SRI DHARMASTHALA MANJUNATHESHWARA COLLEGE, UJIRE-574240

(Autonomous)

(Re-Accredited by NAAC at 'A' Grade with CGPA 3.61 out of 4)

DEPARTMENT OF PG STUDIES IN STATISTICS

Syllabus of

Master's Degree in

STATISTICS

(CREDIT BASED SEMESTER SCHEME)

2020-2021 onwards

Approved by the BOS meeting held on 17th August, 2020
Academic Council meeting, held on 10-11-2020

Preamble:

Revision of the Syllabus for the Two years Master Degree (Choice Based Credit System – Semester Scheme) Programme in Statistics.

The PG BOS in Statistics has prepared the revised Syllabus for M.Sc. Statistics (CBCS based) in its meeting held on 5th September 2019, as per the guidelines suggested by Mangalore University and University Grants Commission, New Delhi. It was resolved to implement this new syllabus from the academic year 2020-21.

In the present revised syllabus, the suggested course pattern includes Hard Core, Soft Core and Open Elective courses with 92 credits for the entire programme. The syllabus consists of 14 Hard Core courses (4 credits each) including 11 theory (3 in I, II, III, and 2 in IV semesters), 2 practicals (in I semester) and one Project work (in IV semester), with a total of 56 credits. It also consists of 10 Soft Core courses (3 credits each) including 5 theory (1 in I, II, III, and 2 in IV semesters) and 5 practicals (2 in II, III, and 1 in IV semesters), with a total of 30 credits. The BOS has also proposed 2 Open Elective courses (1 each in II and III semesters) with 3 credits each (with a total of 6 credits), to be offered to non-Statistics students. But the credits of Open Elective courses are not considered for CGPA. All together total credits come to 92 (including the credits for Open Elective courses), otherwise, a total of 86 credits.

Faculty of PG Studies in Statistics:PGSTAT056

Programme Specific Outcomes:

PSO1: Show the ability to use the knowledge on theoretical foundations for the development of various statistical concepts and procedures.

PSO2: Develop technical skills in probability modelling and statistical inference for the practical application of statistical methods in their future employment.

PSO3: Be able to find solutions to real world problems by applying quantitative modelling and data analysis techniques.

PSO4: Exhibit the skills in the use of computational and statistical software to develop and execute various statistical techniques and statistical computing algorithms.

PSO5: Demonstrate theoretical knowledge and applications of parametric, semi-parametric and non-parametric testing procedures.

PSO6: Design experiments and surveys with a view of providing solutions to real life problems.

PSO7: Be able to use statistical reasoning, formulate a problem in statistical terms, perform exploratory analysis of data, and carry out a variety of advanced inferential procedures.

PSO8: Be familiar with tackling emerging problems through applications of statistics.

List of Hardcorepapers:

i. RealAnalysis
ii. ProbabilityandDistributions–I
iii. Theory ofSampling
iv. ProbabilityandDistributions–II
v. DesignandAnalysisofExperiments
vi. Theory ofEstimation
vii. TheoryofTestingofHypothesis
viii. RegressionAnalysis
ix. MultivariateAnalysis
x. Time SeriesAnalysis
xi. Reliability and SurvivalAnalysis

List of Softcorepapers:

i. LinearAlgebra
ii. DataManagementandStatisticalComputingwithPython
iii. StochasticProcesses
iv. StatisticalModelling
v. Big DataAnalytics
vi. ArtificialIntelligence
vii. Elements of StatisticalComputing
viii. SurvivalAnalysis
ix. StochasticFinance
x. DataMining
xi. BayesianInference
xii. Statistical Methods forReliability
xiii. NonparametricInference
xiv. ActuarialMethods
xv. PatternRecognitionandImageProcessing
xvi. OperationsResearch

Value Added Courses (Certificate Courses):

i. MicrosoftExcel(BasictoAdvance)
ii. RandExcelforDataScience
iii. R for DataScience
iv. RforAdvancedStatisticalMethods&MachineLearning

Course Pattern for M.Sc. Statistics Programme

I Semester:

Course Code	Title of the Paper	Hrs/Week	Credits
STH 411	Real Analysis	4	4
STS 412	Linear Algebra	3	3
STH 413	Probability and Distributions - I	4	4
STH 414	Theory of Sampling	4	4
STP 415	Practical-I(BasedonRprogrammingandSTH 414)	8	4
STP 416	Practical-II(BasedonProgramminginPython)	8	4
	Mini project**		
		Total	23

**Mini project will be incorporated as 'Skill Component'.

II Semester:

Course Code	Title of the Paper	Hrs/Week	Credits
STE 421	Introductory Statistics and Data Analysis (Open Elective Course) Questionnaire Design and Sample Selection Data Visualization	3	3
STH 422	Probability and Distributions - II	4	4
STH 423	Design and Analysis of Experiments	4	4
STH 424	Theory of Estimation	4	4
STS 425	Data Management and Statistical Computing with Python	3	3
STP 426	Practical-III(BasedonSTH423andSTH424 using R)	6	3
STP 427	Practical-IV (Based on STS 425)	6	3
	Mini project**		
		Total	21+3*

*This credit is not included for CGPA.

**Mini project will be incorporated as 'Skill Component'.

I Semester:

Course Code	Title of the Paper	Hrs/Week	Credits
STE 531	Inferential Statistics and Data Analysis (Open Elective Course) Categorical Data Analysis (3 Credits) Demographic Methods and Analysis	3	3
STH 532	Theory of Testing of Hypothesis	4	4
STH 533	Regression Analysis	4	4
STH 534	Multivariate Analysis	4	4
STS 535	Stochastic Processes	3	3
STP 536	Practical-V(BasedonSTH532andSTH533 using R)	6	3
STP 537	Practical-VI (Based on Machine Learning with Python)	6	3
	Mini project**		
		Total	21+3*

*This credit is not included for CGPA.

**Mini project will be incorporated as 'Skill Component'.

II Semester:

Course Code	Title of the Paper	Hrs/Week	Credits
STH 541	Time Series Analysis	4	4
STH 542	Reliability and Survival Analysis	4	4
STS 543	Statistical Modelling	3	3
STS 544	Big Data Analytics	3	3
STP 545	Practical-VII(BasedonSTH541,STH542and STS 543 usingR)	6	3
STP 546	Project Work**	8	4
		Total	21

**Project Work will involve 'Skill Component'.

Scheme of Internal Assessment Evaluation

The scheme of evaluation for internal assessment marks shall be as follows:

i.	TwoInternalTests	20marks
ii.	Seminar/Assignments/ClassroomActivitiesetc.	10marks
	Total:	30marks

Question PaperPattern

The pattern of question paper in theory examinations shall be as follows:

i. Thereshallbetotally8questionsinwhichQ.No.1iscompulsory.Studentshave toanswerany4questionsfromtheremaining7questions.

ii. Q.No.1willcontain8questionsofshortanswertype,eachquestioncarrying3 marks.Studentswillhavetoanswerany6questions.ThusQ.No.1carries 18 marks.

iii. Q.No.2toQ.No.8will be of long answer type, each question carrying13marks. The distribution of marks will be as follows:

Q.No.1	3×6=18
Any 4 questions out of remaining7questions	13×4 =52
	Total=70

(Prof. Shanthiprakash)
Chairman
PG B.O.S. in Statistics

I Semester

STH 411 - Real Analysis (4 Credits)

Rationale/Learning Objectives:

1. Thiscourseprovidesnecessarymathematicalfoundationsrequiredforunderstanding different theoretical aspects instatistics.

Course Outcomes:

CO1: Be able to describe the fundamental properties of real numbers that lead to the formal development of real analysis.

CO2: Show familiarity with necessary mathematical foundations required for understanding different theoretical aspects in statistics.

CO3: Understand the concept of limits and how they are used in sequence, series, differentiation and integration.

CO4: Construct mathematical proofs for basic results involved in real analysis.

Unit 1: Elements of set theory, sets in Euclidean space of k-dimensional R^krectangles. Metric spaces, neighbourhood, interior point and limit point, open andclosedsets,Bolzano-WeierstrasstheoreminR^2,compactset,real-valued functions, Heine-Borel theorem (Statement only), continuity and uniform con- tinuity.(13hrs)

Unit2: SequencesandSeriesofrealnumbers-Cauchysequence,convergence ofboundedmonotonesequence.Limitsuperior,limitinferiorandlimitprop- erties. Series of positive terms - tests for convergence, divergence. Series of arbitraryterms-absoluteandconditionalconvergence.(13hrs)

Unit3: Sequencesoffunctions-uniformconvergenceandpointwiseconver- gence,seriesoffunctions-uniformconvergence,Weierstrass'Mtest.Power series and radius of convergence. Reimann- Stieltjes integration continuous integrand and monotonic/ differentiable integrator.(13hrs)

Unit4: Functionsoftwovariables-partialanddirectionalderivatives.Max- ima and minima of functions, maxima-minima under constraints (Lagrange's

multipliers).Improperintegrals.(13hrs)

Books for Reference:

1. Apostol, T. M. (1985). *Mathematical Analysis*. Narosa IndiaLtd.
2. Bartle,R.G.(1975).*TheElementsofRealAnalysis*(2nded.).Wiley.
3. Courant, R. and John,F.(1965).*Introduction to Calculus and Analysis*. Wiley.
4. Goldberg,R.R.(1970).*MethodsofRealAnalysis*.OxfordPublishingCo.
5. Khuri,A.T.(1993).*AdvancedCalculuswithApplicationsinStatistics*.John Wiley.
6. Rudin, W. (1976). *Principles of Mathematical Analysis*. McGrawHill.

STS 412 - Linear Algebra (3 Credits)

Rationale/Learning Objectives:

1. This course provides necessary mathematical foundations on matrix algebraand vector geometry for better understanding of linear models and multivariateanalysis.

Course Outcomes:

CO1: Be aware of necessary theoretical foundations on matrix algebra and vector geometry, which will help them better understand linear models and multivariate analysis.

CO2: Be able to learn about the implementation of various mathematical aspects in practical problems.

CO3: Develop algebraic skills and knowledge on computational techniques essential for the study of vector spaces, matrix algebra, linear transformations, systems of linear equations, eigenvalues and eigenvectors, and quadratic forms.

CO4: Be familiar with the use of 'R' software in solving computational problems of linear algebra.

Unit 1: Fields, vector spaces, subspaces, linear dependence and independence, basisanddimensionofavectorspace,finitedimensionalvectorspaces,com- pletion theorem. Examples of vector spaces over real fields.Vector spaces with an inner product, Gram-Schmidt orthogonalization process, orthonormal basis.(10hrs)

Unit 2: Row and column spaces of a matrix. Rank and inverse of a matrix, propertiesofinverse.Rankofaproductofmatrices,partitionedsubmatrices, rankfactorizationofamatrix,rankofasum,inverseofapartitionedmatrix. Generallinearsystemofequations,generalizedinverse,Moore-Penroseinverse, idempotentmatrices.Solutionsofmatrixequations.(10hrs)

Unit3: Characteristicrootsandvectors,Cayley-Hamiltontheorem,minimal polynomial,similarmatrices.Algebraicandgeometricmultiplicityofcharac- teristicroots,spectraldecompositionofarealsymmetricmatrix,reductionof apairofrealsymmetricmatrices,singularvaluedecomposition.

(10

hrs)

Unit4: Realquadraticforms,reductionandclassificationofquadraticforms, index and signature. Extrema of quadratic forms.Vector and matrix differ- entiation.(10hrs)

Books for Reference:

1. Biswas, S. (1984). *Topics in Algebra of Matrices*. AcademicPublications.
2. Hadley, G. (1987). *Linear Algebra*.Narosa.
3. Graybill, F. A. (1983). *Matrices with Applications inStatistics*.
4. Rao, A. R. and Bhimasankaran, P. (1992). *Linear Algebra*. Tata McGraw Hill.
5. Rao,C.R.(1973).*LinearStatisticalInferenceanditsApplications*(2nded.). Wiley.
6. Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications*.Wiley.
7. Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*.Wiley.

Practicals based on RProgramming

1. IntroductiontoRprogrammingandbasicsofR.
2. Algebra ofmatrices.
3. Vector space and linear independence ofvectors.
4. Diagonalization of amatrix.
5. Solution to system of linearequations.
6. Transformation and Gram-Schmidtorthogonalization.
7. Eigen values and eigenvectors.
8. Spectraldecompositions.
9. Sketching of p.m.fs, p.d.fs andd.f.
10. Model sampling from univariate and bivariatedistributions.

STH 413 - Probability and Distributions - I (4 Credits)

Rationale/Learning Objectives:

1. Thiscourseprovidesnecessarytheoreticalfoundationsonthedevelopmentsofstatistical concepts and develops problem solvingskills.

Course Outcomes:

CO1: Be familiar with necessary theoretical foundations on the developments of statistical concepts and develop problem solving skills.

CO2: Be able to understand the fundamental aspects and principles of probability theory.

CO3: Exhibit learning about the standard discrete and continuous univariate distributions and its characteristics.

CO4: Show imporvedknowledge on various transformation techniques, order statistics, truncated and mixed distributions.

