMP test for one sided hypothesis concerning a real valued parameter, UMP tests for two sided hypothesis in case of one parameter exponential family. Unbiased tests, UMPU test and LMP tests (concepts only).

MODULE 4

Likelihood ratio test (LRT), asymptotic properties; LRT for the parameters of normal distribution. Connection between tests of hypothesis and confidence sets. Sequential methods of testing: Sequential Probability Ratio Test (SPRT), Wald's fundamental identity, OC and ASN functions (concepts only), Parametric tests (problems only): One sample test for mean, two sample tests for means, ANOVA (one way and two way).

MODULE 5

Non parametric tests : Chi-square goodness of fit test, Kolmogorov-Smirnov test (one sample and two sample tests), Sign test, Wilcoxon signed rank test, run test, Wald-Wolfowitz run test, median test, Mann-Whitney Wilcoxon test; Tests for association based on Kendall's Tau and Spearman's rank correlation coefficient, Kruskal-Wallis test and Friedman test.

TEXT BOOKS

1. Rohatgi, V. K. and Saleh, A.K.M (2003). An Introduction to Probability and Statistics, John Wiley and Sons.
2. Mukhopadhyay, P. (2006). Mathematical Statistics, Books and Allied (P) Ltd., Kolkatta.
3. Rajagopalan, M. and Dhanavanthan, P. (2012). Statistical Inference, PHI Learning Pvt Ltd, New Delhi.
4. Gibbons, J.D. (1985): Non Parametric Statistical Inference, 2nd edn. Marcel Dekker Inc

REFERENCES

- Hogg, R. V. and Craig, A. T. (1989): Introduction to Mathematical Statistics, Macmillan Publishing Company.
- Kale, B. K. (1999): A First Course on Parametric Inference, Narosa Publishing House.
- Lehmann, E. L. (1983): Theory of Point Estimation, John Wiley, New York.
- Mood, A. M., Graybill, F. A. and Boes, D. C.(1972): Introduction to the Theory of
- Statistics, 3rd Edition, Mc-Graw Hill International, New York.
- Ferguson, T. S. (1967): Mathematical Statistics, Academic Press, New York.
- Lehmann, E. L. (1986): Testing of Statistical Hypothesis, John Wiley & Sons.
- Wald, A.(1977): Sequential Analysis, Dover Publications Inc., New York.
- Wetherill, G. B. (1966): Sequential Methods in Statistics, Methuen & Co. Ltd. New Delhi.
- Siegel, S. and Castellan Jr. N. J. (1988): Non-parametric Statistics for the Behavioral Sciences. McGraw Hill, New York.
- Randles, R.H. and Wolfe, D.A. (1979): Introduction to the Theory of Nonparametric statistics, Wiley New York.

SEMESTER III
COURSE CODE : STSD 233(i)

COURSE TITLE : STOCHASTIC FINANCE

COURSE OUTCOMES

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

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

CO2. Explains the basic concepts of financial markets

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

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

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

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
MODULE 1	MO1. Explains the basic concepts of financial markets MO2. Explain the concepts of forward contracts, future contracts, options call and put options, European option and American options, hedgers, speculators, arbitrageurs, MO3. Define the interest rates, compounding, present value analysis, risk free interest rates MO4. Solving problems related to interest rate, compounding, present value analysis, risk free interest rates	Understand Analysis Evaluate 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). MO2. State Arbitrage theorem and The Black-Scholes formula and also describe the Properties of the Black-Scholes option cost.	Apply Analysis
MODULE 4	MO1. Estimating the volatility parameter. MO2. Describe the pricing American options, pricing of European options using Monte Carlo and pricing an American option using finite difference methods.	Understand Evaluate

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

COURSE CONTENT

MODULE 1

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

MODULE 2

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 3

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 4

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 5

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

TEXT BOOKS

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

REFERENCES

1. John C Hull (2008). Options, Futures and other Derivatives, Pearson Education, India
2. Masaaki Kijima (2003). Stochastic Process with Applications to Finance, Chapman and Hall
3. Ruey S Tsay (2005). Analysis of Time Series, Third Ed., John Wiley and Sons.
4. Cuthbertston K and Nitzsche D (2001). Financial Engineering- Derivatives and Risk Management , John Wiley and Sons

SEMESTER: III
COURSE CODE: STSD 233(i)

COURSE TITLE: ORDER STATISTICS

COURSE OUTCOMES

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

CO1. Understand the basic properties of Order statistics.

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

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

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

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

CO6. Explain the estimation and prediction under Order statistics.

