Syllabus
for
4-years B.Sc. in Statistics under NEP-2020

(Effective from Academic Year 2023-24)

Department of Mathematics and Statistics
Aliah University
IIA/27, New Town
Kolkata-160

Outline of 4-yr UG in Statistics (Major & Minor) w.e.f. 2023-24

Semester-I					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC1101	Probability-I	4	0	4
2	STAUGMCC1102	Demography and Vital Statistics	3	1	4
Minor Courses					
3	STAUGMIN1101	Probability	4	0	4
Multi-disciplinary					
4	UCCUGMDC1101	Arabic and Islamic Studies	3	0	3
Ability Enhancement Course (AEC)					
5	UCCUGAEC1101	Bengali/Urdu/Hindi	4	0	4
Skill Enhancement Course (SEC)					
6	STAUGSEC1101	Presentation Skill and Introduction to R	0	3	3
Total					22

Semester-II					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC1203	Descriptive Statistics	3	1	4
2	STAUGMCC1204	Probability-II	3	1	4
Minor Courses					
3	STAUGMIN1202	Descriptive Statistics	3	1	4
Multi-disciplinary					
4	STAUGMDC1202	Economic Statistics and Official Statistics	3	0	3
Value Added Course (VAC)					
5	UCCUGVAC1201	Environmental Science	4	0	4
Skill Enhancement Course (SEC)					
6	STAUGSEC1202	Statistical computations using R	0	3	3
Total					22

Semester-III					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC2305	Inference-I	3	1	4
2	STAUGMCC2306	Sampling Distributions	3	1	4
Minor Courses					
3	STAUGMIN2303	Inference	3	1	4
Multi-disciplinary					
4	STAUGMDC2303	Statistics in Ecology and Epidemiology	3	0	3
Ability Enhancement Course (AEC)					
5	UCCUGAEC2302	English	4	0	4
Skill Enhancement Course (SEC)					
6	STAUGSEC2303	Statistical Computations using Python	0	3	3
Total					22

Semester-IV					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC2407	Sample Survey and Time Series-I	3	1	4
2	STAUGMCC2408	Linear Models	3	1	4
3	STAUGMCC2409	Regression Analysis	3	1	4
Minor Courses					
4	STAUGMIN2404	Sample Survey and Time Series-I	3	1	4
Value Added Courses (VAC)					
5	UCCUGVAC2402	Understanding India	4	0	4
Total					20

Semester-V					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC3510	Multivariate Analysis-I	3	1	4
2	STAUGMCC3511	Inference-II	3	1	4
3	STAUGMCC3512	Mathematical Analysis	3	1	4
Minor Courses					
4	STAUGMIN3505	Multivariate Analysis-I	3	1	4
Internship (SIP)					
5	STAUGSIP3501	Internship			4
Total					20

Semester-VI					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC3613	Survival Analysis	3	1	4
2	STAUGMCC3614	Inference-III	3	1	4
3	STAUGMCC3615	Design of Experiments	4	0	4
Discipline Specific Elective (DSE)					
4	STAUGMDS3601	Statistical Quality Control / Econometrics /Data Structure and Database Management	3	1	4
Minor Courses					
5	STAUGMIN3606	Survival Analysis	3	1	4
Total					20

Semester-VII (with Research)					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC4716	Sample Survey and Time Series-II	3	1	4
2	STAUGMCC4717	Clinical Trials and Epidemiology	3	1	4
Discipline Specific Elective (DSE)					
3	STAUGMDS4702	1. Statistical Learning with Big Data-I 2. Reliability Theory 3. Advanced Probability	3	1	4
4	STAUGMDS4703	1. Advanced Data Analytic Techniques 2. Statistical Genetics and Ecology 3. Demography	3	1	4
Minor Courses					
5	STAUGMIN4707	Statistical Learning with Big Data-I	3	1	4
Dissertation					
6	STAUGPRJ01	Research work			4
Total					24

Semester-VIII (with Research)					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC4818	Stochastic Processes	3	1	4
2	STAUGMCC4819	Multivariate Analysis-II	3	1	4
Discipline Specific Elective (DSE)					
3	STAUGMDS4805	1. Statistical Learning with Big Data-II 2. Statistical Methods for Bio-computing 3. Actuarial Statistics	3	1	4
Minor Courses					
4	STAUGMIN4808	Statistical Learning with Big Data-II	3	1	4
Dissertation					
5	STAUGPRJ01	Research work (continued)			8
Total					24
4-yrs UG honors without Research					