Unit 1: Algebra of sets, sequence of sets and limits, fields and sigma-fields, minimal sigma-field. Events, sample space. Probability measure, probability space, property of probability measure, propertiesrelatedtosequencesof events,independent events, conditional Probability.(13hrs)

Unit2: Measurablefunctions,randomvariables,probabilityinducedbyaran-domvariable.Definitionofsimplerandomvariables.Integrationofmeasurable functions with respect to measures.Expectation, properties of expectation, moments,inequalities.(16hrs)

Unit 3: Standard discrete and continuous univariate distributions andtheir properties, probability generating function and moment generating function. Bivariate normal and multinomial distributions.Transformation techniques.Distributionoffunctionsofrandomvariables.(13hrs)

Unit4: OrderStatistics-theirdistributionsandproperties,jointandmarginal distributions. Distributionofrangeandmedian.Truncatedandmixturedis- tributions.(10hrs)

Books for Reference:

1. Ash, R. B. and CatherineDoleans-Dade(2000).*Probability and Measure Theory*. AcademicPress.
2. Bhat, B. R. (1999). *Modern Probability Theory* (3rd ed.). New Age InternationalPublishers.
3. Johnson, S. and Kotz (1972). *Distributions in Statistics*. Vols. I, II and III, Houghton andMiffin.
4. Mukhopadhyaya,P.(1996).*Mathematical Statistics*. Calcutta Publishing House.
5. Pitman, J. (1993). *Probability*.Narosa.
6. Rao,C.R.(1973).*LinearStatisticalInferenceanditsApplications*(2nded.). WileyEastern.
7. Rohatgi,V.K.andSaleh,A.K.Md.E.(2015). *AnIntroductiontoProbability Theory and Mathematical Statistics*. WileyEastern.
8. Laha, R. G. and Rohatgi, V. K. (1979). *Probability Theory*. WileyEastern.
9. Ross,S.M.(1993).*FirstCourseinProbability*.AcademicPress.
10. Billingsley, P. (1986). *Probability and Measure.* John Wiley andSons.

STH 414 - Theory of Sampling (4 Credits)

Rationale/Learning Objectives:

1. This course provides theoretical knowledge on various sampling techniques used for designing and selecting a sample from a population.

Course Outcomes:

CO1: Be able to understand the basic principles underlying survey design and estimation.
CO2: Exhibit theoretical knowledge on various techniques used for designing and selecting a sample from a population.
CO3: Show an increased learning about how to estimate finite population parameters.
CO4: Be able to implement and analyze various sampling techniques to real life problems.

Unit1: BasicConcepts-samplingdesign,samplingscheme,samplingstrategy, interpenetrating subsampling, concept of non-random sampling. Probabilityproportionaltosizewithreplacement(PPSWR)sampling-selectionof PPSWRsample,estimationofpopulationmean,totalandtheirsamplingvariances. Hansen-Hurwitz strategy, estimation of sampling variance.Comparison with SRSWR, estimation of gain due toPPSWR sampling.(13hrs)

Unit2: Varyingprobabilitywithoutreplacement(PPSWOR)sampling-some properties of sampling design. Horwitz-Thomson estimator, samplingvarianceofpopulationtotalanditsunbiasedestimator.Sen-Midzunosampling scheme, Des-Raj's ordered estimator (general case), Murthy's unorderingprinciple (sample ofsizetwo).(13hrs)

Unit 3: Single stage cluster sampling - concepts, estimation of efficiency of clustersampling,clustersofvaryingsizes.Two-stagesampling-notions,estimationofpopulationtotalanditsvariance,efficiencyoftwo-stagesampling relativetoclusterandunistagesampling.(13hrs)

Unit4: RatioandregressionestimatorsbasedonSRSWOR,methodofsampling,biasandmeansquareerrors,comparisonwithmeanperunitestimator.Two-phasesampling-notion,doublesamplingforratioestimation,double sampling for regression

estimation.Randomized response techniques - Warner's model, related and unrelated questionnaire methods, non-sampling errors.OfficialStatisticsforNationalDevelopment-NSO,CSO,MOSPI,Hu- man Development Index. Measuring inequality in income - Lorenz Curve, Ginicoefficient.(13hrs)

Books for Reference:

1. Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.).Wiley.
2. Des Raj and Chandok (1998). *Sampling Theory*. NarosaPublication.
3. Mukhopadhyay, P. (1998). *Theory and Methods of survey Sampling*. Prentice Hall ofIndia.
4. Murthy, M. N. (1977). *Sampling Theory and Methods*. Calcutta: Statistical PublishingSociety.
5. Sampath, S. (2001). *Sampling Theory and Methods*. NarosaPublishers.
6. Sen, A. (1997). *Poverty andInequality*.
7. Singh, D. and Chaudhary, F. S. (1986). *Theory and Analysis of Sample Survey Designs*. New Age InternationalPublishers.
8. Sukhatme, P.V.,Sukhatme,B.V.,Sukhatme,S.andAshok(1984).*Sampling Theory of Surveys with Applications*. ICARPublication.
9. Banett, V. (2002). *Sample Survey: Methods and Principles*. Arnold Publish- ers.

Practicals based on Theory of Sampling

1. Determination of samplesize.
2. Probabilityproportionaltosizewithreplacement(PPSWR)-I
3. Probabilityproportionaltosizewithreplacement(PPSWR)-II
4. Probabilityproportionaltosizewithoutreplacement(PPSWOR)-I
5. Probability proportional to size without replacement (PPSWOR) -II
6. Des Raj'sorderedestimator.
7. Des Raj'sordered estimator and Murthy's unorderedestimator.
8. Stratified and systematicsampling.
9. Clustersamplingwithclustersofequalsize.
10. Clustersamplingwithclustersofunequalsize.
11. Two stagesampling.
12. Multistagesampling.

13. Ratio method ofestimation.
14. Ratio and regression method ofestimation.

Practicals based on Programming in Python

1. Write a program to perform addition, subtraction, multiplication, division and modulooperationsontwointegernumbers(readtheinputfromkeyboard).
2. (a)Writeaprogramtoreadthreeintegernumbersfromkeyboardandfind the largest among threenumbers.
   (b) Writeaprogramtodeterminewhethertheenteredcharacterisa vowel ornot.
3. Writeapythonprogramtoperformaddition,subtraction,multiplication,di- vision on two matrices (read the input fromkeyboard).
4. Writeapythonprogramtoperformarithmeticoperationinexcel.
5. Writeaprogramtoconvertasquarematrixintoaloweranduppertriangular matrix.
6. (a)Writeaprogramtoidentifywhetheraninputmatrixissymmetricornot.
   (b) Write a program to identify whether an input matrix is binary ornot.
7. Writeaprogramtocountthenumberofdigits,uppercasecharacters,lowercase charactersandspecialcharactersinagivenstring.
8. Write a python program to find mean, median, mode of a entered list of numbers.
9. Write a program that has dictionary of names of students and a list of their marks in four
   subjects. Create another dictionary from this dictionary that has name of the student and their total marks. Find out topper and his/her score.
10. Write a program that has a class student that stores roll number, name and marks(inthreesubjects)ofthestudents.
    Displaytheinformation(rollnumber,name,andtotalmarks)storedaboutthestudent.

II Semester

STE 421 - Introductory Statistics and Data Analysis (3Credits)

(OpenElective Course)

Rationale/Learning Objectives:

Thiscourseprovidesknowledgeonfundamentalsinvolvedindescriptivestatistics, probabilityandsampling,whicharewidelyusedindataanalysis.

Unit1: Statistics-introduction,meaning,definitionandscopeofthesubject as a science of decision making against uncertainty. Data types, methods of collection,presentationintheformoftablesandgraphs.DescriptiveStatistics-measuresofcentraltendency,positionalaverages,measuresofdispersion, skewnessandkurtosis.Methodsofsummarizingcategoricaldataunivariate and bivariate contingency tables.Box plots - construction and interpreta- tions. Exploratory data analysis using descriptive measures and graphical tools.(13hrs)

Unit2: Theconceptofrandomexperiment,simpleevents,samplespace,types ofevents,probabilityofanevent,rulesofprobability,conditionalprobability, Baye'srule,exercisesoncomputationofprobabilitiesusingtheserulestofix theideas.Theconceptofrandomvariables-discreteandcontinuoustype,Binomial,PoissonandNormaldistributions-theiruseinpracticalapplications, computingprobabilitiesusingthesedistributions.(13hrs)

Unit3: Samplingmethods-populationandsample,parameterandstatistic, concept of a random sample, simple random sampling, stratifiedsampling, systematic, sampling, sample size determination. The concept of sampling distributionofastatisticandstandarderror. (14hrs)

Books for Reference:

1. Campbell, R. C. (1974). *Statistics for Biologists*. Cambridge UniversityPress.
2. Chatfield, C. (1981). *Statistics for Technology*. Chapman andHall.
3. Frank, H. and Athoen, S. C. (1997). *Statistics: Concepts & Applications*. Cambridge UniversityPress.
4. Medhi, J. (1992). *Statistical Methods: An Introductory Text*. WileyEastern Limited.
5. Ross, S. M. (2017). *Introductory Statistics*. AcademicPress.
6. Rice, J. A. (2006). *Mathematical Statistics and Data Analysis*. Singapore: Thomson-Duxbury.

Questionnaire Design and Sample Selection(3Credits)

(OpenElective Course)

Rationale/Learning Objectives

- To understand the characteristics of a well-designed questionnaire.
- To design questionnaires in a way that make them attractive towards target respondents.
- To ask specific questions in a way that will encourage responses.
- To choose the best method of distributing survey and get the right people to answer it.

Unit 1: Introduction, qualities of a good questionnaire, types of questionnaires: exploratory questionnaire (qualitative) and formal standardized questionnaire (quantitative). Questionnaire question types: open-ended questions, multiple choice questions, dichotomous questions, scaled questions, and pictorial questions; questions to avoid in a questionnaire. (14 hours)

Unit 2: Steps involved in the development of a questionnaire, methods of reaching target respondents: personal interviews, group or focus interviews, mailed questionnaires, and telephone interviews. Advantages and disadvantages of questionnaires.Examples of questionnaires.Introduction to pilot surveys and its use in questionnaire development and modification.
(13 hours)

Unit 3: Sampling: introduction, techniques: probability sampling – simple random, systematic, stratified, and cluster sampling; non-probability sampling – convenience, quota, judgement, and snowball sampling. Applications of sampling.Sample size determination.
(13 hours)

References

1. http://www.fao.org/3/w3241e/w3241e05.htm
2. https://www.kyleads.com/blog/questionnaire/
3. https://www.digitalvidya.com/blog/sampling-techniques/

Data Visualization (3 credits)

(OpenElective Course)

Rationale/LearningObjectives

- Data Preparation, Basic Concepts and Methods of Data Visualization.
- Develop simple summaries and exploratory graphs that optimize data visualization.
- Understand the characteristics and purposes of visualizing data.
- Understand data distributions and relationship between variables.
- Employ best practices in data visualization to develop charts, maps, tables, and other visual representations of data.

Unit 1: Tables and Univariate Graphs

Introduction, Tables, Q-Q Plot, Categorical (Bar chart, Pie chart, Tree map), Quantitative (Histogram, Frequency polygon, Frequency Curve, Ogives, Stem and leaf plot, Dot chart). Applications and examples.(12 hours)

Unit 2: Bivariate Graphs

Categorical vs. Categorical (Stacked bar chart, Grouped bar chart, Segmented bar chart), Quantitative vs. Quantitative (Scatterplot, Line plot, Area chart), Categorical vs. Quantitative (Bar chart on summary statistics, Box plots). Applications and examples.(14 hours)

Unit 3: Multivariate Graphs & Statistical Models

Grouping, Faceting, Correlation plots, Linear Regression, Scatter Plot Matrix, Parallel coordinates plot, Star plot, Chernoff faces, Growth curve. Applications and examples.(14 hours)

References

1. https://rkabacoff.github.io/datavis/
2. https://www.analyticsvidhya.com/blog/2015/07/guide-data-visualization-r/
3. https://towardsdatascience.com/a-guide-to-data-visualisation-in-r-for-beginners-ef6d41a34174
4. https://www.tutorialspoint.com/excel_data_analysis/excel_data_analysis_visualization.htm

STH 422 – Probability and Distributions – II (4 Credits)

Rationale/Learning Objectives:

1. Thiscourseprovidesadequateconceptualbasisfortheasymptotictheoryandcharacterization properties of thedistributions.

Course Outcomes:

CO1: Be able to understand the conceptual basis for the asymptotic theory.

CO2: Be familiar with the characterization properties of the distributions.

CO3: Understand the applications of theoretical aspects.

CO4: Be able to solve problems by using the theoretical knowledge.

Unit1: Measure,probabilitymeasure,propertiesofameasureandprobability, Carathedory extension theorem (statement only). Lebesgue and Lebesgue- Stieltjes measure on the real line. Absolute continuity, definition of Radon- Nikodym derivativeandillustrations.(13hrs)

Unit2: Monotoneconvergencetheorem,Fatou'slemmaanddominatedconvergencetheorem.Borel-Cantellilemma,convergenceinprobability,convergence almost surely, convergence in distribution, convergence in r^{th}mean, convergencetheorem for expectations.Slutsky'stheorem.(13hrs)

Unit3: Weaklawoflargenumbers–Kolmogorov'sgeneralizedWLLN(proof ofsufficientconditiononly),Khintchine'sWLLNasspecialcase,Chebyshev'sWLLN. Kolmogorov's strong law of large number sequence of independentand iidrandomvariables.Kolmogorov'sinequality.(13hrs)

Unit 4: Characteristic function – properties, inversion theorem (statement only and proof for density version), uniqueness theorem, continuity theorem (statement only). Central limit theorem, Lindeberg-Levy and Liapounov central limittheorems.StatementofLindeberg-Fellerform(statementonly).Application of these theorems.Sampling distributions, chi-square, *t*, *F* and non- centralchi-square,mgfofnon-centralchi-squaredistribution, reproductive property.Non-central *t*and non-central *F*.(13hrs)

Books for Reference:

1. Bhat, B. R. (1999). *Modern Probability theory* (3rded.). New Age International Publishers. (To be used asText.)
2. Mukhopadhyaya,P.(1996).*Mathematical Statistics*. Calcutta Publishing House.