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

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

MODULE OUTCOMES

Sl. No:	Outcomes On completion of each module, students should be able to:	Taxonomy level
Module 1	MO1. Explains the basic concepts of distribution of single order statistic, joint distribution of two or more order statistics. MO2. Verify the Markov property of Order statistics. MO3. Examine the various properties of Order statistics. MO4. . Explain the concept of order statistics	Understand Remember Understand Understand
Module 2	MO1. Explain the moments of order statistic and its properties. MO2. Derive the recurrence relations on the single and product moments of order statistics. MO3. Describe Order statistics from symmetric population.	Understand Remember Understand
Module 3	MO1. Examine order statics of various types of discrete and continuous distributions and articulate their properties	Apply
Module 4	MO1. Explain the different estimators used in order statistics MO2.Examine prediction of order statistics. MO3. Find the confidence interval using sample quantile.	Evaluate Apply Understand
Module 5	MO1. Explains the concept of Record value and its application. MO2. Explains the concept of Concomitant Order statistics and application	Understand Understand

COURSE CONTENT

MODULE 1

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

MODULE 2

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

MODULE 3

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

MODULE 4

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

MODULE 5

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

Text Books

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

SEMESTER III
COURSE CODE : STSD 233(ii)

COURSE TITLE : MULTIVARIATE ANALYSIS

COURSE OUTCOME

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

CO1: Describe multivariate normal distribution and its properties

CO2: Find the marginal and conditional distribution of multivariate normal distribution

CO3: Find the distribution of quadratic forms of multivariate normal vectors

CO4: Describe Wishart distribution and its properties

CO5: Obtain the estimators for parameters of multivariate normal distribution

CO6: Describe multiple and partial correlation coefficients

CO7: Define sample multiple and sample partial correlation coefficients for multivariate normal vector

CO8: Test the hypothesis regarding parameters of multivariate normal distribution

CO9: Use Hotelling's T^2 and Mahalanobis D^2 statistics for testing hypothesis CO10: Perform Multivaraite data analysis

MODULE OUTCOME

MODULE	**Module outcomes** *On completion of each module, students should be able to:*	**Taxonomy Level**
MODULE 1	MO1: Define the multivariate normal density function. MO2: Obtain the characteristic function of multivariate normal density MO3: Find the distribution of linear combination of multivariate normal random vector using characteristic function MO4: Find regression of multivariate normal distribution using conditional distribution MO5: Characterize quadratic forms of multivariate distribution. MO6: Obtain the distribution of sums and quotients of quadratic forms. MO7: Apply Cocharan's theorem to find distribution of quadratic forms of multivariate normal random vector	Remember Understand Apply Apply Apply Apply Apply
MODULE 2	MO1: Obtain the MLEs of mean and variance of multivariate normal distribution MO2 : Find the characteristic function of Wishart distribution MO3 : Show that Whishart distribution possess additive property MO4 : Find the distribution of sample dispersion matrix	Evaluate Apply Understand Remember
MODULE 3	MO1: Find the distribution of sample multiple correlation for multivariate normal distribution MO2: Find the distribution of sample partial correlation coefficients for multivariate normal distribution MO3: Obtain the distribution of test statistic for the test the significance of correlation coefficient and partial correlation coefficient.	Remember Apply Apply

MODULE 4	MO1: Test the mean vector of a multivariate normal distribution MO2: Test the equality of means of two or more multivariate normal distributions MO3: Use Hotelling's T^2 and Mahalanobis D^2 statistics in testing hypothesis regarding multivariate normal distributions. MO4: Find the relationship between Hotelling's T^2 and Mahalanobis D^2 statistics	Apply Apply Apply Understand
MODULE 5	MO1: Perform principal component analysis and factor analysis MO2: Classify individuals/items in to one of k multivariate normal populations MO3: Identify canonical variables and quantify canonical correlation	Analysis Analysis Analysis

COURSE OUTCOME

MODULE 1

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

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

MODULE 2

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 3

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 4

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 5

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.

TEXT BOOKS

1. Anderson, T.W. (2003) :An Introduction to Multivariate Statistical Analysis, John Wiley, New York.
2. Muirhead, R.J. (1982): Aspects of Multivariate Statistical Theory, John Wiley & Sons New York
3. Johnson, R and Wychern (1992): Applied Multivariate Statistical Analysis, Prentice hall, London.

REFERENCES

1. Kendall, M. G. (1958) :A Course in Multivariate Analysis, Griffin, London.
2. Khatri, C.G. and Srivastava (1979) : An Introduction to Multivariate Statistics, North – Holland, New York.
3. Muirhead, R.J. (1982): Aspects of Multivariate Statistical Theory, John Wiley & Sons New York.

SEMESTER III

COURSE CODE : STSD 233(iii)

COURSE TITLE : DATA MINING AND THEIR APPLICATIONS

COURSE OUTCOMES

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

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

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

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

CO4. Apply statistical methods for any given raw data.