Semester-VII (without Research)					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC4716	Sample Survey and Time Series-II	3	1	4
2	STAUGMCC4717	Clinical Trials and Epidemiology	3	1	4
Discipline Specific Elective (DSE)					
3	STAUGMDS4702	1. Statistical Learning with Big Data-I 2. Reliability Theory 3. Advanced Probability	3	1	4
4	STAUGMDS4703	1. Advanced Data Analytic Techniques 2. Statistical Genetics and Ecology 3. Demography	3	1	4
5	STAUGMDS4704	1. Statistical Computation using SAS 2. Industrial Statistics 3. Decision Theory			4
Minor Courses					
6	STAUGMIN4707	Statistical Learning with Big Data-I	3	1	4
Total					24

Semester-VIII (without Research)					
Sl. No.	Course Code	Course Title	Credit		
			Theory	Practical	Total
Major Courses					
1	STAUGMCC4818	Stochastic Processes	3	1	4
2	STAUGMCC4819	Multivariate Analysis-II	3	1	4
Discipline Specific Elective (DSE)					
3	STAUGMDS4805	1. Statistical Learning with Big Data-II 2. Statistical Methods for Bio-computing 3. Actuarial Statistics	3	1	4
4	STAUGMDS4806	1. Advanced Regression Analysis 2. Linear Programming Problems and Game Theory 3. Machine Learning	3	1	4
5	STAUGMDS4807	1. Simulation and Data Analysis 2. Financial Statistics 3. Data Mining Techniques	3	1	4
Minor Courses					
6	STAUGMIN4808	Statistical Learning with Big Data-II	3	1	4
Total					24

Detailed Syllabus:

Semester-I

STAUGMCC1101 Probability-I / STAUGMIN1101 Probability

Course objectives	In this course students will learn 1. Various concepts of probability: definition of probability, independence of two or more events, finding probabilities 2. Random variables and its distributions 3. Expectations and moments of a random variable 4. Probability inequalities and their applications 5. Generating function and its role as an alternate characterization of a distribution. 6. Characterization and identification of distribution using generating functions 7. Deriving distributions of some linear function of a set of independent random variables
Learning Outcomes	At the end of the course a student will be able to 1. Find probability using set theory, permutations, and combinations 2. Solve problems related to the distribution of a random variable, cumulative distribution function etc. 3. Find various moments of a distribution and their role 4. Solve problems related to probability inequalities 5. Solve problems related to generating functions 6. Derive the distribution of sum and average of independent random variables 7. Identify a distribution from generating functions

Detailed Syllabus:

Theory	
Unit 1	**20L**
Probability: Introduction, random experiments, sample space, events, and algebra of events. Definitions of Probability – classical, statistical regularity. Limitations of Classical definition. Probability of union and intersection of events, Probability of occurrence of exactly m and at least m events out of n events, Examples based on classical approach and repeated trials, Kolmogorov's Axiomatic definition.	
Unit 2	**10L**
Conditional Probability, laws of addition and multiplication, theorem of total probability, Bayes' theorem and its applications, independent events.	
Unit 3	**20L**
Random variables, distribution function and properties, p.m.f., p.d.f., illustrations and properties of random variables. Introduction to some important distributions: uniform, binomial, Poisson, geometric, exponential, normal. Mathematical Expectation and properties. Probability generating function. Moments, Dispersion, Skewness, Kurtosis and Quantiles. Cauchy-Swartz Inequality, inequalities related to moments and measures of skewness and kurtosis.	
Unit 4	**10L**
Moment generating function, Cumulant generating function and Characteristic function. Uniqueness and inversion theorems (without proof) along with applications. Gambler's ruin problem.	