3. Pitman, J. (1993). *Probability*.Narosa.
4. Billingsley, P. (1986). *Probability and Measure*. John Wiley andSons.
5. Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.
6. Ash, R. B. and CatherineDoleans-Dade(2000).*Probability and Measure Theory*. AcademicPress.
7. Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*.
8. Rao,C.R.(1973).*LinearStatisticalInferenceanditsApplications*(2nd ed.). WileyEastern.
9. Rohatgi,V.K.andSaleh,A.K.Md.E.(2015).*AnIntroductiontoProbability Theory and Mathematical Statistics*. WileyEastern.

STH 423 – Design and Analysis of Experiments (4 Credits)

Rationale/Learning Objectives:

This course provides theoretical foundations on the fundamentals and principles involved in designed experiments.

Course Outcomes:

CO1: Demonstrate necessary theoretical foundations on the fundamentals and principles involved in designed experiments.

CO2: Exhibit theoretical knowledge on various experimental designs such as BIBD, nested designs, factorial experiments, split-plot designs, strip-plot designs, complete and partial confounding.

CO3: Be able to construct standard experimental designs and identify the appropriate statistical models to analyze the data.

CO4: Be able to understand the importance and applications of experimental designs in analyzing real life problems.

Unit 1: Gauss-Markov setup, normal equations and least squares estimates, estimablefunctionandestimationspace,varianceandcovarianceofleastsquares estimates,estimationoferrorvariance,estimationwithcorrelatedobservations,simultaneouses timatesoflinearparametricfunctions.Testsofhypothesisforoneandmorethanonelinearparame tricfunctions.Confidenceintervals and regions, analysis of variance, power of F-test, multiplecomparison tests–TukeyandBonferroni,simultaneousconfidenceinterval.(13hrs)

Unit2: Introductiontodesignedexperiments,generalblockdesignanditsin- formation matrix, criteria for connectedness, balance and orthogonality. Intra- block analysis –estimability, best point estimates/interval estimates of estimable linear parametric function sandtesting of linear hypotheses, estimation f parameters.(13hrs)

Unit 3: BIBD – definition, concept of connectedness, balancing, properties, estimability,recoveryofinter-blockinformation.Analysisofcovarianceina generalGauss-Markovmodel,applicationstoCRDandRCBD.Fixed,mixed and random effects models, variance components estimation, study of various methods.(13hrs)

Unit 4: General factorial experiments, factorial effects – best estimates and testing the significance of factorial effects, study of $2^n$and 3^nfactorial experimentsinrandomizedblocks.Completeandpartialconfounding.Nested designs. Split-plot, stripplotdesigns.(13hrs)

Books for Reference:

1. Bapat, R. B. (2012). *Linear Algebra and Linear Models*. Hindustan Book Agency.
2. Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley Eastern.
3. AlokeDey (1986). *Theory of Block Designs*. WileyEastern.
4. Dean,A.andVoss,D.(1999).*DesignandAnalysisofExperiments*.Springer.
5. Chakrabarti, M. C. (1962). *Mathematics of Design and Analysis of Experiments*.Asia.
6. Cochran and Cox, D. R. (1957). *Experimental Designs*. JohnWiley.
7. Das, M. N. and Giri, N. (1979). *Design and Analysis of Experiments*.Wiley Eastern.
8. Giri, N. (1986). *Analysis of Variance*. South AsianPublishers.
9. John, P. W.M.(1911).*Statistical Design and Analysis of Experiments*. Macmillan.
10. Joshi, D. D. (1987). *Linear Estimation and Design of Experiments*. Wiley Eastern.
11. Montgomery,C.D.(1976).*DesignandAnalysisofExperiments*.NewYork: Wiley.
12. Mukhopadhyay, P. (1998). *Applied Statistics*. Books and Allied (P)Ltd.
13. Pearce, S. C. (1984). *Design of Experiments*. New York:Wiley.
14. Rao, C. R. and Kleffu, J. (1988). *Estimation of Variance Components and Applications*. NorthHolland.
15. Searle, S. R., Casella, G. and McCullugh, C. E. (1992). *Variance Components*. Wiley.

Practicals based on Design and Analysis of Experiments

1. One way classification and multiple comparisontests.
2. Twowayclassificationwithequalnumberofobservationspercell(modelwith interaction).
3. Twowayclassificationwithunequalnumberofobservationspercell(model withoutinteraction).
4. Estimability andcompleteness.
5. Analysis of general blockdesign.
6. Analysis of LSD andBIBD.
7. Analysis of covariance in one way and two waymodel.
8. 2^kfactorial experiments and analysis of single replicate of 2^k.
9. Total and partial confounding in 2^kfactorialexperiments.
10. Analysis of 2^kfractional factorialexperiments.
11. Analysis of 3^2 factorialexperiments.
12. Analysis of one way classification random effectsdata.

STH 424 – Theory of Estimation (4 Credits)

Rationale/Learning Objectives:

1. This course provides necessary theoretical foundations on the developments and applications of various estimation techniques.

Course Outcomes:

CO1: Be able to understand the random phenomenon of the character of interest.
CO2: Be familiar with the estimation techniques.
CO3: Be able to understand the asymptotic 18ehavior of estimation.
CO4:Be able to understand the applications of theoretical aspects.

Unit 1: Parametric models, likelihood function, example from standarddiscreteandcontinuousmodels.Plottinglikelihoodfunctions.Sufficiency,Neyman factorization criterion, Fisher information for single and several parameters.Minimal sufficient statistic, likelihood equivalence.Exponential families and Pitmanfamilies.Completeness, Ancillary Statistics, Basu'stheoremand applications.(13hrs)

Unit2: Minimumvarianceunbiasedestimation,unbiasedness,locallyunbiased estimators, minimum variance, locally minimum various, mean squared error, Cramer-Raolower bound approach. Minimum variance unbiased estimators (MVUE), Rao-Blackwell theorem, completeness, Lehman-Scheffe'stheorem,necessaryandsufficientconditionforMVUE.Bhattacharyabounds (without proof). Introduction to interval estimation, construction of confidence intervalsusingpivot.(13hrs)

Unit3: Consistency,estimationofrealandvectorvaluedparameters,invari- ance properties. Consistency of estimators by method of moments and method ofpercentiles,meansquarederrorcriterion,asymptoticrelativeefficiency,consistent asymptoticnormal(CAN).(13hrs)

Unit 4: Method of maximum likelihood – notion, MLE in location and scale family, exponential family, Cramer family (statement only). Cramer-Huzurbazar theorem.Solutions to

likelihood equations method of scoring, Newton-Raphsonandotheriterativeprocedures.Fisherlowerboundtoasymptoticvariance,ex-tensiontomulti-parametercase(withoutproof).(13hrs)

Books for Reference:

1. Casella,G.andBerge,R.L.(2002).*StatisticalInference*(2nd ed.).Singapore: Thomson-Duxbury.
2. Kale, B. K. (1999). *A First Course on Parametric Inference*. Narosa Publish-ingHouse.
3. Lehman, E. L. (1986). *Theory of Point Estimation*. JohnWiley.
4. Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley Eastern.
5. Rohatgi, V. K. and Saleh, A. K. L. (2001). *An Introduction to Probability and Mathematical Statistics*. WileyEastern.
6. Rajagopalan, M. and Dhanavanthan, P.(2012).*Statistical Inference*. Phi Learning Pvt.Ltd.
7. Zacks, S. (1981). *Parametric Statistical Inference*. PergamonPress.

Practicals based on Theory of Estimation

1. Estimation by the methods of moments andpercentile.
2. Estimation by the methods of MLE for non-regular andmulti-parameters.
3. Estimation by the methods of MLE using iterativemethod.
4. Construction ofUMVUE.
5. Asymptotic behavior ofestimation.

STS 425 – Data Management and Statistical Computing with Python (3 Credits)

Rationale/Learning Objectives:

1. This course provides comprehensive knowledge of python programming paradigms required for data management and statistical computing.

Course Outcomes:

CO1: Be able to gain comprehensive knowledge on Python programming paradigms.
CO2: Show an increased learning about the implementation of Python programming in data management and statistical computing.
CO3: Exhibit insights on using Pandas in Python required for data manipulation.
CO4: Be able to explore about how to generate powerful data visualizations using Python.

Unit1: UsingNumpy–BasicsofNumPy-ComputationonNumPy-Aggregations- Computation on Arrays Comparisons, Masks and Boolean Arrays-Fancy Indexing- Sorting Arrays-Structured Data: NumPy'sStructuredArray. (14hrs)

Unit 2: Data Manipulation with Pandas – Introduction to Pandas Objects-DataindexingandSelection-OperatingonDatainPandas-HandlingMissing Data-HierarchicalIndexing–CombiningDataSets.(14hrs)

Unit 3: Visualization and Matplotlib– Basic functions of matplotlib-Simple LinePlot,ScatterPlot-DensityandContourPlots-Histograms,Binningsand Density-Customizing Plot Legends, Colour Bars-Three- Dimensional Plotting inMatplotlib.(12hrs)

Books for Reference:

1. VanderPlas, J. (2016). *Python Data Science Handbook – Essential Tools for Working with Data*. O'Reily Media,Inc.
2. Zhang, Y. (2016). *An Introduction to Python and Computer Programming*. SpringerPublications.
3. Thareja, R. *Python Programming using Problem Solving Approach*. Oxford UniversityPress.
4. Grus, J. (2016). *Data Science from Scratch First Principles with Python,*. O'ReillyMedia.
5. Padmanabhan, T. R. (2016). *Programming with Python*. Springer Publications.

Practicals based on Data Management and Statistical Computing with Python

1. WriteaNumPyprogramtocomputesumofallelements,sumofeachcolumn and sum of each row of a givenarray.
2. Write a NumPy program to find rank, determinant, trace and eigenvalueof anarray.
3. WriteaNumpyprogramtosortarraycontentinrowandcolumnwise.
4. Writeaprogramtopreprocessthedatausing Numpyandsklearn.preprocessingpackages.
5. Writeapythonprogramtoimportpandaslibrarytoperformjoining,merging and concatenating differentdataframe.
6. Write program to do thefollowing:

(a) Createdataframedfconsisting10rowsand4columnsofrandomlygen- erated numbers between 1 to100.

(b) Create a new column such that, each row contains the row number of nearest row-record by Euclideandistance.

7. Use Iris data set, write program to answer the followingquestions:

(a) Find the mean, median, standard deviation of iris's sepal-length (1^{st} column).

(b) Createanormalizedformofiris'ssepal-lengthwhosevaluesrangeexactly between0and1sothattheminimumhasvalue0andmaximumhasvalue 1.

(c) Findthenumberandpositionofmissingvaluesiniris'ssepal-length(1^{st} column).

8. Use Automobile dataset, write program to answer the followingquestions:

(a) CleanthedataandupdatetheCSVfile.

(b) Find the most epesensivecar.

(c) Find each company's highest pricecar.

9. UseCompaniessalesdataset,writeprogramtoanswerthefollowingquestions:

(a) Readtotalprofitofallmonthsandshowitusingalineplot.

(b) Readallproductsalesdataandshowitusingamultilineplot.

(c) Calculatetotalsaledataforlastyearforeachproductandshowitusing a piechart.

10. Use SAHeart dataset, write program to answer the belowquestions:

(a) Draw a bar plot to show the number of person having CHD or not in comparison to they having family history of the disease ornot.

(b) FindoutthenumberofCHDcasesindifferentagecategories.DoaBar Plotandsortthemintheorderofagegroups.

III Semester

STE 531 – Inferential Statistics and Data Analysis (3 Credits)

(Open Elective Course)

Rationale/Learning Objectives:

1. This course provides fundamentals for developing various tests for the validityof different hypotheses, which are widely used in data analysis.

Course Outcomes:

CO1: Be able to identify the basics of hypothesis testing and perform hypothesis test for mean, proportion and difference between means and proportions from two populations.

CO2: Construct confidence intervals for mean and proportion.

CO3: Conduct one-way analysis of variance hypothesis test.

CO4: Apply non-parametric tests, correlation and regression techniques to real life problems.

Unit1: Theconceptofhypothesisandtestsofhypothesis:nullhypothesis,alternatehypothesis,teststatistic,levelofsignificance,*p*-value,testinghypothesisaboutpopulationmeans,andpopulationproportions,confidenceintervals.Nonparametrictests–signtest,Wilcoxon-Mann-Whitneytest,Wilcoxon signedranktest.Contingencytables,chisquaretestforindependenceofat- tributes.(16hrs)

Unit 2: Testing for the equality of several population means. The conceptof analysisofvariance,onewayanalysisofvariance,itsutilityintheanalysisof surveydataanddataobtainedfromdesignedexperiments.(10hrs)

Unit3: Regressionandcorrelation–bivariatedata,correlation,scatterplot, correlationcoefficientanditsproperties,testingforcorrelationcoefficient,rank correlation. Regression – use of simple linear regression model to study the linearrelationshipbetweentwovariables,fittingthesimplelinearregression model, testing significance of regression coefficient, coefficient of determination.(14hrs)

Books for Reference:

1. Campbell, R. C. (1974). *Statistics for Biologists*. Cambridge UniversityPress.
2. Chatfield, C. (1981). *Statistics for Technology*. Chapman andHall.
3. Frank, H. and Athoen, S. C. (1997). *Statistics: Concepts & Applications*. Cambridge UniversityPress.
4. Medhi, J. (1992). *Statistical Methods: An Introductory Text*. WileyEastern Limited.
5. Ross, S. M. (2017). *Introductory Statistics*. AcademicPress.
6. Rice, J. A. (2006). *Mathematical Statistics and Data Analysis*. Singapore: Thomson-Duxbury.