MODULE OUTCOMES

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

COURSE CONTENT

MODULE 1

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

MODULE 2

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

MODULE 3

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

MODULE 4

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

MODULE 5

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

TEXT BOOKS

1. Box, G.E.P., Jenkins, G.M. and Reinsel, G.C (2004). Time Series Analysis-Forecasting and Control, Pearson Education, Singapore.
2. Daniel T. Larose, (2006): Data Mining: Methods and Models, Wiley-Interscience, New Jersey.
3. Hawkins, D.M, (1980): Identification of Outliers, Chapman and Hall, London.
4. Jiawei Han, Micheline Kamber, (2006): Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, second edition, San Francisco.

REFERENCES

1. Krzysztof J.Cios, Wiltold Pedrycz, Roman W.Swiniarski, Lukasz A.Kurgan, (2007): Data Mining: A Knowledge Discovery Approach, Springer Science +Business Media, New York.
2. Montgomery, D.C. and Johnson, L.A. (1977) Forecasting and Time Series Analysis, McGraw Hill, New York.
3. Paolo Giudici, (2005): Applied Data Mining: Statistical Methods for Business and Industry, John Wiley & Sons Ltd, England.
4. Peter J.Rousseeuw and Annick M.Lorey, (1987): Robust Regression and Outlier Detection, John Wiley & Sons, United States.
5. Vic Barnett and Toby Lewis, (1978): Outliers in Statistical Data, John Wiley & sons.

SEMESTER : III

COURSE CODE : STSD 233(iv)

COURSE TITLE : CATEGORICAL DATA ANALYSIS

COURSE OUTCOMES

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

CO1. Analyze categorical data and make conclusion from it.

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

CO3. Model Binary and multinomial response variables.

CO4. Explain logistic regression and its analysis.

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

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
MODULE 1	MO1. How to model binary response variables. MO2. Explain logit model.	Understand Analysis
MODULE 2	MO1. Model multinomial response variables. MO2. Analyse repeated categorical response data.	Understand Analysis
MODULE 3	MO1. Analyse two way and three way contingency tables.	Understand
	MO2. Model data with repeated measurements. MO3. Explain logistic regression.	Analysis Remember
MODULE 4	MO1. Tests for independence and homogeneity of proportions for contingency tables. MO2. Explain Fishers exact test. MO3. Understand Simpson's Paradox.	Understand Analysis Understand
MODULE 5	MO1. Explain Polytomous logit models. MO2. How to model for multiway tables.	Understand Analysis

COURSE CONTENT

MODULE 1
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 2
Multinomial Response Models - Models for Matched Pairs- Analyzing Repeated Categorical Response Data - Asymptotic Theory for Parametric Models - Estimation Theory for Parametric Models.

MODULE 3
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 4

Introduction to contingency tables: 2 × 2and r × 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 5

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.

TEXT BOOKS

1. Agresti, Alan (1996). An Introduction to Categorical Data Analysis, Wiley.Breiman, L.
2. Friedman, J.H. Olshen, R.A. and Stone, C.J. (1984): Classification and Regression Trees.
3. Wadsworth and Brooks/Cole.
4. Fienberg, S.E. (1980). The Analysis of Cross-Classified Categorical Data. MIT Press.

REFERENCES

1. Bergsma, W., Croon, M.A. and Hagenaars, J.A. (2009). Marginal Models: ForDependent, Clustered, and Longitudinal Categorical Data. Springer.
2. Bishop, Y.M., Fienberg, S.E. and Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice, MIT Press.
3. Wasserman, L. (2004). All of Statistics: A Concise Course in Statistical Inference, Springer.
4. Whittaker, J. (1990). Graphical Models in Applied Multivariate Statistics. Wiley.

SEMESTER : III

COURSE CODE : STSD 233(v)

COURSE TITLE : LONGITUDINAL DATA ANALYSIS

COURSE OUTCOMES

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

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

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

MODULE OUTCOMES

Sl. No:	Outcomes On the completion of each module, students should be able to:	Taxonomy level
MODULE 1	MO1. Explain and articulate the background of linear, count and binary regression analysis. MO2. Exemplify the various concepts on Linear, log-linear and logistic regression models for continuous, count and binary data. MO3. Explain the different methods of estimation and comment on the asymptotic properties of these estimators for regression effects.	Understand Remember Understand
MODULE 2	MO1. Exemplify the concept of longitudinal/panel data. MO2. Illustrate the different types of longitudinal data with examples.	Remember Apply
MODULE 3	MO1. Explain how to fit a binary dynamic logits (BDL) model to the longitudinal stationary/non-stationary binary data. MO2. Distinguish between binary dynamic logits (BDL) and binary auto-correlation models. MO3. Comment of the concept of correlation structure under the BDL model. MO4. Exemplify the concept of GQL and likelihood estimation. Also, illustrate and estimate the model using asthma data.	Apply Evaluate Analysis 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. MO2. Explain the AR(1) type mixed effect models for count and binary data. MO3. Exemplify the method of GQL estimation for regression and random effects variance parameters and illustrate it with suitable data.	Remember Evaluate Analysis

COURSE CONTENT

MODULE 1

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 2

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 3

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 4

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 5

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.