References:

1. Chung, K.L. (1983): Elementary Probability Theory with Stochastic Process, Springer / Narosa.
2. Feller, W. (1968): An Introduction to Probability Theory & its Applications, John Wiley.
3. Goon, A.M., Gupta, M.K. &Dasgupta, B. (1994): An Outline of Statistical Theory (Vol-1), World
4. 4. Press. Parzen, E. (1972): Modern Probability Theory and its Applications, John Wiley.
5. Uspensky, J.V. (1937): Introduction to Mathematical Probability, McGraw Hill.
6. Cacoullos, T. (1973): Exercises in Probability. Narosa.
7. Rahman, N.A. (1983): Practical Exercises in Probability and Statistics
8. Griffen. Ross, S. (2002): A First Course in Probability, Prentice Hall.

STAUGMCC1102 Demography and Vital Statistics

Course Objectives	1. Understand the basics of demography. 2. Understand the core social demographic variables, and how these variables influence population growth, composition, and structure. 3. Use of different mortality rate as demographic tools in understanding public health issues knowledge attitude and practices. 4. Know the concept and construction of abridged life tables. 5. Identify appropriate sources of data, perform basic demographic analyses using various techniques and ensure their comparability across populations.
Learning Outcomes	1. To find Standardized death rate by: - (i) Direct method (ii) Indirect method 2. To construct a complete life table 3. To fill in the missing entries in a life table 4. Calculate probabilities of death at pivotal ages and use it construct abridged life table using (i) Reed-Merrell Method, (ii) Greville's Method and (iii) King's Method 5. Calculate CBR, GFR, SFR, TFR for a given set of data 6. Calculate Crude rate of Natural Increase and Pearle's Vital Index for a given set of data 7. Calculate GRR and NRR for a given set of data and compare them

Theory	
Unit 1	**13**
Population Theories: Coverage and content errors in demographic data, use of balancing equations and Chandrasekharan-Deming formula to check completeness of registration data. Adjustment of age data, use of Myer and UN indices, Population composition, dependency ratio.	
Unit 2	**12**
Introduction and sources of collecting data on vital statistics, errors in census and registration data. Measurement of population, rate and ratio of vital events. Measurements of Mortality: Crude Death Rate (CDR), Specific Death Rate (SDR), Infant Mortality, Rate (IMR) and Standardized Death Rates.	
Unit 3	**10**
Stationary and Stable population, Central Mortality Rates and Force of Mortality. Life (Mortality) Tables: Assumption, description, construction of Life Tables and Uses of Life Tables.	
Unit 4	**10**

Abridged Life Tables; Concept and construction of abridged life tables by Reed-Merrell method, Greville's method and King's Method. Measurements of Fertility: Crude Birth Rate (CBR), General Fertility Rate (GFR), Specific Fertility Rate (SFR) and Total Fertility Rate (TFR). Measurement of Population Growth: Crude rates of natural increase, Pearl's Vital Index, Gross Reproduction Rate (GRR) and Net Reproduction Rate (NRR).	
PRACTICAL/LAB. WORK:	30

1. To calculate CDR and Age Specific death rate for a given set of data.
2. To find Standardized death rate by: - (i) Direct method (ii) Indirect method.
3. To construct a complete life table.
4. To fill in the missing entries in a life table.
5. To calculate probabilities of death at pivotal ages and use it construct abridged life table using (i) Reed-Merrell Method, (ii) Greville's Method and (iii) King's Method.
6. To calculate CBR, GFR, SFR, TFR for a given set of data.
7. To calculate Crude rate of Natural Increase and Pearle's Vital Index for a given set of data.
8. Calculate GRR and NRR for a given set of data and compare them.

References:

1. Mukhopadhyay, P. (1999): Applied Statistics, New Central Book Agency, Calcutta.
2. Goon, A.M., Gupta, M.K. and Dasgupta, B. (2008): Fundamentals of Statistics, Vol. II,9th Edition World Press, Kolkata.
3. Gupta, S. C. and Kapoor, V.K. (2008): Fundamentals of Applied Statistics, 4th Edition (Reprint), Sultan Chand & Sons.

STAUGSEC1101 Presentation Skill and Introduction to R Programming

Course Objectives:

In this course students will

1. Learn different parts of a scientific report and how to prepare using microsoft word/Google docs/Latex.
2. Learn how to prepare a good presentation using power point/Beamer.
3. Learn how to develop presentation skills.
4. Learn and practice about Writing and compiling codes in R, writing outputs in word.
5. Use loops and controls in R, functions in R.