Categorical Data Analysis (3 Credits)

(Open Elective Course)

Rationale/Learning Objectives

- Conceptual understanding and application of statistical procedures.
- Identify designs of contingency tables and recommend appropriate measures of association and statistical tests.
- Develop models for binary, polytomous and multivariate categorical responses, interpret results regardless of model parameterization, and diagnose model fits.
- Interpret and communicate categorical data methods.

Unit 1: Introduction and Probability distributions

What is categorical data analysis, Scales of measurement, A brief history of categorical methods, Probability distributions for categorical variables, Frequency distribution tables for discrete variables, The hypergeometric distribution, The Bernoulli distribution, The binomial distribution, The multinomial distribution, The Poisson distribution. Maximum likelihood estimation: a single proportion, Hypothesis testing for a single proportion, Confidence intervals for a single proportion, Goodness-of-fit: comparing distributions for a single discrete variable. (12 hours)

Unit 2: Analyzing Contingency Tables

Probability Structure for Contingency Tables, Comparing Proportions in 2 × 2 Contingency Tables, The Odds Ratio, Chi-Squared Tests of Independence, Testing Independence for Ordinal Variables, Contingency tables for three categorical variables, Marginal and conditional independence, Inferential statistics for three-way tables. (14 hours)

Unit 3: Generalized Linear Models

Components of a Generalized Linear Model, Generalized Linear Models for Binary Data, Generalized Linear Models for Counts and Rates, Statistical Inference and Model Checking, Fitting Generalized Linear Models. (14 hours)

References

1. Agresti, Alan. *An introduction to categorical data analysis* (Third edition). Hoboken, NJ: John Wiley & Sons, 2019. Series: Wiley series in probability and statistics, ISBN 9781119405269.
2. Azen, Razia, (1969). *Categorical Data Analysis for the Behavioural and Social Sciences*/ RaziaAzen, Cindy M. Walker. p. cm. ISBN 978-1-84872-836-3

Demographic Methods and Analysis (3 Credits)

(Open Elective Course)

Rationale/Learning Objectives

- To understand the key measures and techniques used in studying population behaviour and change.
- To explore the different sources of demographic data.
- To understand the basic demographic indicators and their applications.

Unit 1: Demography: introduction, purpose, nature of demographic information: births, fertility, fecundity, deaths, mortality, life expectancy, migration. Data collection methods: census, sample surveys, registration of vital events, population registers, and administrative records.
(10 hours)

Unit 2: Statistical measures: measures of central tendency – arithmetic mean, median, mode; normal and skewed distributions; measures of dispersion – variance and standard deviation, quantiles; correlation and linear regression. (16 hours)

Unit 3: Measurement of population, measures of fertility: crude birth rate, age-specific fertility rate, general fertility rate, and total fertility rate; measures of mortality: crude death rate, age-specific death rate, standardized death rate, infant mortality rate, neo-natal mortality rate, and maternal mortality rate. Life table and its components. (14 hours)

References

1. Yusuf, F., Martins, J. M., and Swanson, D. A. (2014). *Methods of Demographic Analysis*. Springer, New York, London.
2. Carmichael, G. A. (2016). *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*. Springer, New York, London.

STH 532 - Theory of Testing of Hypothesis (4 Credits)

Rationale/Learning Objectives:

1. This course provides necessary theoretical foundations on developments and appli- cations of various tests for the validity of different hypotheses.

Course Outcomes:

CO1: Develop various tests for the validity of different kinds of hypotheses.

CO2: Show acquisition of adequate foundations on the fundamentals involved in testing of hypothesis and understand its importance.

CO3: Show learning about the theoretical aspects of most powerful, uniformly most powerful, unbiased, likelihood ratio tests, interval estimation and its implementation in practical problems.

CO4: Exhibit knowledge on various non-parametric tests and its applications in real life problems.

Unit1: Framingofnullhypothesis,criticalregion,levelofatest,randomized and non-randomized tests, two kinds of error, size of a test, p-value, power function.Mostpowerfultestsinclassofsizeatest,Neyman-Pearsonlemma, MP test for simple null against simple alternative hypothesis. UMP tests for one sided null against one sided alternatives, monotone likelihood ratio property. Extension of these results in Pitman family when only upper or lowerendpointsdependontheparameter.(13hrs)

Unit 2: Non-existence of UMP test. Neyman-Pearson generalized lemma (statement only), concept of UMP for simple null against two sided alter- natives in one parameter exponential family and UMPU tests with applicationtooneparameterexponentialfamily, UMPfortwosidednull(statement only). Likelihood ratio test (LRT), asymptotic distribution of LRT statistic, Pearson'schisquaretestforgoodnessoffit,Bartlett'stestforhomogeneityof variances, largesampletests.(16hrs)

Unit 3: Interval estimation, confidence level, construction of confidence intervalsbyinvertingacceptanceregion.Shortestexpectedlengthconfidenceinterval,evaluatin gintervalestimatorsusingsizeandcoverageprobabilityand test related optimality, uniformly most accurate one-sided confidence interval anditsrelations to UMP test

foronesidednullagainstonesidedalternative hypothesis.(10hrs)

Unit4:U-statistics,propertiesandasymptoticdistributions(inoneandtwo sample case). Nonparametric tests: One sample test - test based on total numberofruns,theordinarysigntest,theWilcoxonsignedranktest,theKolmogorov-Smirnov one sample goodness of fit test. Two sample tests - the median test, the Wilcoxon-Mann-Whitney test, Kolmogorov-Smirnov two sampletest.

(13hrs)

Books for Reference:

1. Casella, G. and Berger, R. L. (2002). *Statistical Inference*. WadsworthGroup.
2. Gibbons, J. D. (1971). *Nonparametic Inference*. McGrawHill.
3. Kale, B. K. (1999). *A First Course on Parametric Inference*. Narosa Publish-ingHouse.
4. Lehmann, E. L. and Romano, J. (2008). *Testing Statistical Hypotheses*. John Wiley.
5. Pratt, T. W. and Gibbons, J. D. (1981). *Concepts of Nonparametric Theory*. Springer.
6. Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley Eastern.
7. Rohatgi, V. K. and Saleh, A. K. L. (2001). *An Introduction to Probability and Mathematical Statistics*. WileyEastern.
8. Rajagopalan, M. and Dhanavanthan, P.(2012).*Statistical Inference*. Phi Learning Pvt.Ltd.

Practicals based on Theory of Testing of Hypothesis

1. MPtests-I
2. MPtests-II
3. UMPtests-I
4. UMPtests-II
5. UMPUtests.
6. Likelihood ratiotests.
7. Confidenceintervals.
8. Large sample tests using variance stabilizing transformation.
9. Non-parametrictests.
10. Goodness of fittests.
11. Kolmogorov-Smirnovtests.

STH 533 - Regression Analysis (4 Credits)

Rationale/Learning Objectives:

This course provides theoretical foundations on regression techniques, which are extensively used in data analysis.

Course Outcomes:

CO1: Understand the relationship between the response and the predictors and how the variation in response is explained by the predictors.

CO2: Show an acquisition of necessary theoretical foundations on different regression techniques and its extensive use in data analysis.

CO3: Be skilled in model adequacy checking and regression diagnostics.

CO4: Be familiar with the theoretical aspects of simultaneous equation models and identification problem.

Unit 1: Simple linear regression, multiple linear regression, basic assumptions, ordinaryleastsquares(OLS)-estimationandtheirproperties,testsofhypothesis about regression coefficients, likelihood ratio criterion. Dummy variables.Prediction-bestlinearunbiasedpredictor.(10hrs)

Unit 2: Regression diagnostics and specification tests - residual analysisfor identifying influential observations, recursive residuals and their applications, specification tests, subset selection of explanatory variables, Mallows $C_p$statistic.Useofpriorinformation.Restrictedleastsquaresestimatorsandmixed regressionestimator.(10hrs)

Unit3: Violationofbasicidealconditions-disturbancewithnon-zeromean, asymptotically uncooperative regressors. Multicollinearity - its consequences and testing. Ridge estimator and its properties, ridge regression.Stochastic regressors, autoregressive models, instrumental variables, errors in variables.Distributedlagmodels.(10hrs)

Unit 4: Heteroscedasticity - tests for heteroscedasticity. Generalized least squares (GLS) estimators and its properties, feasible generalized least squares estimators. Grouping of observations.Sets of Regression Equations. Auto cor- relation - its consequences and testing

for autocorrelation, estimation and prediction.Autoregressiveconditionalheteroscedasity(ARCH)models.(10hrs)

Unit 5: Simultaneous equation models. Identification problem, identification usinglinearhomogeneousrestrictionsonstructuralparameters,rankandorder conditions,estimationinsimultaneousequationmodels.Indirectleastsquares, twostageleastsquares,structuralequationmodelling.(12hrs)

Books for Reference:

1. Cook, R. D. and Weisberg, S. (1982). *Residual and Influence in Regression*. London: Chapman andHall.
2. Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis* (3rd ed.). New York:Wiley.
3. Gunst,R.F.andMason,R.L.(1980).*RegressionAnalysisanditsApplication - A Data Oriented Approach*. Marcel Dekker.
4. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). *Introduction to Linear Regression Analysis*. JohnWiley.
5. Ryan, T. P. (1997). *Modern Regression Methods*. New York: JohnWiley.
6. Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis* (2nd ed.). New York: JohnWiley.
7. Fomby, T. B., Hill, C. R. and Johnson, S. R. (1988). *Advanced Econometric Methods*.Springer.
8. Greene, W. H. (2002). *Econometric Analysis* (5th ed.). New York: Prentice Hall.
9. Johnston,J.andDinardo,J.(1996).*EconometricMethods*(4thed.).McGraw- Hill.
10. Maddala, G. S. (1992). *Introduction to Econometrics* (2nd ed.). New York: Macmillan.
11. Gujarati, D. N. (2004). *Basic Econometrics* (4th ed.).McGraw-Hill.

STH 534 - Multivariate Analysis (4 Credits)

Rationale/Learning Objectives:

1. Thiscourseprovidestheoreticalfoundationsonmultivariatetechniques,whichare extensively used in dataanalysis.

Course Outcomes:

CO1: Show an acquisition of necessary theoretical foundations on various statistical techniques for analyzing vector-valued random entities.

CO2: Exhibit theoretical knowledge on various multivariate techniques such as principal component analysis, cluster analysis, classification and discrimination.

CO3: Apply the multivariate techniques in solving real life problems.

CO4: Understand the extensive use of multivariate techniques in data analysis.

Unit 1: Nature of a multivariate problem, main types of multivariate problems, objectives of multivariate analysis. Organization of multivariate data, descriptive statistics, visualization techniques.Multivariate normal distributionproperties,maximumlikelihoodestimatorsoftheparameters.Independence of sample mean vector and sample covariance matrix. Assessing the assumptionsofnormalityQ-Qplot,chi-squareplot,transformationstonear normality.(13hrs)

Unit 2: Inference problems in multivariate normal distribution, HotellingsT^2, MahalanobisD^2statistics, likelihood ratio tests for collinearity, q-sample problem.Roy'sunionandintersectiontest.Testforsymmetry.Confidenceregions, simultaneous confidence statements.Independence of subvectors, sphericitytest.Wishart matrix, statement of Wishart distribution, its properties and applications.(13hrs)

Unit3: Principalcomponentanalysis(PCA)-definitionandproperties,graphing the principal components, sample principal components, interpretation of zero, small and repeated eigenvalues, component loadings and component correlations, the problem of scaling, tests of hypotheses. Canonical correlation analysis - canonical variates and canonical correlations, sample canonical variates,samplecanonicalcorrelations,inferenceproblems. Factoranalysis-or- thogonal factor model, factor loadings, estimation of factor loadings, factor scores.(13hrs)

Unit 4: Classification and discrimination problems - concepts ofseparation and classification, Bayes and Fisher's criteria, classification rules based on expectedcostofmisclassification(ECM)andtotalprobabilityofmisclassification(TPM),classification withtwomultivariatenormalpopulations(equal and unequal covariance matrices), evaluating classification rules, classification withseveralpopulations, Fisher'slineardiscriminantfunction,testsassociatedwithdiscriminantfunctions.ClusterAnalysis-distancesandsimilarity measures, hierarchical clustering methods,*k*-meansmethod.(13hrs)

Books for Reference:

1. Anderson, T. W. (1984). *An Introduction to Multivariate Analysis* (2nd ed.). JohnWiley.
2. Flury, B. (1997). *A First Course in Multivariate Statistics*. Springer Texts in Statistics.
3. Kshirasagar, A. M. (1972). *Multivariate Analysis*. MarcelDekker.
4. Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivatiate Analysis*. AcademicPress.
5. Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley Eastern.
6. Johnson, R. A. and Wichern, D. W. (1986). *Applied Multivariate Statistical Analysis* (6th ed.). Prentice Hall ofIndia.
7. Rencher, A. C. (2003). *Methods of Multivariate Analysis*.Wiley.

Practicals based on Multivariate Analysis

1. Graphical representation of multivariatedata.
2. Model sampling from multivariate normaldistribution.
3. Applications of Hotelling'sT^2.
4. Testing equality of covariancematrices.
5. Principal componentanalysis.
6. Canonical correlationanalysis.
7. Classification.
8. Discriminantanalysis.
9. Clusteranalysis.

STS 535 - Stochastic Processes (3 Credits)

Rationale/Learning Objectives:

1. Thiscourseprovidestheoreticalknowledgeonmodellingforsequenceofnon-independent randomvariables,whichareextensivelyusedintheanalysisoftimedependentdata.