TEXT BOOKS

1. Longitudinal Categorical Data Analysis (2014) by B. C. Sutradhar, Springer.
2. Analysis of Longitudinal Data (2002) by P. Diggle, P. Heagerty, K-Y Liang, and S. L. Zeger, 2nd edition, Oxford University Press, Oxford.

REFERENCES

1. Applying Generalized Linear Models (1997) by J. K. Lindsey, Springer.
2. Generalized Linear Models (1989) by P. McCullagh and J. A. Nelder, Chapman and Hall.

SEMESTER : III

COURSE CODE : STSD 233(vi)

COURSE TITLE : OPERATIONS RESEARCH

COURSE OUTCOMES

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

CO1. Describe Simplex method to solve the linear programming problem.

CO2. Explain the steps in solving a linear programming problem by two-phase method.

CO3. Explain the concept of duality in linear programming problem.

CO4. Give the outline of dual simplex method.

CO5. Describe the computational procedure of optimality test in a transportation table.

CO6. Explain the Hungarian method to solve the Assignment problem.

CO7. To understand basic structure of quadratic programming problem.

CO8. Give an account of different types of inventory models and inventory cost.

CO9. Derive an EOQ formula for different rate of demand in different cycles.

CO10. Formulate and solve the purchase inventory problem with one price break.

CO11. Derive the steady state solution of M/M/1 queue model.

CO12. Obtain expected number of units in the M/G/1 queueing system under steady state .

CO13. Derive an expression of the average annual cost of an item over a period of n years.

CO14. Describe Bellmen's principle of optimality.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy Level**
MODULE 1.	MO1. Explain the concepts of linear programming problem. MO2. Solve the linear programming problem by using Simplex method. MO3. Computational steps of Big-m-Method MO4. Write the dual of the given linear programming problem.	Understand Apply Remember Evaluate

MODULE: 2	MO1. Find the initial basic feasible solution to the given Transportation problem. MO2. To determine the optimum assignment problem. MO3. State and prove Kuhn-Tucker necessary and sufficient conditions in a non-linear programming problem. MO4. Describe briefly the Beal's method for solving quadratic programming problem.	Evaluate Evaluate Remember Remember
MODULE 3	MO1. Explain with examples the probabilistic models in inventory. MO2. Derive an expression for EOQ for Harri's or Wilson model. MO3. Discuss the purchase inventory model with three price break.	Understand Understand Remember
MODULE 4	MO1. State some of the important distributions of arrival interval and service time. MO2. To evaluate expected value of queue length in various models. MO3. Obtain the differential difference for M/M/1 queue model.	Remember Evaluate Understand
MODULE 5	MO1. Describe the problem of replacement of items whose maintains cost increase with time. MO2. Solve Dynamic programming problem by using Linear programming problem.	Understand Apply

COURSE CONTENT

MODULE 1

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

MODULE 2

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

MODULE 3

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

MODULE 4

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 5

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

TEXT BOOKS

1. Kanti Swarup, Manmohan and Gupta, M.M. (1985): Operations Research, Sultan Chand & Sons.
2. Ravindran, A, Philips, D.T. and Soleberg, J.J. (1997): Operations Research – Principles and **i.** Practise
3. S, D, Sharma(2012) : Operations Research Theory Methods & Applications.
4. Taha, H.A. (1997): Operations Research, Mc.Millian.

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. Mittal, K.V. (1990): Optimization Methods.
5. Saaty, T.L. (1961): Elements of Queuing Theory with Applications, Mc Graw Hill.

SEMESTER: III
COURSE CODE: STSD 234

COURSE TITLE: PRACTICAL II (USING R AND PYTHON)

(Problems covering papers STSD 231, STSD 232 and STSD 233)

SEMESTER: IV
COURSE CODE: STSD 241

COURSE TITLE: STOCHASTIC PROCESSES AND TIME SERIES

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, martingales, Wiener process, Gaussian process.

CO2. Explain Markov chains: Definition, transition probability matrix, n-step transition Probability and Chapman-Kolmogorov equation, Calculate n-step transition probabilities, Classify states of a finite Markov chain .

CO3. Describe periodicity and ergodicity of chains, Describe limiting beha vior of n-step transition probabilities, obtain the stationary distribution of a Markov chain

CO4. Explain and exemplify continuous time Markov chain, Poisson process, pure birth process, birth and death processes, compound Poisson process, Markov Process with discrete states.

CO5. Explain and exemplify renewal processes, renewal equation. Describe and apply renewal theorem.