Learning Outcomes:

At the end of the course a student will be able to

1. Prepare report using Microsoft office word/Google docs/Latex

2. Deliver presentations using power point/Beamer.
3. Exercise their report writing and presentation skills.
4. Write and compile codes using R.
5. Create different objects in R like scalar, vector, matrix, data frame, array and their applications.
6. Write codes based on controls, loops and create their own functions.

Unit 1	**25 L**
Formats of Reports: introduction, parts of a report, cover and title page, introductory pages, text, reference section, typing instructions, copy reading, proof reading. Presentation of a report: introduction, communication dimensions, presentation package, audio-visual aids, presenter's poise. Application to Microsoft office word, Google docs, Microsoft power point, Latex/Rmarkdown Project to create a report and give some presentation.	
Unit 2	**10 L**
Introduction to R; R help; help.search(), R mailing list, contributed documentation on CRAN. Saving workspace/history. Writing programs in R markdown. Data types in R: numeric/character/logical; real/integer/complex, strings and the paste command, R as a calculator: The four basic arithmetic operations. Use of parentheses nesting up to arbitrary level. The power operation. Evaluation of simple expressions. Quotient and remainder operations for integers. Standard functions, e.g., sin, cos, exp, log. Creating a vector using c(), seq() and colon operator. Operations on vectors such as addition, subtraction, multiplication, Functions to summarize a vector: sum, mean, sd, median etc.	
Unit 3	**10 L**
Creation of matrix and operations on matrices such as addition, subtraction, multiplication, determinant, trace, matrix inverse, Solution of linear equations, eigenvalues and eigenvectors Other R objects: data frames, lists, factor, array etc. and their creation and various operations Uses of if, if-else and nested if-else and related exercises. Loops in R and writing codes using loops, Writing own functions in R and related exercises.	

References:

1. Gardener, M (2012) Beginning R: The Statistical Programming Language, Wiley Publications.
2. Braun W J, Murdoch D J (2007): A First Course in Statistical Programming with R. Cambridge University Press. New York.
3. A simple introduction to R by Arnab Chakraborty (freely available at http://www.isical.ac.in/~arnabc/) R for beginners by Emmanuel Paradis (freely available at https://cran.r-project.org/doc/contrib/Paradisrdebuts_en.pdf).

Semester-II

STAUGMCC1203/ STAUGMIN1202 Descriptive Statistics

Course Objectives	In this course will learn: ● Data visualization. ● Define and compute the various measures of central tendency and different measures of dispersion, their use in comparing variability of different distribution. ● Explain the concept of skewness and kurtosis and the significance of its study. ● Understand the meaning of the term correlation and significance of its study. ● Understand the meaning of regression, and its role in statistical analysis. ● Differentiate between variables and attributes.
Learning Outcomes	Upon completion of this course, students will be able to: ● Have visual perception of different type of data, with help of Tables, Histogram, Charts, Graphs, etc. using Excel. ● Apply correctly a variety of descriptive statistical techniques. ● Develop the analytical understanding of cause and effect of regression models. ● Interpret, in plain language, the application and outcomes of statistical techniques.

Theory	
Unit 1	**15L**
Statistics: Definition and scope. Concepts of statistical population and sample. Data: quantitative and qualitative, cross-sectional and time-series, discrete and continuous. Scales of measurement: nominal, ordinal, interval, and ratio. Presentation of data: tabular and graphical. Frequency distributions, cumulative frequency distributions and their graphical representations. Stem & Leaf diagram. Box Plot. Outliers.	
Unit 2	**15L**
Measures of Central Tendency: Mean, Median, Mode. Measures of Dispersion: Range, Mean deviation, Standard deviation, Coefficient of variation, Lorenz Curve. Moments, skewness, and kurtosis. Quantiles and measures based on them.	
Unit 3	**18L**
Bivariate data: Definition, scatter diagram, expectation and covariance, simple correlation, linear regression, principle of least squares, fitting of polynomial and exponential curves, correlation ratio, correlation index, intraclass correlation. Rank correlation – Spearman's and Kendall's measures.	
Unit 4	**12L**
Analysis of Categorical Data: Contingency table, independence and association of attributes, measures of association - odds ratio, Pearson's and Yule's measure, Goodman-Kruskal gamma.	
PRACTICAL/LAB. WORK:	
1. Diagrammatic representation of data. 2. Problems based on construction of frequency distributions, cumulative frequency distributions and their graphical representations, stem, and leaf plot.	