Course Outcomes:

CO1: Elucidate the power of stochastic processes and their range of applications.

CO2: Exhibit theoretical knowledge on modelling for sequence of non-independent random variables, which are extensively used in the analysis of time dependent data.

CO3: Demonstrate essential stochastic modelling tools including Markov chains, renewal theory and branching process.

CO4: Formulate and solve problems which involve setting up stochastic models.

Unit1: Introductiontostochasticprocesses-classificationaccordingtostate spaceandtimedomain.Stationaryprocess-weaklystationaryandstrongly stationary processes. Countable state Markov chains (MCs), Chapman Kol-mogorovequations,calculationof*n*-steptransitionprobabilityanditslimit. Stationarydistribution,classificationofstates,randomwalkandgamblersruin problem, estimation of TPM for finite statesofMC.(10hrs)

Unit2: Discrete state space continuous time MC, Kolmogorov-Feller differ- ential equations, Poisson process, birth and death process, applications to queues. Wiener process as a limit of random walk, first passage time and otherproblems.(10hrs)

Unit3: Renewaltheory-elementaryrenewaltheoremandapplications.State- ment and uses of key renewal theorem, study of residual life time process. (10hrs)

Unit 4: Branching process - Galton-Watson branching process, probability ofultimateextinction,distributionofpopulationsize.Martingaleindiscrete time-definitionandelementaryproperties,convergencetheorem,applications.(10hrs)

Books for Reference:

1. Basu, A. K. (2003). *Introduction to Stochastic Processes*. Narosa Publications.
2. Bhat, B. R. (2000). *Stochastic Models: Analysis and Applications*. NewAge International.
3. Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*. AcademicPress.
4. Medhi, J. (1982). *Stochastic Processes*. WileyEastern.
5. Ross,S.M.(1983).*StochasticProcesses*.JohnWiley&Sons.
6. Lawler, G. F. (2006). *Introduction to Stochastic Processes* (2nd ed.). Chap- man andHall.

Practicals based on Stochastic Processes

1. Realization of stochasticprocesses.
2. Calculation of *n*-step transitionprobabilities.
3. Classificationofstatesandmeanrecurrencetimeofstate.
4. SimulationofMarkovchainandestimatingthestationarydistributionofer- godicMarkovchain.
5. Simulation of Poissonprocesses.
6. Realization of queues and computations of typical eventslimiting.
7. Simulation of branching process and estimating its mean andvariance.

Practicals based on Machine Learning with Python

1. Fashion trends online (FTO) is an e-commerce company that sells women apparel.Itisobservedthat10%oftheircustomersreturntheitemspurchased by them for many reasons (such as size, color and material mismatch). On a specific day, 20 customers purchased items from FTO. Write program to answer thefollowing:
   (a) Probabilitythatexactlyfivecustomerswillreturntheitems.
   (b) Probability that a maximum of five customers will return theitems.
   (c) Probabilitythatmorethanfivecustomerswillreturntheitemspurchased bythem.
   (d) Averagenumberofcustomerswhoarelikelytoreturntheitemsandthe variance and the standard deviation of the number ofreturns.
2. Thenumberofcallsarrivingatacallcenterfollowsa Poissondistributionat 10 calls per hour. Write program to answer thefollowing:

(a) Calculatetheprobabilitythatthenumberofcallswillbemaximumfive.

(b) Calculate the probability that the number of calls over a 3 hour period will exceed30.

3. Aspersurveyofpesticidesamong1000farmersingrapefarmingforaround 10 acres of grape farmland, it was found that the grape farmers spray 38 litersofpesticidesinaweekonanaveragewiththecorrespondingstandard deviationof5liters.Assumethatthepesticidesprayperweekfollowsanormal distribution. Write program to answer the followingquestions:

   (a) Whatproportionofthefarmersissprayingmorethan50litersofpesticide in aweek?

   (b) Whatproportionofthefarmersissprayinglessthan10liters?

   (c) Whatproportionofthefarmersissprayingbetween30litersand60liters?

4. Designapythonprogramtoperformprincipalcomponentanalysisforasample training dataset.

5. Designapythonprogramtoperformdiscriminantanalysisforasampletrain- ing dataset.

6. Designaprogramtoimplementthesimplelinearregressionmodelforasample trainingdatasetstoredasaCSVfile.

7. Design a program to implement the multiple linear regression model for a sampletrainingdatasetstoredasaCSVfile.

8. DesignaprogramtoimplementtheBayesianclassifierforasampletraining datasetstoredasaCSVfile.

9. Writeaprogramtoimplementthegradientdescentalgorithmforpredicting future sales using the datasetAdvertising.csv.

10. Designaprogramtoimplementthek-meansclusteringforasampletraining datasetstoredasaCSVfile.

IVSemester

STH 541 - Time Series Analysis (4 Credits)

Rationale/Learning Objectives:

This course provides theoretical knowledge on the developments and applications of various techniques used in analyzing time series data and also forecasting.

Unit1: Simpledescriptivetechniques-timeseriesplots,trend,seasonaleffect. Testsfortrendandseasonality-estimationandeliminationoftrendandseasonalcomponents.Exponentialandmovingaverages moothing.Timeseries asdiscreteparameterstochasticprocess.Stationarity,autocovarianceandautocorrelationfuncti onandtheirproperties.Partialautocorrelationfunction.

(13hrs)

Unit 2: Probability models - White noise model, random walk, linear processes,movingaverage(MA),autoregressive(AR),ARMAandARIMAmodels, invertibility, ACF and PACF of these processes. Spectral properties of stationarymodels-periodogram,spectrum.(13hrs)

Unit3: Spectraldensityfunction-estimationofspectraldensitiesofAR,MA and ARMA models. Sample ACF and PACF for model identification.Model building-estimationofmean,autocovariancefunctionandautocorrelation function.EstimationinARmodels,Yule-Walkerequations,estimationinMA model and ARMA models.Order selection in AR and MA models. (13hrs)

Unit4: Forecasting-forecastmeansquareerror(FMSE),leastsquaresprediction,BLUP,innovationalgorithm.Box-JenkinsforecastingforARMAMod- els.Forecasting through exponential smoothing and Holt Winters smoothing.Residual analysis and diagnostic checking.Non-stationary time series models andtheiridentification.IntroductiontoARCHandGARCHmodels. (13hrs)

Books for Reference:

1. Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: HoldenDay.
2. Brockwell, P. J. and Davis, R. S. (2002). *Introduction to Time series and Forecasting*(2nd ed.).Springer.
3. Chatfield,C.(1996).*TheAnalysisofTimeSeries:AnIntroduction*.Chapman Hall.
4. Kendall, M. G. and Ord, J. K. (1990). *Time Series* (3rd ed.). EdwardArnold.
5. Montgomory, D. C. and Johnson, D. A. (1977). *Forecasting and Time Series Analysis*. McGrawHill.
6. Tanaka, K. (1996). *Time Series Analysis*. WileySeries.
7. Tsay, R. S. (2005). *Analysis of Time series*. John Wiley &Sons.

Practicals based on Time Series Analysis

1. Timeseriesplotsandeliminationoftrendandseasonality.
2. Estimation of ACF andPACF.
3. Model identification and estimation of ARMAmodel.
4. Model identification and estimation of ARIMAmodel.

STH 542 - Reliability and Survival Analysis (4 Credits)

Rationale/Learning Objectives:

Thiscourseprovidestheoreticalknowledgeonreliabilitytechniquesandstatistical lifetime models, which are being used in medical sciences andindustries.

Unit 1: Coherent structures, representation of coherent systems in terms of paths and cuts, duals systems, modules of coherent systems. Reliability of systemofindependentofcomponents,associationofrandomvariables,bounds on system reliability, improved bounds on system reliability using modular decompositions,lifetimedistributionof*k*outof*n*system.(13hrs)

Unit2: Measuresofreliability,survival/failurerate,hazardfunction,cumulativehazardfunction ,lackofmemoryproperty,graphsofthesystemreliability functions.Notionofaging,lifedistributionsofcoherentsystems,classesoflife distributions - parametric and nonparametric models, mean residual lifetime withsurvivalfunction.NBU,NBUE,NWU,NWUEclassesoflifedistributions andtheirimplications.(13hrs)

Unit 3: Complete and censored samples, type I, II and random censoring, lifedistributions-Exponential,Gamma,Weibull,Lognormal,Paretofamily. Estimationofparameterforexponentialandgammadistributionundervariouscensoringsituatio ns.ConfidenceintervalforparametersofExponential, Weibull, and Lognormal distributions.Wald, Score and LR tests for Exponential against GammaandWeibull.(13hrs)

Unit 4: Estimation of survival function - Kaplan-Meier estimator, Nelson-Aalenestimator,Greenwoodsformula. Otherlifetableestimators.Actuarial method of estimation of survival function. Semi-parametric regression for failure rate, Cox's proportional hazards model with one and more number of covariates, log likelihood function, log linear hazards, test for regression coefficients with andwithoutties.(13hrs)

Books for Reference:

1. Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart and WinstonInc.
2. Barlow, R. E. and Proschan, F. (1996). *Mathematical Theory of Reliability*. JohnWiley.
3. Tobias,P.A.andTrindane,D.C.(1995).*AppliedReliability*(2nded.).CRC Press.
4. Lawless, J. R. (1982). *Statistical Models and Methods for LifetimeData*.
5. Bain,L.J.andEngelhardt(1991).*StatisticalAnalysisofReliabilityandLife TestingData*.
6. Zacks,S.(1992).*IntroductiontoReliabilityAnalysis:ProbabilityModelsand Statistical Methods*.Springer.
7. Cox, D. R. Oakes, D. (1984). *Analysis of Survival Data*. New York: Chapman andHall.
8. Kalbfeish, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure TimeData*(2nded.).JohnWiley&Sons,Inc.
9. Deshpande, J. V. and Purohit, S. G. (2005). *Lifetime Data: Statistical Models and Methods*. WorldScientific.

STS 543 - Statistical Modelling (3 Credits)

Rationale/Learning Objectives:

Thiscourseprovidestheoreticalknowledgeon Bayesianstatistics,nonparametric techniques and generalized linear models for data analysis andinference.

Course Outcomes;

CO1: Exhibit theoretical knowledge on Bayesian and non-parametric techniques for the data analysis and inference.

CO2: Explain the Bayesian framework for data analysis and demonstrate when the Bayesian approach can be beneficial.

CO3: Understand the importance of some advanced statistical concepts such as non-parametric density estimation, non-parametric regression and resampling techniques.

CO4: Exhibit theoretical knowledge on some advanced regression techniques such as logistic, multilogit, count data, and log linear regression and understand its applications.

Unit 1: Introduction to Bayesian theory and philosophy - loss function and risk, foundations of optimal decision making, Bayes rule, minimax rule, admissibility. Prior and posterior distributions, conjugate families, non-informative priors - uniform and Jeffrey's prior. Bayesian estimation.Introduction to crediblesets,Bayesianhypothesistesting,Bayesianprediction.(15hrs)

Unit2: Nonparametricdensityestimation(kernelbased),nonparametricregressiontechniques-kernel,nearestneighbour,localpolynomial(LOESSregression) and spline based methods. Concept of Resampling techniques -boot- strap and jackknife methods. Bootstrap procedure - hypothesis testing and bootstrapintervalestimation.(12hrs)

Unit3: Introductiontogeneralizedlinearmodel(GLM),logisticregression, multilogit regression, count data regression, loglinearregression.(13hrs)

Books for Reference:

1. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York:Springer-Verlag.
2. Ghosh, J. K., Delampady, M. and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer Texts in Statistics.
3. Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Massachusetts: Addison-Wesley,Reading.
4. Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.
5. Hardle, W. (1991). *Smoothing Techniques*. Springer Science & Business Media.
6. Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer Science & BusinessMedia.
7. Wasserman, L. (2005). *All of Nonparametric Statistics*. Springer Science & BusinessMedia.
8. Dobson, A. J. (1983). *Introduction to Statistical Modelling*. Chapman and Hall.
9. Agresti, A. (1990). *Categorical Data Analysis* (3rd ed.).Wiley.
10. Myers, R. H., Montgomery, D. C., Vining, G. G. and Robinson, T. J. (2010). *Generalized Linear Models: with Applications in Engineering and the Sciences* (2nd ed.). John Wiley &Sons.
11. Davison, A. C. and Hinkley, D. V. (1991). *Bootstrap Methods and their Application*. Cambridge UniversityPress.
12. Higgins, J. J. (2004). *An Introduction to Modern Nonparametric Statistics*. Brooks/Cole.

STS 544 - Big Data Analytics (3 Credits)

Rationale/Learning Objectives:

1. Thiscourseenablesthestudentstounderstandaboutbigdataand Hadoopecosys- tem tools for large dataanalytics.

Course Outcomes:

CO1: Be able to understand the theoretical aspects involved in big data.

CO2: ImplementHadoop ecosystem tools in solving big data problems.

CO3: Explore the concepts and techniques involved in business intelligence used for decision making purpose.

CO4: Show an understanding of theoretical knowledge on various data mining techniques such as neural networks, association rule mining, text mining, web mining and social network analysis.