CO6. Describe Branching processes, offspring distribution, extinction probabilities

CO7. Distinguish between strict and weak (covariance or wide sense) stationarity,

CO8. Predict the time series data using different stationary time series data

MODULE OUTCOMES

Sl. No.	Outcomes On completion of each module, students should be able to:	Taxonomy Level
MODULE 1.	MO1. Articulate and exemplify the concepts of Stochastic processes, time space and state space. MO2. Construction of examples of Stochastic processes MO3. Explain the concepts of particular types of stochastic processes like process with stationary independent increments, Markov process, martingales, Wiener process, Gaussian process etc. MO4. Articulate concepts of Markov chains, transition probability matrix, n-step transition probabilities MO5. Calculate n-step transition probabilities MO6. Describe and exemplify classification of states in a Markov Chain MO7. Calculate the periodicity of a Markov Chain MO8. Explain the concepts of recurrence, ergodic chains MO9. Explain and exemplify concepts of limiting behaviour of n-step transition probabilities. MO10.Describe stationary distributions and solve problems	Understand Understand Create, Apply Understand Evaluation Understand Evaluation Understand Understand Apply
MODULE 2.	MO1. Describe and exemplify: Continuous time Markov chains, Poisson process, pure birth process, birth and death processes. MO2. Derive of steady state probabilities/differential difference equations in case of Poisson process, pure	Understand

	birth process, birth and death processes. MO3. Describe and exemplify: Compound Poisson process, MO4 Derive properties of Poisson process and Compound Poisson process MO5. Explain the concept of Markov Process with discrete states. MO6. Illustrate these processes with examples	Evaluation Understand Evaluation Understand Apply
MODULE 3.	MO1. Describe and exemplify: renewal processes, renewal equation. MO2. Explain the statement and applications of renewal theorem MO3. Solve problems based on the applications of renewal theorem MO4. Concepts of stopping time, Wald's equation, residual and excess life times, backward and forward recurrence times MO5. Poisson process as a renewal process MO6. Describe and exemplify Galton-Watson branching processes MO7. Explain concepts of offspring distribution and its implications MO8. Interpret the concept of extinction probabilities MO9. Compute the probability extinction in case of a particular offspring distribution.	Understand Apply Apply Understand Analyse Understand Apply Apply Evaluation
MODULE 4.	MO1. Explain the Stationary time series, Autocorrelation, partial auto correlation function MO2. Develop Linear stationary models: auto regressive, moving average and mixed processes. MO3.Apply Linear non-stationary models- Autoregressive integrated moving average (ARIMA) models	Understand Create Apply
MODULE 5.	MO1 Forecasting using ARMA and ARIMA models: MMSE methods MO2. Forecasting using different forms of the ARIMA models	Apply Apply

COURSE CONTENT

MODULE 1

Introduction to stochastic processes - time and state space, classification of stochastic processes, processes with independent increments, Gaussian process, Martingales, Markov process, random walk and Wiener process (examples). Markov chain, transition probabilities and stationary transition probabilities, transition probability matrix, Chapman - Kolmogorov equation: classification of states, first passage time distribution, stationary distribution, irreducible Markov chain, aperiodic chain, ergodic theorem and Gamblers ruin problem.

MODULE 2

Poisson process - Properties of Poisson process and related distributions, compound Poisson process, pure birth process, birth immigration process, time dependent Poisson process, pure death process and birth and death process.

MODULE 3

Renewal process - definition and examples, renewal function and renewal density, renewal equation, statement and applications of renewal theorems, stopping time, Wald's equation, residual and excess life times, backward and forward recurrence times, Poisson process as a renewal process, branching process - definitions and examples, generating function of branching process, Galton - Watson branching process, probability of ultimate extinction, distribution of total number of progeny.

MODULE 4

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

MODULE 5

Forecasting using ARMA and ARIMA models: MMSE methods, using different forms of the ARIMA models

Text books:

1. Karlin, S. and Taylor, H.M. (1975). A First Course in Stochastic Processes, Academic Press.
2. Medhi, J. (2009). Stochastic Processes, New Age International Publishers, New Delhi.
3. Box, G.E.P., Jenkins G.M. and Reinsel, G.C. (2007) Time Series Analysis, Forecasting and Control, Pearson Education.
4. Brockwell,P J. and David R. A. (2002). Introduction to time series and forecasting, 2nd edition, Springer.

References:

1. Bhat, U.N. (1972). Elements of Applied Stochastic Processes, John Wiley, New York
2. Cinlar, E. (1975). Introduction to Stochastic Processes, Prentice Hall, Inc, New York.

3. Makridakis, S and Wheelwright, S C. Forecasting methods and applications, John Wiley and Sons
4. Feller, W. (1968). Introduction to Probability Theory and Applications, Vol. I, John Wiley, New York
5. Feller, W. (1971). Introduction to Probability Theory and Applications, Vol. II, John Wiley, New York

SEMESTER: IV
COURSE CODE: STSD 242

COURSE TITLE: BIG DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

COURSE OUTCOMES

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

CO1. Understand when a data becomes Big Data

CO2. Choose appropriate technology for processing Big Data problems

CO3. store and scale large fast growing of Data

CO4. Build architecture for solving complex data

CO5. Get good understanding of Hadoop and HDFS Ecosystem CO3.