3. Problems based on measures of central tendency.
4. Problems based on measures of dispersion.
5. Problems based on combined mean and variance and coefficient of variation.
6. Problems based on moments, skewness, and kurtosis.
7. Problems related to quantiles and measures based on them, construction of box plot.
8. Problems based on analysis of bivariate data.
9. Problems based on measures of rank correlation.
10. Problems based on analysis of categorical data.

References:

1. Goon, A.M., Gupta, M.K. and Dasgupta, B. (2002): Fundamentals of Statistics, Vol. I& II, 8th Edn. The World Press, Kolkata.
2. Miller, Irwin and Miller, Marylees (2006): John E. Freund's Mathematical Statistics with Applications, (7th Edn.), Pearson Education, Asia.
3. Mood, A.M., Graybill, F.A. and Boes, D.C. (2007): Introduction to the Theory of Statistics, 3rd Edn. (Reprint), Tata McGraw-Hill Pub. Co. Ltd.
4. Tukey, J.W. (1977): Exploratory Data Analysis, Addison-Wesley Publishing Co.
5. Agresti, A. (2010): Analysis of Ordinal Categorical Data, 2nd Edition, Wiley.
6. Freedman, D., Pisani, R. and Purves, R. (2014): Statistics, 4th Edition, W. W. Norton & Company

STAUGMCC1204 Probability-II

Course objectives	In this course students will learn ● Various univariate discrete and continuous distributions and their examples in real life and their statistical properties ● Two dimensional random variables: discrete as well as continuous type, joint, marginal, and conditional distributions ● Convergence of sequence of random variables: Almost sure convergence, Convergence in probability, Convergence in distribution, mean-square convergence ● Weak law and Strong law of large numbers ● Central limit theorem for iid random variables and applications.
Learning Outcomes	At the end of the course a student will be able to 8. Solve problems related to various discrete as well as continuous distributions 9. Fit a data set to a distribution such as binomial, Poisson etc. 10. Solve problems related to bivariate distributions 11. Solve problem related to probability inequalities 12. To check convergence of a sequence of random variables 13. Derive the asymptotic distributions

Theory	
Unit 1	**10L**

Standard discrete probability distributions: Binomial, Poisson, geometric, negative binomial, hypergeometric, uniform. Standard continuous probability distributions: uniform, normal, exponential, Cauchy, beta, gamma, lognormal, logistic, double exponential and Pareto along with their properties and limiting/approximation cases. Truncated distributions.	
Unit 2	**5L**
Standard Probability Inequalities (Univariate Cases): Markov's & Chebyshev's (one- and two- sided) inequalities, Jensen's Inequality, Holder's Inequality.	
Unit 3	**10L**
Two dimensional random variables: discrete type, joint, marginal, and conditional p.m.f and c.d.f., statement of properties of c.d.f, independence of variables, Sum-law and Product-law of expectation, trinomial distribution. Bivariate c.d.f and p.d.f. and generating functions in continuous case. Marginal and Conditional distributions, Independence, Conditional Expectation, Correlation and Regression. Theorems on sum and product of expectations of random variables. Bivariate Normal Distribution (BVN).	
Unit 4	**15L**
Limit laws: Sequence of random variables, convergence in probability, almost sure convergence, convergence in mean square and convergence in distribution and their interrelations, W.L.L.N., S.L.L.N and their applications, De-Moivre Laplace Limit theorem, Statement of Central Limit Theorem (C.L.T.) for i.i.d. variates, applications of C.L.T.	
PRACTICAL/LAB. WORK:	30L

1. Fitting of binomial distributions for n and p = q = ½.
2. Fitting of binomial distributions for given n and p.
3. Fitting of binomial distributions after computing mean and variance.
4. Fitting of Poisson distributions for given value of mean
5. Fitting of Poisson distributions after computing mean.
6. Fitting of negative binomial distribution.
7. Fitting of suitable discrete distributions
8. Application problems based on binomial distribution.
9. Application problems based on Poisson distribution.
10. Application problems based on negative binomial distribution.
11. Problems based on area property of normal distribution.
12. To find the ordinate for a given area for normal distribution.
13. Application based problems using normal distribution.
14. Fitting of normal distribution when parameters are given.
15. Fitting of normal distribution when parameters are not given.
16. Problems like those in 11 to 15 in cases of other continuous distributions.
17. Application based Problems on trinomial distributions.
18. Application based Problems on bivariate normal distributions.