Unit 1: Introduction to Big Data - Classification of Digital Data, Character- istics of Data, Evolution of Big Data, Challenges with Big Data, Business Intelligence Vs Big Data, A Typical Data Warehouse Environment. BigData Analytics - Introduction to Big Data Analytics, Classification of Analytics, Importance of Big Data Analytics, Data Science, Terminology used in Big DataEnvironment.(10hrs)

Unit 2: Hadoop - Hadoop Distributed File System Basics, HadoopMapRe- duce Framework, MapReduce Programming. Hadoop Essential Tools - Apache HIVE, Apache PIG, Sqoop,ApacheFlume.(10hrs)

Unit3: BusinessIntelligenceConceptsandApplication,DataWarehousing, Data Mining,DataVisualization.(10hrs)

Unit4: ArtificialNeuralNetwork,AssociationRuleMining,TextMining,Web Mining, SocialNetworkAnalysis.(10hrs)

Books for Reference:

1. Acharya, S. and Chellappan, S. (2015). *Big Data and Analytics*. Wiley Pub- lications.
2. Eadline, D. (2016). *Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem* (1st ed.). Pearson Education. ISBN-13:978-9332570351.
3. Maheshwari, A. (2017). *Data Analytics* (1st ed.). McGraw Hill Education. ISBN-13:978-9352604180.
4. White, T. (2015). *Hadoop: The Definitive Guide*. (4th ed.). O'Reilly Media. ISBN-13:978-9352130672.
5. Lam, C. *Hadoop in Action*. Manning. ISBN9781935182191.

Artificial Intelligence (3 Credits)

Rationale/Learning Objectives:

This course enables the students to identify the problems where artificial intelligence is required and the different methods available.

Unit1: Whatisartificialintelligence?Problems,ProblemSpacesandsearch, Heuristicsearchtechnique.(10hrs)

Unit 2: Knowledge Representation Issues, Using Predicate Logic, representing knowledgeusingRules,SymbolicReasoningunderUncertainty.(10hrs)

Unit 3: Statistical reasoning, Weak Slot and Filter Structures, Strong slot- and-fillerstructures.(10hrs)

Unit4: GamePlaying,NaturalLanguageProcessing,Learning.(10hrs)

Books for Reference:

1. Rich, E., Knight, K. and Nair, S. B. *Artificial Intelligence* (3rd ed.). McGraw Hill.
2. Rusell, S. and Norving, P. *Artificial Intelligence: A Modern Approach* (2nd ed.). PearsonEducation.
3. Patterson, D. W. *Introduction to Artificial Intelligence and Expert Systems*. Prentice Hall ofIndia.
4. Luger, G. (2002). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (4th ed.). PearsonEducation.
5. Rolston, D. W. *Artificial Intelligence and Expert Systems Development*.Mc-GrawHill.
6. Padhy, N. P. (2015). *Artificial Intelligence and Intelligent Systems*. Oxford UniversityPress.

Elements of Statistical Computing (3 Credits)

Rationale/Learning Objectives:

This course provides foundations for statistical simulation and validation of models.

Unit1: Randomnumbergeneration,requisitesofagoodrandomnumbergenerator, methods of random number generation such as linear congruential, mixedcongruentialandmultiplicativecongruential.Testingofrandomnumbergenerator,runtest,Kolmogrov-Smirnovtest,signtest,ranktest,gaptest, digitfrequencytestandserialcorrelation,selectionofarandomnumbergenerator. Methods of generating random observations such as inverse transforms, composition, convolutionandacceptance-rejection.(10hrs)

Unit2: Simpleoptimizationmethod,directsearch,gridsearch,interpolatorysearch, gradient search. Newton-Raphson method, Muller's method, Aitken's extrapolation,simpleproblemsandapplications.(10hrs)

Unit3: Methodstocomputeintegrals-quadratureformula,doubleintegration, singularity, Gaussian integration. Monte Carlo Methods - MonteCarlo integrationandsimplecasestudies,applicationsofMonteCarlomethodsto compute expected values of functions of random variables such as Laplace transform,fouriertransformetc.,somecasestudies.(10hrs)

Unit4: Approximatingprobabilitiesandpercentagepointsinselectedprobability distribution, verification of WLLN and CLT using random number generator,simulatingnulldistributionofvarioustessstatistics,simpleapplications andcasestudies.(10hrs)

Books for Reference:

1. Kennedy, W. J. Gentle, J. E. (1980). *Statistical Computing*. MarcelDekker.
2. Sen, K. V. (1993). *Numerical Algorithm Computation in Science and Engineering*(2nd ed.). Affiliated East WestPress.
3. Law,A.M.andKelton,W.D.(2000).*Simulation,ModelingandAnalysis* (3rd ed.). Tata McGraw Hill.

4. Rajaraman,V.(1993).*Computer Oriented Numerical Methods* (4th ed.). PrenticeHall.
5. Ripley, B. D. (1987). *Stochastic Simulation*. JohnWiley.
6. Ross, S. M. (2000). *Introduction to Probability Models*. AcademicPress.
7. Ross, S. M. (2013). *Simulation*. AcademicPress.
8. Thisted,R.A.(1988).*ElementsofStatisticalComputing*.ChapmanandHall.

Practicals based on Elements of Statistical Computing

1. Generation of random numbers by acceptance rejectionmethod.
2. Generation of random numbers by linear congruentialmethod.
3. Solution of the equations using iterativemethods.
4. Numericalintegration.
5. Monte-Carlointegration.
6. Empirical distributions of the teststatistics.
7. Applications ofCLT.

Survival Analysis (3 Credits)

Rationale/Learning Objectives:

1. This course provides theoretical foundations on statistical lifetime models being used in medical sciences and industries.

Unit 1: Complete and censored samples, type I, II and random censoring, lifedistributions-Exponential,Gamma,Weibull,Lognormal,Paretofamily. Estimationofparameterforexponentialandgammadistributionundervariouscensoringsituations.ConfidenceintervalforparametersofExponential, Weibull, and Lognormal distributions.Wald, Score and LR tests for Exponential against GammaandWeibull.(10hrs)

Unit 2: Life tables - standard methods for uncensored and censored data, asymptoticpropertiesofestimatesunderarandomcensorshipmodel.Failure rate, mean residual life and their elementary properties.Estimation of survival function - Kaplan-Meier estimator, Greenwoods formula.Other life table estimators.Actuarialmethodofestimationofsurvivalfunction.(10hrs)

Unit3: Fullyparametricanalysisofdependencyacceleratedlifemodel-simple form,loglogisticacceleratedlifemodel,proportionalhazardsmodelinrelationwithacceleratedlifemodel.Semi-parametricregressionforfailurerate, Cox'sproportionalhazardsmodelwithoneandmorenumberofcovariates, loglikelihoodfunction,loglinearhazards,testforregressioncoefficientswith and withoutties.(10hrs)

Unit 4: Two sample problem - Gehan test, log rank test, Mantel-Haenszel test. Competing risks model - parametric and nonparametric inference for thesemodel.(10hrs)

Books for Reference:

1. Cox, D. R. Oakes, D. (1984). *Analysis of Survival Data*. New York: Chapman andHall.
2. Kalbfeish, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure TimeData*(2nded.).JohnWiley&Sons,Inc.
3. Lawless, J. F. (2002). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons,Inc.
4. Miller, R. G. (1981). *Survival Analysis*. John Wiley & Sons,Inc.
5. Hosmer,D.W.,Lemeshow,S.andMay,S.(2008).*AppliedSurvivalAnalysis: Regression Modeling of Time-to-Event Data* (2nd ed.). John Wiley & Sons, Inc.
6. Deshpande, J. V. and Purohit, S. G. (2005). *Lifetime Data: Statistical Models and Methods*. WorldScientific.

Practicals based on SurvivalAnalysis

1. Estimation of parameters undercensoring.
2. Construction of lifetables.
3. Test for classproperties.
4. Kaplan-Meierestimators.
5. Accelerated failure timemodel.
6. Cox proportional hazardmodel.
7. Two-sample tests undercensoring.

Stochastic Finance (3 Credits)

Rationale/Learning Objectives:

Thiscourseprovidesfoundationsonfundamentalsoffinancialmarketsandstocks andtoanalyzethedataonfinance.

Unit 1: Basic concepts of financial markets and stocks, types of traders,forwardcontractsandfutures,callandputoptions,EuropeanoptionandAmerican options. Interest rates, continuous compounding, presentvalue analysis, bond pricing, risk free interest rates. Returns, gross returns, log returns. (10hrs)

Unit2: Portfoliotheory,meanvarianceportfoliotheory.Oneriskyassetand oneriskfreeasset,tworiskyassets.Sharpesratio,tangencyportfolio,optional mix of portfolio. Market portfolio, beta, security market line, and capital asset pricing model (CAPM) and their assumption.Value at risk (VAR), nonparametric and parametric estimation of VAR, VAR for a derivative and for a portfolio of assets, delta normal method, simulation of VAR models. (10hrs)

Unit3: Financialderivatives,options,pricingviaarbitrage,lawofoneprice. Risk neutral valuation, arbitrage theorem.Convexity of cost of call option, binomialmodelsingleandmultiperiodbinomialmodel.Modelingreturns- lognormal model, random walk model, modeling through geometric Brownian motion process. Ito lemma (without proof).Arbitrage theorem. The Black Scholesformulaandassumptions,propertiesoftheBlackScholesoptioncost. (10hrs)

Unit 4: Black Scholes Merton differential equations and assumptions, the delta hedgingarbitragestrategy,volatilityandestimatingthevolatilityparameter, implied volatility. Pricing American options, pricing of an European option usingMonteCarloandpricinganAmericanoptionusingfinitedifferencemethods.Calloptionsondividendpayingsecurities.(10hrs)

Books for Reference:

1. Ross, S. M. (2003). *An Elementary Introduction to Mathematical Finance*. Cambridge UniversityPress.
2. Ruppert, D. (2004). *Statistics and Finance: An Introduction*. Springer InternationalEdition.
3. Hull, J. C. (2008). *Options, Futures and Other Derivatives*. India: Pearson Education.
4. Cuthbertson, K. and Nitzsche, D. (2001). *Financial Engineering: Derivatives andRiskManagement*.JohnWiley&SonsLtd.
5. Leuenberger, D. G. (1998). *Investment Science*. Oxford UniversityPress.
6. Wilmott, P. (2000). *Quantitative Finance*. John Wiley &Sons.
7. Tsay, R. S. (2005). *Analysis of Time series*. John Wiley &Sons.

Data Mining (3 Credits)

Rationale/Learning Objectives:

Thiscourseprovidesfoundationsonvariousstatisticalmethodsusedindataanalysis, including artificial intelligence and machine learningtechniques.

Unit 1: Data mining - motivations and importance, knowledge discovery in databases(KDD)processsearch-introduction,querying,approximationand compression. Kinds of data considered for data mining, basic data mining tasks, data mining issues. Data mining models - predictive and descriptive, inter connections between statistics, data mining, artificial intelligenceand machinelearning,applicationsofdatamining.(10hrs)

Unit2: Datamarts,databasesanddatawarehouses,OLTPsystems,multidi- mensionalmodels-datacubes,OLAPoperationsondatacubes,multidimen- sional schemes. Data Pre-processing - data cleaning, data integration, data transformation and data reduction. Visualization techniques for multidimen- sional data - scatter plot matrix, star plots, Andrews plots, Chernoff faces, parallelaxisplots.(10hrs)

Unit3: Supervisedlearning-classificationandprediction,statisticalclassification, linear discriminants, Mahalanobis linear discriminant, Fisher's linear discriminant,Bayesianclassifier,regressionbasedclassification,*k*-NN(nearest neighbour) classifier. Tree classifiers - decision trees, ID3 algorithm,CART.Artificialneuralnetworks(ANN)-thelearningproblem,perceptron,thedelta rule,multilayerfeedforwardneuralnetwork,backpropagationlearningalgo- rithm.Supportvectormachines-Lagrangianformulationandsolution,mea- suringclassifieraccuracy.(10hrs)

Unit 4: Unsupervised learning - clustering problem, similarity anddistance measures,partitioningalgorithms-*k*-means,*k*-medoids(PAM)algorithms. Density based clustering algorithms (DBSCAN). Association rule mining - marketbasketanalysis,frequentitemsets,supportandconfidenceofanassociationrule,apriorial gorithm,partitionalgorithm.(10hrs)

Books for Reference:

1. Han, J. and Kamber, M. (2002). *Data Mining: Concepts and Techniques*. USA: Morgan KaufmanPublishers.
2. Dunham, M. H. (2005). *Data Mining: Introductory and Advanced Topics*. PearsonEducation.
3. Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York:Springer.
4. Berthold, M. R. and Hand, D. J. (2003). *Intelligent Data Analysis: An Introduction*(2nd ed.).Springer.
5. J. P. Marques de Sa (2001). *Pattern Recognition: Concepts, Methods and Applications*.Springer.
6. Chattamvelli, R. (2009). *Data Mining Methods*. Narosa PublishingHouse.

Bayesian Inference (3 Credits)

Rationale/Learning Objectives:

This course provides theoretical knowledge on Bayesian techniques for data analysis and inference.

Unit1: Introductionandphilosophy-lossfunctionandrisk,foundationsof optimaldecisionmaking,Bayesrule,minimaxrule,admissibility,sufficiency andRao-Blackwellization.(10hrs)

Unit 2: Utility theory, utility and loss, personal utility function, prior and posterior, conjugate families, non-informative priors - uniform prior, Jeffrey's prior, left and rightinvariantprior.(10hrs)

Unit 3: Bayesian analysis - the posterior distribution, Bayesian estimation, credible sets, Bayesian hypothesis testing, Bayesian prediction, empiricalBayes analysis,hierarchicalBayesanalysis,Bayesianrobustness.(10hrs)

Unit 4: Bayesian computation - analytic approximation, the EM algorithm, MonteCarlosampling,MarkovChainMonteCarlomethods.(10hrs)

Books for Reference:

1. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York:Springer-Verlag.
2. Ghosh, J. K., Delampady, M. and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer Texts in Statis- tics.
3. Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Anal- ysis*. Massachusetts: Addison-Wesley,Reading.