Can implement Distributed computing to process large data

MODULE OUTCOMES

Sl. No:	**Outcomes** **On completion of each module, students should be able to:**	**Taxonomy Level**
Module 1.	MO1. Introduction to Big Data MO2. Understanding the properties of Big Data MO3. Explanation of HDFS and Map Reduce MO4. Setup a Hadoop Cluster	Understand Analysis Remember Apply
Module 2.	MO1. Deep understanding of HDFS architecture MO2. Understanding Hadoop YARN, MO3. Describe Hadoop Ecosystems and components	Remember Understand Analysis

Module 3.	MO1. Understand distributed computation with map reduce MO2. Explore mapreduce configuration MO3. Store in nosql database using Hbase MO4. Implement MapReduce	Understand Analysis Understand Apply
Module 4.	MO1. Understanding the architecture of YARN	Understand
Module 5.	MO1. Building machine learning models using mahout MO2. Understand how Machine learning is applied on large scale data MO3. What is Nosql and when to use Nosql	Apply Understand Apply

COURSE CONTENT

MODULE 1

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

MODULE 2

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

MODULE 3

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 4

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 5

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

Text Books

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. Acharjya, D.P. , Sachidananda, D. And Sugata, S (2016): Computational Intelligence for Big

Data Analysis: Frontier Advances and Applications. Sprmger

References

1. D, Cielen, Arno, D B Meysman and M Ali (2016) :Introducing Data Science. Dreamlech
2. F, Flach.(2015): Machine Learning. Cambridge University Press.
3. J, Kumar and N, S. Gill(2020). Artificial Intelligence and Deep Learning for Decision Makers, BPB Publications.

SEMESTER IV

COURSE CODE : STSD243(i)

COURSE TITLE : RELIABILITY MODELING AND STATISTICAL QUALITY CONTROL

COURSE OUTCOMES

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

.

CO1. Define reliability including the different types and how they assessed.

CO2. Ensure the validity and precision of statistical analysis.

CO3. Explains reliability in discrete set up.

CO4. Understand the concept of quality control statistical process control.

CO5. Measure the performance of a process.

CO6. Identify assignable causes

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
MODULE 1	MO1. Explain basic concepts of reliability MO2. Describe reliability of a coherent system MO3. Describe several measures of reliability MO4. Calculate measures of reliability based on several distributions	Apply Analysis Understand Apply
MODULE 2	MO1. Explain reliability in discrete set up MO2. Derive expression of the relation connecting various	Understand Evaluate

	measures of reliability	
MODULE 3	MO1. Explain inference in reliability models MO2. Describe estimation of parameters based on censored sampling MO3. Solve problems related to estimation of parameters based on censored sampling	Understand Apply Apply
MODULE 4	MO1. Explain statistical process control MO2. Describe various control charts	Understand Analysis
MODULE 5	MO1. Describe different sampling inspection techniques MO2. Explains six-sigma concepts MO3. Distinguish chain sampling and continuous sampling	Evaluate Analysis Remember

COURSE CONTENT

MODULE 1

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 2

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

MODULE 3

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.

MODULE 4

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

MODULE 5

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

TEXT BOOKS

1. Barlow, R.E. and Proschan, F. (1985): Statistical Theory of Reliability and Life Testing, Holt, Rinehart and Winston.
2. Montgomery, D.C. (2005): Introduction to Statistical Quality Control, 5th edition, John Wiley.
3. Cox, D.R. and Oakes, D. (1984): Analysis of Survival Data, Chappman Hall.
4. Duncan, A. J. (1959): Quality Control and Industrial Statistics (5th edition), Irwin, Homewood I.

REFERENCES

1. Galambos, J. and Kotz, S. (1978) Characterization of Probability Distributions.
2. Klefjo, B. (1982) The HNBUE and HNWUE Classes of Life distributions, Naval Research Logistic Quarterly, 29, 331-344.
3. Lawless, J. F. (2003): Statistical Models and Methods for Lifetime Data, John Wiley.
4. Nelson, W. (1982): Applied life data analysis, Wiley.
5. Sinha, S. K. (1986) Reliability and Life Testing, Wiley.

SEMESTER : IV
COURSE CODE : STSD 243(ii)

COURSE TITLE : OFFICIAL STATISTICS

COURSE OUTCOME

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

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

CO2. Understand population and related issues.

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

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

CO5. Explain the procedure of national income estimation.