References:

1. Chung, K.L. (1983): Elementary Probability Theory with Stochastic Process, Springer / Narosa.
2. Feller, W. (1968): An Introduction to Probability Theory & its Applications, John Wiley.
3. Goon, A.M., Gupta, M.K. &Dasgupta, B. (1994): An Outline of Statistical Theory (Vol-1), World Press.
4. Parzen, E. (1972): Modern Probability Theory and its Applications, John Wiley.

5. Uspensky, J.V. (1937): Introduction to Mathematical Probability, McGraw Hill.
6. Cacoullos, T. (1973): Exercises in Probability. Narosa.
7. Rahman, N.A. (1983): Practical Exercises in Probability and Statistics.

STAUGMDC1202 Economic Statistics and Official Statistics

Course objectives	In this course a student will learn ● Understand institutional, legal, and organizational aspects of official statistics in India. ● Understand the functioning of official statistics.
Learning Outcomes	At the end of the course a student will be able to 14. How index numbers are created and their interpretations 15. Understand the role of Statistics in official purpose

Economic Statistics: [L25]
Index Numbers: Price, Quantity and Value indices. Price Index Numbers: Construction, Uses, Limitations, Tests for index numbers, Various formulae and their comparisons, Chain Index Number. Some Important Indices: Consumer Price Index, Wholesale Price Index, and Index of Industrial Production methods of construction and uses.
Measurement of income inequality: Gini's coefficient, Lorenz curves, Application of Pareto and Lognormal as income distributions.

Official Statistics: [l20]
The Present Statistical System in India: The central and State Government organizations, the functions of the Central Statistical Organization (CSO), the National Sample Survey Organization (NSSO), Ministry of Statistics & Program Implementation (MoSPI), and National Statistical Commission.

Methods of collection of official statistics, their reliability, and limitations. Government of India's Principal publications containing data on the topics such as population, industry, and finance.

National Income Statistics: Income, expenditure, and production approaches. Their applications in various sectors in India.

STAUGSEC1202 Statistical commutations using R

Course Objectives:

In this course students will

1. Learn how to import and export datasets in R
2. Learn how to subset a dataset using rows/columns/groups or any combination of these.
3. How create various plots in R.
4. How to generate samples using R.
5. Learn various applications based on samples generated from a distribution.

Learning Outcomes:

At the end of the course a student will be able to

1. Import and export datasets.
2. Subset a dataset using rows/columns/groups or any combination of these.
3. Create various plots in R.
4. Generate samples using R.
5. Apply samples generated from a distribution to verify central limit theorems and to estimate the value of pi.

Unit 1	**20**
Importing and exporting datasets, sub-setting datasets, various datasets in R Graphs in R: the plot command, histogram, bar plot, box plot, scatter plot, lines, segments, arrows, inserting mathematical symbols in a plot, pie diagram, Customization of plot: setting graphical parameters, adding text, saving to a file; Adding a legend.	
Unit 2	**20**
Using the computer for random number generation. (treated as a black box) A brief look at some popular approaches (no mathematical justification needed). Simulating a coin toss, a die roll and a card shuffle. CDF inversion method. Simulation from standard distributions. Finding probabilities and moments using simulation. Monte Carlo integration. Basic idea of importance sampling.	
Unit 3	**20**
Generating from Binomial and Poisson distributions, and comparing the histograms to the PMFs. Generating from Uniform(0,1) distribution, and applying inverse CDF transforms. Simulating Gaussian distribution using Box-Muller method. Approximating the expectation of a given function of a random variable using simulation. Graphical demonstration of the Law of Large Numbers. Approximating the value of pi by simulating dart throwing.	