Practicals based on BayesianInference

1. Computation of riskfunctions.
2. Computation of Bayes and minimaxrules.
3. EMalgorithm.

Statistical Methods for Reliability (3 Credits)

Rationale/Learning Objectives:

This course provides theoretical knowledge on reliability techniques for data analysis and inference.

Unit 1: Coherent structures, representation of coherent systems in terms of paths and cuts, duals systems, modules of coherent systems. Reliability of systemofindependentofcomponents,associationofrandomvariables,bounds on system reliability, improved bounds on system reliability using modular decompositions.(10hrs)

Unit2: Measuresofreliability,graphsofthesystemreliabilityfunctions.Notion of aging, life distributions of coherent systems, distributions with in- creasing failure rate average arising from shock models, preservation oflife distributionclassesunderreliabilityoperations.Reliabilitybounds,lifetime distribution of *k* out of*n*system.(10hrs)

Unit3: Classesoflifedistributions-parametricandnonparametricmodels, mean residual lifetime with survival function. Applicable in replacement models,NBU,NBUE,NWU,NWUEclassesoflifedistributionsandtheirimplications.Shockm odelsleadingtoNBU.Agereplacementandblockreplacement policies.Renewaltheoryusefulinreplacementmodels.(10hrs)

Unit4: Replacementpolicycomparisons,preservationoflifedistributionclasses underreliabilityoperations.Reversedhazardrate,cumulativereversedhazardfunction,relationb etweenhazardfunctionandreversedhazardfunction. Lack ofmemoryproperty.(10hrs)

Books for Reference:

1. Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart and WinstonInc.
2. Barlow, R. E. and Proschan, F. (1996). *Mathematical Theory of Reliability*. JohnWiley.
3. Tobias,P.A.andTrindane,D.C.(1995).*AppliedReliability*(2nded.).CRC Press.
4. Lawless, J. R. (1982). *Statistical Models and Methods for LifetimeData*.
5. Bain,L.J.andEngelhardt(1991).*StatisticalAnalysisofReliabilityandLife TestingData*.
6. Zacks,S.(1992).*IntroductiontoReliabilityAnalysis:ProbabilityModelsand Statistical Methods*.Springer.

Nonparametric Inference (3 Credits)

Rationale/Learning Objectives:

This course provides theoretical knowledge on various nonparametric testing procedures for dataanalysis.

Unit 1: Empirical distribution function, Glivenko-Cantelli theorem, Kolmogorov goodnessoffittest.OnesampleU-statistics,kernelandsymmetrickernel,two sampleU-statistics,asymptoticdistributionofU-statistics.UMVUEproperty of U-statistics, asymptotic distribution of linear function of order statistics. (10hrs)

Unit 2: Rank tests, locally most powerful rank test, linear rank statistics and their distributional properties under null hypothesis, Pitman's asymptotic relativeefficiency.(10hrs)

Unit 3: One sample location problem, sign test and signed rank test, two sampleKolmogorov-Smirnovtests.Twosamplelocationandscaleproblems. Wilcoxon-Mann-Whitney test, normal score test, ARE of various testbased linear rank statistics. Kruskal-Wallis *k*sampletest.(10hrs)

Unit4: Cox'sproportionalhazardsmodel,ranktest(partiallikelihood)forregressioncoefficients, conceptsofJackknifingmethodofquenouilleforreducing bias,bootstrapmethods,confidenceintervals.(10hrs)

Books for Reference:

1. Cox, D. R. and Oakes, D. (1983). *Survival Analysis*. Chapman andHall.
2. Davison, A. C. and Hinkley, D.V.(1991).*Bootstrap Methods and their Application*. Cambridge UniversityPress.
3. Fraser,D.A.S.(1957).*NonparametricMethodsinStatistics*.JohnWiley.
4. Gibbons, J. D. (1985). *Nonparametric Statistical Inference* (2nd ed.). Marcel Dekker.
5. Hajek,J.andSidak,Z.(1961).*TheoryofRankTests*.AcademicPress.
6. Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*.Wiley.
7. Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Non-parametric Statistics*.Wiley.

Actuarial Methods (3 Credits)

Rationale/Learning Objectives:

This course provides theoretical knowledge on various models used in insurance, risk analysis and theory of credibility.

Unit 1: Review of decision theory and actuarial applications. Loss distributions-modelingofindividualandaggregatelosses,moments,fittingdistributionstoclaimsdata,deductiblesandretentionlimits,proportionaland excess of loss reinsurance, share of claim amounts, parametric estimation with incompleteinformation.(10hrs)

Unit 2: Risk models - models for claim number and claim amount in short term contracts, moments, compound distributions, moments of insurer's and reinsurer'sshareofaggregateclaims.(10hrs)

Unit3: ReviewofBayesianstatistics-estimationandapplication.Experience rating-ratingmethodsininsuranceandbanking,claimprobabilitycalculation,stationarydistributionofproportionofpolicyholdersinvariouslevelsof discount.(10hrs)

Unit 4: Delay/run-off triangle - development factor, basic andinflation-adjusted chain-laddermethod,alternativemethods,averagecostperclaimandBorn-huetterFergusonmethodsforoutstandingclaimamounts,statisticalmodels. (10hrs)

Books for Reference:

1. Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A. and Nesbitt, C.J. (1997). *Actuarial Mathematics* (2nd ed.). Society ofActuaries.
2. Klugman,S.A.,Panjer,H.H.,Willmotand,G.E.andVenter,G.G.(1998). *Loss Models: From Data to Decisions*. John Wiley & Sons.
3. Daykin, C. D., Pentikainen, T. and Pesonen, M. (1994). *Practical Risk Theory for Actuaries*. ChapmanHall.

Pattern Recognition and Image Processing (3 Credits)

Rationale/Learning Objectives:

This course provides theoretical knowledge on various statistical techniques involved in pattern recognition and image processing for data analysis and inference.

Pattern Recognition

Unit1: ReviewofBayesclassification-errorprobability,errorbounds,Bhat-tacharyabounds,errorratesandtheirestimation.Parametricandnonpara- metric learning, density estimation.Classification trees.*k*-NN rule and its errorrate.(8hrs)

Unit2: Neuralnetworkmodelsforpatternrecognition-learning,supervised andunsupervisedclassification.Unsupervisedclassification-split/mergetech-niques,hierarchicalclusteringalgorithms,clustervalidity,estimationofmix-turedistributions.(8hrs)

Unit 3: Feature selection - optimal and suboptimal algorithms. Some of the other approaches like the syntactic, the fuzzy set theoretic, the neurofuzzy, the evolutionary (based on genetic algorithms), and applications. Some recent topicslikedatamining,supportvectormachines,etc.(8hrs)

Books for Reference:

1. Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition* (2nd ed.). New York: AcademicPress.
2. Devijver, P. A. and Kittler,J.(1982).*Pattern Recognition: A Statistical Approach*. PrenticeHall.
3. Jain, A. K. and Dube, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall.
4. Everitt, B. S. (1993). *Cluster Analysis*. Halsted Press.
5. Fu, K. S. (1982). *Syntactic Pattern Recognition and Applications*. Prentice Hall.
6. Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. PlenumPress.
7. Hastie, T., Tibshirani, R. and Friedman, J. H. (2001). *Elements of Statistical Learning*.Springer-Verlag.
8. Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge UniversityPress.
9. Theodoridis, S. and Koutroumbas, K. (1999). *Pattern Recognition*. Academic Press.

Image Processing

Unit4: Introduction,imagedefinitionanditsrepresentation.TypicalIPoperations like enhancement, contrast stretching, smoothing andsharpening, greylevelthresholding, edge detection, medial axis transform, skeletonization/thinning,warping.(8hrs)

Unit 5: Segmentation and pixel classification, object recognition, some statistical(includingBayesian)approachesfortheabove,likeBesag'sICMalgorithm, deformabletemplatesapproachofGrenander.(8hrs)

Books for Reference:

1. Young, T. Y. and Fu, K. S. (1986). *Handbook of Pattern Recognition and Image Processing*. Vols. 1 & 2, AcademicPress.
2. Jain, A. (1989). *Fundamentals of Digital Image Processing*. PrenticeHall.
3. Castleman, K. R. (1996). *Digital Image Processing*. PrenticeHall.
4. Mardia, K. V. and Kanji, G. K. (1993). *Statistics and Images*.Carfax.

Operations Research (3 Credits)

Rationale/Learning Objectives:

Thiscourseprovidestheoreticalfoundationsonoptimizationtechniquesformanagerial decision makingprocess.

Unit 1: Linear programming problem (LPP) - definition, formulation. Simplexmethod-canonicalform,improvingnon-optimalbasicfeasiblesolution (B.F.S), conditions for optimality, conditions for unboundedness.Convex sets, geometryofsimplexmethodextremepointandB.F.S,existenceofB.F.S,existence of optimal B.F.S. Two phase method, big *M*method.(10hrs)

Unit2: DualitytheoryofLPP-weakdualitytheoremanditsproperties,the fundamentaldualitytheorem,complementaryslacknesstheorem.Dualsimplex method.Sensitivity analysis.Integer programming cutting plane technique, Gomory's algorithm for pure integer program.Dynamic Programming, multistage decision making problems, Bellman's principle of optimality, recursivenatureofcomputation,application,applicationsofdynamicprogramming. (10hrs)

Unit3: Inventorytheory-natureofinventoryproblem,motivesforcarrying inventory, deterministic inventory model with decay. Probabilisticinventory models, continuous review and periodic review systems, (*s, S*) policy,heuristic solutionoflotsizereorderpointmodel((*Q,r*)policy).(10hrs)

Unit 4: Queuing theory - characteristics of queues, *M/M/*1 system, steady state solution, measures of effectiveness, waiting time distributions,Little's formula,*M/M/*1/*K*system,*M/M/C*system,machineinterferenceproblem, *M/G/*1system,PollaczekKhintchineformula.(10hrs)

Books for Reference:

1. Gross, D. and Harris, C. M. (1985). *Fundamentals of Queuing Theory* (2nd ed.). JohnWiley.
2. Hadley, G. (1975). *Linear and Combinatorial Programming*. John Wiley & Sons.
3. Murty, K. G. (1976). *Linear and Combinatorial Programming*. John Wiley & Sons.
4. Kambo, N. S. (1991). *Mathematical Programming Techniques*. Affiliated East WestPress.
5. Taha, H. A. (2001). *Operations Research: An Introduction* (6th ed.). India: PrenticeHall.
6. Sivazlian, B. D. and Stanfel, L. E. (1975). *Analysis of Systems in Operations Research*. PrenticeHall.
7. Daellenbach, H. G. and George, J. A.(1978).*Introduction to Operations Research Techniques*. Allyn and BaconInc.

Certificate Course on Microsoft Excel (Basic to Advance)

This Course helps you to master in Excel, and learn the powerful features Excel has to offer to analyze the data.

Learning Objectives:

- Understand the practicality ofexcel.
- Knowledge of formatting, functions &formulas.
- Learn to use advanced features, graphs & presentation techniques to maximizeimpact.
- Perform data cleaning, processing & manipulation techniques using superpower functions & formulas.
- Build a dashboard / summary report with dynamic charts &tables.
- How macros and VBA automate your spreadsheets and increaseinteractivity.

Course Outcome:

- Apply visual elements and advanced formulas to a worksheet to display data in various formats.
- Learn to use advanced functions & features of excel to improve productivity, enhance spreadsheets with templates, charts, graphics, and formulas and streamline the operational work.
- Automate common tasks & apply more advanced analysis techniques to more complexdata.

Course Syllabus:

Section 1: IntroductiontoExcelHours:1
a) Purpose & application of Excel, Understanding the Excel interface - Menu Options, Create & Save Spreadsheets, Save As Formats, Limitations, Insert & delete rows / columns, Printing.
b) Navigation & Editing: Moving around the spreadsheets, Entering information into cells, Types of data, Clipboard, Transformation, Hide rows/columns.
c) Protecting & Sharing: Protect sheet / workbook - Locking cells.

Section 2:DataHandlingHours:12
a) Sorting, Filters & Advanced Filters, Remove duplicates, Text to columns, Cellreference.
b) Presentation: Formatting - Cell (Alignment, Height & width, Wrapping, Merging), Numbers (Currency, %, Decimal, negative), Custom Format.
c) Conditional Formatting - Changing the format of the values depends upon the cellvalue, conditional format formulas
d) Data Cleaning - Extracting / Combining text, for typos & bugs – LEFT (), RIGHT (), LEN (), FIND ()
e) Performing Math with Date & Time: TODAY (), NOW (), DATEVALUE (), YEAR (), MONTH (), DAY (), TEXT ()
f) Lookup & Reference: VLOOKUP, HLOOKUP, INDEX, MATCH, OFFSET & NAMED RANGES, INDIRECT

g) Logical Functions: Automatic decision making - IF ELSE, AND, OR, NOT, NESTED IF ELSE.
h) Information Functions : ISERROR, ISBLANK, CELL, ISTEXT
i) Text Functions: TRIM, MID, LOWER, UPPER, PROPER, REPT,TRUNC, CONCATENATE etc…
j) Formula Evaluation: Debugging errors informula.

Section 3:DataValidationHours:1
a) Controlling user inputs to reduce the risk of error & increase efficiency – Validation Criteria (List, Date, Time, Text Length etc…)

Section 4:DataAnalysisHours:16
a) Summarizing Data - SUM () Family, COUNT () Family, AVERAGE, MEDIAN,MIN, MAX, STDEV etc…
b) Array Formulas – Perform multiple calculations in one cell - SUMPRODUCT
c) Pivot Tables & Pivot Charts, Adding Slicers - Value Field Settings, Filtering,Grouping, Sorting, Changing layout & format etc…
d) Data Visualization - Charts & Sparkline's: Static & Dynamic charts, formatting & Designing.
e) Analysis Tools - Apply various statistical methods to analyze thedata. o Correlation Analysis o OLS Regression: Simple & Multiple Linear RegressionAnalysis. o ANOVA: Single / TwoFactor o Random NumberGeneration **T - test (Paired / Twosamples)** o Understanding & Interpretation of statisticalresults.