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

MODULE OUTCOMES

Sl. No:	Outcomes On completion of each module, students should be able to:	Taxonomy level
MODULE 1	MO1. A general understanding of the official statistics System that prevails in our country and others. MO2. Responsibilities of various central and state divisions under the Indian official statistics system.	Understand Evaluate

	MO3. Dissemination and use of data collected.	Apply
MODULE 2	MO1. Learn about population and related issues. MO2. Learn various health and family welfare schemes. MO3. Understand in detail the Population Census in India.	Understand Analysis Remember
MODULE 3	MO1. Understand various other social statistics. MO2. Understand various index numbers used in the official statistics system.	Understand Analysis
MODULE 4	MO1. Learn the economic aspects of growth and development. MO2.Estimation of national income through various approaches.	Understand Apply
MODULE 5	MO1. Understand various income inequality measures. MO2. Understand various poverty measurement methods.	Understand Apply

.

COURSE CONTENT

MODULE 1

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 2

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 3

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

MODULE 4

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 5

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.

Text Books

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. ublications.

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 : STSD 243(iii)

COURSE TITLE : STOCHASTIC FINANCE

COURSE OUTCOMES

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

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

CO2. Explains the basic concepts of financial markets

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

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

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

MODULE OUTCOMES

Sl. No:	Outcomes On completion of each module, students should be able to:	Taxonomy level
MODULE 1	MO1. Explains the basic concepts of financial markets MO2. Explain the concepts of forward contracts, future contracts, options call and put options, European option and American options, hedgers, speculators, arbitrageurs, MO3. Define the interest rates, compounding, present value analysis, risk free interest rates MO4. Solving problems related to interest rate, compounding, present value analysis, risk free interest rates	Understand Analysis Evaluate 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). MO2. State Arbitrage theorem and The Black-Scholes formula and also describe the Properties of the Black-Scholes option cost.	Apply Analysis
MODULE 4	MO1. Estimating the volatility parameter. MO2. Describe the pricing American options, pricing of European options using Monte Carlo and pricing an American option using finite difference methods.	Understand Evaluate
MODULE 5	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. MO2. Examine the ARCH(1), ARCH(p), GARCH(p,q) models and their estimation. MO3.Distinguish between ARMA and GARCH processes	Understand Apply Analysis

COURSE CONTENT

MODULE 1

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

MODULE 2

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 3

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 4

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 5

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

TEXT BOOKS

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

REFERENCES

1. John C Hull(2008). Options, Futures and other Derivatives, Pearson Education, India
2. Masaaki Kijima (2003). Stochastic Process with Applications to Finance, Chapman and Hall
3. Ruey S Tsay (2005). Analysis of Time Series, Third Ed., John Wiley and Sons.

4. Cuthbertston K and Nitzsche D (2001). Financial Engineering- Derivatives and Risk Management , John Wiley and Sons

SEMESTER IV
COURSE CODE : STSD0243(iv)

COURSE TITLE : SPATIAL DATA ANALYTICS

COURSE OUTCOMES

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

CO1. Analyze spatial temporal data.

CO2. Explain domain specific data analysis and its applications.

CO3. Handle heterogeneous data.

CO4. Distinguish between Quadrant and nearest neighbor methods.

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

CO6. Interpret scatter plots.

CO7. Construct Estimation Criteria.

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

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

CO10. Find the applications of point process models.

MODULE OUTCOMES

Sl. No:	**Outcomes** On completion of each module, students should be able to:	**Taxonomy level**
MODULE 1	MO1. Explain the theoretical and practical meanings of spatial patterns and relationships. MO2. Explain the use of spatial data analysis in geographic domain (Geospatial data analysis). MO3.Explain the spatial aspects of computational, statistical, and visual analytical methodologies for big data analytics.	Understand Evaluate Remember

MODULE 2	MO1. Understand various applications of domain specific data analysis. MO2. Solve simple problems related to simulation. MO3. Handle heterogeneous data. MO4. Explain various aspects of predictive analytics and its applications. MO5. Explain the needs of complexity measures and complexity modeling.	Understand Apply Apply Understand Understand
MODULE 3	MO1. Explain the concept of spatial clustering. MO2. Distinguish between quadrant and nearest neighbor methods. MO3. Find the difference between K-function and L-function. MO4. Find the applications of point process models.	Remember Remember Analysis Analysis
MODULE 4	MO1. Explain the applications of exploratory spatial data analysis (ESDA) MO2. Understand and plot experimental variogram. MO3. Interpret scatter plots. MO4. Understand stationarity and isotropy with examples.	Understand Understand Apply Understand
MODULE 5	MO1. Explain deterministic estimation. MO2. Articulate Estimation Criteria. MO3. Evaluate the uses of Geostatistical Estimation. MO4.Understand the multivariate variant of Ordinary Kriging Operation (CoKriging).	Understand Apply Evaluate Understand