References:

1. Shonkwiler, Ronald W. and Mendivil, Franklin (2009): Explorations in Monte Carlo Methods (Undergraduate Texts in Mathematics)
2. Carsey, Thomas M. and Harden, Jeffrey J. (2014): Monte Carlo Simulation and Resampling Methods for Social Science.

Semester-III

STAUGMCC2305 Inference-I / STAUGMIN2404 Inference

Course Objective	1. Introduce the concept of estimation of parameters 2. Know the basic concepts of Testing of Hypotheses (Large Sample Tests and small sample test) 3. Know the confidence interval construction methods. 4. Introduce the concept of variance stabilizing transformation and large sample distributions.
Learning Outcomes	1. Calculate the problems related to point estimation. 2. Formulate hypothesis tests in some common models (including Normal models), correctly using the terms null hypothesis, alternative hypothesis, test statistic, rejection region and significance level. 3. Justify and make use of the Likelihood Ratio Test and the Generalized Likelihood Ratio tests. 4. Testing of significance and confidence intervals using variance stabilizing transformations. 5. Tests for goodness of fit, independence and homogeneity using Pearsonian chi-square statistic.

Theory	
Unit 1	**15L**
Estimation: Concepts of estimation, unbiasedness, sufficiency, consistency and efficiency. Factorization theorem. Complete statistic, Minimum variance unbiased estimator (MVUE), Rao-Blackwell and Lehmann-Scheffe theorems and their applications. Cramer-Rao inequality and MVB estimators (statement and applications).	
Unit 2	**10L**
Methods of Estimation: Method of moments, method of maximum likelihood estimation, method of minimum Chi-square, basic idea of Bayes estimators.	
Unit 3	**20L**
Principles of test of significance: Null and alternative hypotheses (simple and composite), Type-I and Type-II errors, critical region, level of significance, size and power, best critical region, Most powerful test, uniformly most powerful test, Neyman Pearson Lemma (statement and applications to construct most powerful test) and its applications to construct uniformly most powerful test, unbiased test (definition only). Likelihood ratio test, properties of likelihood ratio tests (without proof). Large sample tests, use of CLT for testing single proportion, difference of two proportions, single mean, difference of two means, standard deviation, and difference of standard deviations by classical and p-value approaches.	
PRACTICAL/LAB. WORK:	30L
1. Maximum Likelihood Estimation 2. Estimation by the method of moments, minimum Chi-square 3. Most powerful critical region (NP Lemma)	

4. Uniformly most powerful critical region
5. Unbiased critical region
6. Power curves
7. Likelihood ratio tests for simple null hypothesis against simple alternative hypothesis
8. Likelihood ratio tests for simple null hypothesis against composite alternative hypothesis
9. Asymptotic properties of LR tests
10. Testing of significance and confidence intervals for single proportion and difference of two proportions using CLT.
11. Testing of significance and confidence intervals for single Poisson mean and difference of two Poisson means using CLT.
12. Testing of significance and confidence intervals concerning sample standard deviation, coefficient of variation and correlation coefficient (both single sample and two sample cases).
13. Testing of significance and confidence intervals using variance stabilizing transformations.
14. Determination of the minimum sample size required to achieve normality by sample proportion, mean and standard deviation.
15. Tests for goodness of fit, independence and homogeneity using Pearsonian chi-square statistic.

References:

1. Goon, A.M., Gupta, M.K. and Dasgupta, B. (2003): An Outline of Statistical Theory, Vol. I, 4th Edn. World Press, Kolkata.
2. Rohatgi V. K. and Saleh, A.K. Md. E. (2009): An Introduction to Probability and Statistics. 2ndEdn. (Reprint) John Wiley and Sons.
3. Hogg, R.V. And Tanis, E.A. (2009): A Brief Course in Mathematical Statistics. Pearson Education.
4. Johnson, R.A. and Bhattacharya, G.K. (2001): Statistics-Principles and Methods, 4th Edn. John Wiley and Sons.
5. Mood, A.M., Graybill, F.A. and Boes, D.C. (2007): Introduction to the Theory of Statistics, 3rd Edn. (Reprint). Tata McGraw-Hill Pub. Co. Ltd.