Section 5: Visual BasicandMacroHours:10
a) Recording Macro and understanding the code behind.
b) Create macros by writing VBA scripts.
c) Creating user forms & recording the data.
d) User defined functions with VBA.
e) Adding Add-Ins in Excel.

Learning Method:

- Hands-on Training - Classroom (70%)
- Homework(10%)
- Project(20%)

Assessment:

- Practical Exam:100Marks **40**

Certificate Course on R and Excel for Data Science

This Course helps you to master in R and Excel, and learn the powerful functions and features to analyze and visualize the data effectively.

Learning Objectives:

- Learn R language fundamentals and basicsyntax.
- Learn how to program in R, How to use R for effective dataanalysis.
- Explore R syntax, functions & packages.
- Analyze real world challenges in data management; explore general practices of data science.
- Understand the practicality ofexcel.
- Knowledge of formatting, functions &formulas.
- Learn to use advanced features, graphs & presentation techniques to maximizeimpact.
- Perform data cleaning, processing & manipulation techniques using superpower functions & formulas.
- Build a dashboard / summary report with dynamic charts &tables.

Course Outcome:

- Become familiar with the major R datastructures.
- Create your own functions &visualizations.
- Learn to write your own project syntax ranging from importing data into R to apply standard and more advanced statistical analysismethods.
- Apply visual elements and advanced formulas to a worksheetto display data in various formats.
- Learn to use advanced functions & features of excel to improve productivity, enhance spreadsheetswithtemplates,charts,graphics,andformulasandstreamlinetheoperationalwork.

Course Syllabus: R Programming

Section 1: Getting StartedwithRHours:1
a) History of R, Installation of R & R studio, Loading Add-on packages, choosingrepositories, Accessing data in packages.
b) Help & Documentation – Help for Functions/Packages/Data Sets.
c) Data Types - Vectors, Lists, Matrices, Arrays, Factors, Data Frames.
d) Variables - Variable Assignment, Finding & Deleting Variables, Data Type of a Variable.
e) Operators - Arithmetic, Relational, Logical, Assignment & Miscellaneous Operators.

Section 2: ProgrammingLanguageBasicsHours6
a) Simple Manipulations: Numbers & Vectors, Vectors & Assignment, Vector Arithmetic, Logical Vectors, Character Vectors.
b) Generating Sequences & Missing Values.

c) Index Vectors: Selecting & Modifying subsets of a data set.
d) Objects, their modes & attributes: Intrinsic attributes: mode and length, changing the length of an object, class of an object.
e) Ordered and unordered factors: A specific example, The function tapply () and ragged arrays.
f) Arrays and matrices: Arrays, Array indexing - Subsections of an array, Index matrices.
g) The array () function: Mixed vector and array arithmetic. The recycling rule, The outer product of two arrays, Generalized transpose of an array.
h) Matrix facilities: Matrix multiplication, Linear equations and inversion, Forming partitioned matricescbind () and rbind (), The concatenation function c() with arrays, Frequency tables from factors.
i) Lists and Data frames: Lists, Constructing and modifying lists, Concatenating lists. Data Frames, Making data frames, working with data frames.

Section3:FunctionHours:2
a) Built-in functions (Numeric/Character/Statistical/Other), User define function, Calling function, Defining new binary operators, Assignments within functions, more advanced examples, Applying functions to matrices & data frames

Section 4: AdvancedDataManagementHours:9
a) Data Input & Output: Changing directories, Managing files & workspace, Reading data from files, writing data from R, Connection to External data sources.
b) View / Edit Data, Objects / Variable types, Converting Objects / Variables, Selecting Variables / Observations.
c) Applying Functions: lappy, sapply, tapply, apply, mapply
d) Combining Variables with c, cbind, rbind functions
e) Combining data with Vector / Matrix / Data Frame / List Function
f) Working with Date & Time
g) Finding NA / NaN& Replacing
h) Conditional Transformation / Decision Making
i) Control Flow (Repetition & Looping / ConditionalExecution)
j) Variables: Renaming Variables / Observations, Creating New / Recoding Variables,Keeping & Dropping Variables
k) Generating Random Numbers
l) Data Sets : Stacking/Concatenating/Adding Datasets, Joining / Merging DataFrames
m) Summary: Creating Summarized / Aggregated Datasets (dplyr)
n) Reshaping the data (Reshape Package)
o) Removing Duplicates, Sorting the Data Frames
p) Value Labels or Formats (and Measurement Level)

Section 5:DataVisualization	**Hours:2**
a) Traditional Graphics, Graphics with ggplot2 & Advanced Graph Types.	

Section 6:DataAnalysis	**Hours:5**
b) Basic Statistics: Descriptive Statistics, Frequency & Contingency tables, T – tests, Non-parametric tests of group differences	
c) Predictive Modeling: Correlation Analysis, Splitting data into training & validation, OLS Regression - Simple / Multiple Linear Regression.	
d) Regression diagnostics: Non-Normality, Multicolinearity, Non-linearity, Non-constant error variance.	
e) Unusual Observations & Corrective measures: Outliers, High-leverage points &Influential Observations.	
f) Choosing Best Regression Model: Comparing Models, Variable selection	
g) Model Validation: Cross-Validation	
h) Assessment of Regressors: Relative Importance.	

Course Syllabus: Excel

Section 1: IntroductiontoExcel	**Hours:1**
a) Purpose & application of Excel, Understanding the Excel interface - Menu Options, Create & Save Spreadsheets, Save As Formats, Limitations, Insert & delete rows /columns, Printing.	
b) Navigation & Editing: Moving around the spreadsheets, Entering information into cells, Types of data, Clipboard, Transformation, Hide rows/columns.	
c) Protecting & Sharing: Protect sheet / workbook - Locking cells.	

Section 2:DataHandling	**Hours:7**
a) Sorting, Filters & Advanced Filters, Remove duplicates, Text to columns, Cellreference.	
b) Presentation: Formatting - Cell (Alignment, Height & width, Wrapping, Merging), Numbers (Currency, %, Decimal, negative), Custom Format.	
c) Conditional Formatting - Changing the format of the values depends upon the cell value, conditional format formulas	
d) Data Cleaning - Extracting / Combining text, for typos & bugs – LEFT (), RIGHT (), LEN (), FIND ()	
e) Performing Math with Date & Time: TODAY (), NOW (), DATEVALUE (), YEAR (), MONTH (), DAY (), TEXT ()	
f) Lookup & Reference: VLOOKUP, HLOOKUP, INDEX, MATCH	

g) Logical Functions: Automatic decision making - IF ELSE, AND, OR, NOT, NESTED IF ELSE.
h) Information Functions : ISERROR, ISBLANK, CELL, ISTEXT
i) Text Functions: TRIM, MID, LOWER, UPPER, PROPER, REPT, TRUNC, CONCATENATE etc…

Section 3:DataAnalysis **Hours:7**
a) Summarizing Data - SUM () Family, COUNT () Family, AVERAGE, MEDIAN,MIN, MAX, STDEV etc…
b) Array Formulas – Perform multiple calculations in one cell - SUMPRODUCT
c) Pivot Tables & Pivot Charts, Adding Slicers - Value Field Settings, Filtering,Grouping, Sorting, Changing layout & format etc…
d) Data Visualization - Charts & Sparkline's: Static & Dynamic charts, formatting & Designing.
e) Analysis Tools - Apply various statistical methods to analyze thedata. • Correlation Analysis • ANOVA : Single / TwoFactor • Random NumberGeneration • T - test (Paired / Two samples) • Understanding & Interpretation of statisticalresults.

Learning Method: Total Hours Required: 40

Certificate Course on R for Data Science

This course will offer you to learn Data Science in R from scratch.

Learning Objectives:

- Learn R language fundamentals and basicsyntax.
- Learn how to program in R, How to use R for effective dataanalysis.
- Explore R syntax, functions & packages.
- Analyze real world challenges in data management; explore general practices of data science.

Course Outcome:

- Become familiar with the major R datastructures.
- Create your own functions &visualizations.
- Learn to write your own project syntax ranging from importing data into R to apply standard and more advanced statistical analysismethods.

Course Syllabus: R Programming

Section 1: Getting StartedwithR	**Hours:2**
a) History of R, Installation of R & R studio, Loading Add-on packages, choosingrepositories, Accessing data in packages.	
b) Help & Documentation – Help for Functions/Packages/Data Sets.	
c) Data Types - Vectors, Lists, Matrices, Arrays, Factors, Data Frames.	
d) Variables - Variable Assignment, Finding & Deleting Variables, Data Type of a Variable.	
e) Operators - Arithmetic, Relational, Logical, Assignment & Miscellaneous Operators.	

Section 2: ProgrammingLanguageBasics	**Hours9**
a) Simple Manipulations: Numbers & Vectors, Vectors & Assignment, Vector Arithmetic, Logical Vectors, Character Vectors.	
b) Generating Sequences & Missing Values.	
c) Index Vectors: Selecting & Modifying subsets of a data set.	
d) Objects, their modes & attributes: Intrinsic attributes: mode and length, changing the lengthof an object, class of an object.	
e) Ordered and unordered factors: A specific example, The function tapply () and ragged arrays.	
f) Arrays and matrices: Arrays, Array indexing - Subsections of an array, Index matrices.	
g) The array () function: Mixed vector and array arithmetic. The recycling rule, The outer product of two arrays, Generalized transpose of an array.	
h) Matrix facilities: Matrix multiplication, Linear equations and inversion, Forming partitioned matrices cbind () and rbind (), The concatenation function c() with arrays, Frequency tables	

from factors.
i) Lists and Data frames: Lists, Constructing and modifying lists, Concatenatinglists. Data Frames, Making data frames, working with data frames.

Section3:Functions **Hours:3**
a) Built-in functions (Numeric/Character/Statistical/Other), User define function, Calling function, Defining new binary operators, Assignments within functions, more advanced examples, Applying functions to matrices & data frames

Section 4: AdvancedDataManagement **Hours:12**
a) Data Input & Output: Changing directories, Managing files & workspace, Reading data from files, writing data from R, Connection to External data sources.
b) View / Edit Data, Objects / Variable types, Converting Objects / Variables, Selecting Variables / Observations.
c) Applying Functions: lappy, sapply, tapply, apply, mapply
d) Combining Variables with c, cbind, rbind functions
e) Combining data with Vector / Matrix / Data Frame / List Function
f) Working with Date & Time
g) Finding NA / NaN& Replacing
h) Conditional Transformation / Decision Making
i) Control Flow (Repetition & Looping / ConditionalExecution)
j) Variables: Renaming Variables / Observations, Creating New / Recoding Variables,Keeping & Dropping Variables
k) Generating Random Numbers
l) Data Sets : Stacking/Concatenating/Adding Datasets, Joining / Merging DataFrames
m) Summary: Creating Summarized / Aggregated Datasets (dplyr)
n) Reshaping the data (Reshape Package)
o) Removing Duplicates, Sorting the Data Frames
p) Value Labels or Formats (and Measurement Level)

Section 5:DataVisualization **Hours:5**
a) Traditional Graphics, Graphics with ggplot2 & Advanced Graph Types.

Section 6:DataAnalysis	**Hours:9**
b) Basic Statistics: Descriptive Statistics, Frequency & Contingency tables, T – tests, Non-parametric tests of group differences	
c) Predictive Modeling: Correlation Analysis, Splitting data into training & validation, OLS Regression - Simple / Multiple Linear Regression.	
d) Regression diagnostics: Non-Normality, Multicollinearity, Non-linearity, Non-constant error variance.	
e) Unusual Observations & Corrective measures: Outliers, High-leverage points &Influential Observations.	
f) Choosing Best Regression Model: Comparing Models, Variable selection	
g) Model Validation: Cross-Validation	
h) Assessment of Regressors: Relative Importance.	

Learning Method:

- Classroom(70%)
- Homework(10%)
- Project(20%)

Assessment:

- Practical Exam:100Marks **40**

Certificate Course on R for Advanced Statistical Methods & Machine Learning

This course will offer you to learn how to apply advanced statistical methods & machine learning algorithms.

Learning Objectives:

- Apply advanced statistical methods which include discovery & exploration of complex multivariate relationships amongvariables.

Course Outcome:

- Become familiar with the major R datastructures.
- Create your own functions &visualizations.
- Learn to write your own project syntax ranging from importing data into R to apply standard and more advanced statistical analysismethods.

Course Syllabus:

Section 1: Advanced Statistical MethodsUsingR	**Hours:20**
a) Generalized Linear Models: Logistic Regression, Multinomial Logistic Regression, Poisson Regression	
b) Ridge Regression	
c) Forecasting: Time Series Analysis – ARIMA	
d) Cluster Analysis: Hierarchical & Portioning cluster analysis (K- Means)	
e) Classification: Decision Tree & CHAID	
f) Text Mining: Word Cloud	
g) Dimensionality Reduction: PCA & Factor Analysis	

Section 2: Machine LearningwithR	**Hours:20**
a) Classification: Support Vector Machine, Random Forest Method, Navie Bayes	
b) Gradient Boosting Model	
c) Artificial Neural Network – Single Layer & Multiple Layer	

Learning Method:

- Classroom(70%)
- Homework(10%)
- Project(20%)

Assessment: Practical Exam: 100 Marks