COURSE CONTENT

MODULE 1

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 2

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 3

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 4

Exploratory Spatial Data Analysis, Spatial Continuity Analysis, Experimental Variogram, hScatterplot, Variogram versus Univariate Statistics, Stationarity and isotropy, Higher

Dimensions & Statistical Anisotropy, Exploring Anisotropy, Spatial Continuity Analysis, Spatial dependence, Spatial Auto correlation

MODULE 5

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

TEXT BOOKS

1. Li, Deren. Spatial Data Mining: Theory and Application. S.l.: Springer, 2016.
2. Bivand, Roger S, Edzer J Pebesma, and Virgilio Gómez-Rubio. Applied Spatial Data Analysis With R. New York: Springer, 2008.
3. Lloyd, Christopher D. Spatial Data Analysis: An Introduction for GIS Users. Oxford: Oxford UP, 2010.

SEMESTER : IV
COURSE CODE : STSD 243(v)

COURSE TITLE : ACTUARIAL STATISTICS

COURSE OUTCOMES

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

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

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

CO3. Describe estimation procedures for lifetime distributions.

CO4 Explain the concept of survival models.

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

CO6. Describe Net premiums and its various types.

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

MODULE OUTCOMES

Sl. No:	Outcomes On completion of each module, students should be able to:	Taxonomy level
MODULE 1	MO1. Explains the utility theory and insurance. MO2. Explain survival function and application. MO3. Examine the properties of force of mortality. MO4 Define Life tables and its relation with survival function, examples.	Understand Understand Apply Remember
MODULE 2	MO1.Explain Multiple life functions and its properties. MO2. Articulate the insurance and annuity benefits through multiple life functions evaluation for special mortality laws. MO3. Explains the Multiple decrement tables. MO4.Describe net single premiums and their numerical evaluations.	Understand Analysis Understand Remember
MODULE 3	MO1.Define Distribution of aggregate claims . MO2. Derive the compound Poisson distribution and explain its applications. MO3. Explain Principles of compound interest and its attributes.	Understand Remember Understand
MODULE 4	MO1. Explain the Life insurance and its types. MO2 : Describe Insurance payable at the moment of death and at the end of the year of death-level benefit insurance MO3. Explain the Life annuities and its types.	Understand Analysis Understand
MODULE 5	MO1. Explain Net premiums and its importance MO2. Distinguish between Continuous and discrete premiums MO3. Accumulation type benefits.	Understand Analysis Apply

COURSE OUTCOME

MODULE 1

Utility theory, insurance and utility theory, models for individual claims and theirnsums, survival function, curtate future lifetime, force of mortality. Life tables and its

relation with survival function, examples, assumptions for fractional ages, some analytical laws of mortality, select and ultimate tables.

MODULE 2

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

MODULE 3

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

MODULE 4

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

MODULE 5

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

TEXT BOOKS

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

SEMESTER : IV
COURSE CODE : STSD 243(vi)

COURSE TITLE : BAYESIAN INFERENCE

COURSE OUTCOMES

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

CO1. Use relative frequencies to estimate probabilities.

CO2. Calculate conditional probabilities

CO3. Calculate posterior probabilities using Bayes' theorem.

CO4. Calculate simple likelihood functions

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

MODULE OUTCOMES

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

COURSE CONTENT

MODULE 1

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

MODULE 2

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

MODULE 3

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

MODULE 4

Bayesian interval estimation: Credible intervals. Highest posterior density regions. Interpretation of the confidence coefficient of an interval and its comparison with the interpretation of the confidence coefficient for a classical confidence interval.

MODULE 5

Bayesian testing of Hypothesis: Specification of the appropriate form of the prior distribution for a Bayesian testing of hypothesis problem. Prior odds, Posterior odds, Bayes factor for various types of testing hypothesis problems depending upon whether the null hypothesis and the alternative hypothesis are simple or composite. Specification of the Bayes tests in the above cases. Discussion of Lindley's paradox for testing a point hypothesis for normal mean against the two-sided alternative hypothesis.

REFERENCES

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

SEMESTER: IV
COURSE CODE: STSD 244

COURSE TITLE: PRACTICAL III (USING R AND PYTHON)

(Problems covering papers STSD 241, STSD 242 and STSD 243)

SEMESTER IV
COURSE CODE : STSD 245

COURSE TITLE : PROJECT/ INTERNSHIP

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.

Approved by

09. 10. 2020

[Dr C. Satheesh Kumar
Chairman: Board of Studies–Statistics(PG)]

Dr. C. SATHEESH KUMAR
CHAIRMAN
P.G. BOARD OF STUDIES IN STATISTICS
UNIVERSITY OF KERALA
THIRUVANANTHAPURAM

DEPT. OF STATISTICS
KARIAVATTOM CAMPUS
THIRUVANANTHAPURAM - 695 581
UNIVERSITY OF KERALA