STAUGMCC2306 Sampling Distributions

Course Objectives	Distribution of sample statistic (such as, sample mean, sample variance and sample proportion) and the pattern of variability of sample statistic will be explored in this course. Different properties of chi-square distribution, t-distribution and F-distribution will be discussed.
Learning Outcomes	Students will learn different properties of sample statistic and it can be used in testing of hypothesis. They can perform the hypothesis testing for population mean, variance and the testing for goodness of fit.

Theory	
Unit 1	**15L**
Definitions of random sample, parameter and statistic, sampling distribution of a statistic, sampling distribution of sample mean, standard errors of sample mean, sample variance and sample proportion.	
Unit 2	**8L**
Order Statistics: Introduction, distribution of the r^{th} order statistic, smallest and largest order statistics. Joint distribution of r^{th} and s^{th} order statistics, distribution of sample median and sample range.	
Unit 3	**12L**

Exact sampling distribution: Definition and derivation of p.d.f. of χ^2 with n degrees of freedom (d.f.) using m.g.f., nature of p.d.f. curve for different degrees of freedom, mean, variance, m.g.f., cumulant generating function, mode, additive property and limiting form of χ^2 distribution. Tests of significance and confidence intervals based on distribution.	
Unit 4	**10L**
Exact sampling distributions: Student's and Fishers t-distribution, Derivation of its p.d.f., nature of probability curve with different degrees of freedom, mean, variance, moments and limiting form of t distribution. Snedecore's F-distribution: Derivation of p.d.f., nature of p.d.f. curve with different degrees of freedom, mean, variance and mode. Distribution of $1/F(n_1,n_2)$. Relationship between t, F and $\chi 2$ distributions. Test of significance and confidence Intervals based on t and F distributions.	
PRACTICAL/LAB. WORK:	30L
1. Testing of significance and confidence intervals for single proportion and difference 1. of two proportions 2. Testing of significance and confidence intervals for single mean and difference of two means and paired tests. 3. Testing of significance and confidence intervals for difference of two standard deviations. 4. Exact Sample Tests based on Chi-Square Distribution. 5. Testing if the population variance has a specific value and its confidence intervals. 6. Testing of goodness of fit. 7. Testing of independence of attributes. 8. Testing based on 2 X 2 contingency table without and with Yates' corrections. 9. Testing of significance and confidence intervals of an observed sample correlation coefficient. 10. Testing and confidence intervals of equality of two population variances	

References:

1. Goon, A.M., Gupta, M.K. and Dasgupta, B. (2003): An Outline of Statistical Theory, Vol. I, 4th Edn. World Press, Kolkata.
2. Rohatgi V. K. and Saleh, A.K. Md. E. (2009): An Introduction to Probability and Statistics. 2ndEdn. (Reprint) John Wiley and Sons.
3. Hogg, R.V. And Tanis, E.A. (2009): A Brief Course in Mathematical Statistics. Pearson Education.
4. Johnson, R.A. and Bhattacharya, G.K. (2001): Statistics-Principles and Methods, 4th Edn. John Wiley and Sons.
5. Mood, A.M., Graybill, F.A. and Boes, D.C. (2007): Introduction to the Theory of Statistics, 3rd Edn. (Reprint). Tata McGraw-Hill Pub. Co. Ltd.

STAUGMDC2303 Statistics in Ecology and Epidemiology

Course Objectives	In this course students will learn: 1. Various aspects of Ecology. 2. Study various growth models. 3. Estimation of population density. 4. Various aspects of Epidemiology.

Learning Outcomes	After completion of the course a student will 1. Apply growth models in real life. 2. Apply various techniques to estimate population density. 3. Apply various techniques of epidemiology.

Unit-1: [L15]
Introduction to ecology and evolution, population dynamics: single species Exponential, Logistic and Gompertz models, Leslie matrix model for age and stage Structured population, survivorship curves- Constant, monotone and bath tub shaped hazard rates.

Unit-2: [L10]
Population density estimation: Capture-recapture, Nearest Neighbor, line transect sampling, indirect methods. Ecological Diversity: Species abundance curve, indices of diversity (Simpson's index, Shannon Wiener index).

Unit-3: [L10]
Basic concepts and ethics of epidemiology, Measures of exposure and outcome: History of Epidemiology, Emergence of modern epidemiology, Measures of Exposures, Types of exposures, Sources of exposures, Measures of outcome. Measures of exposure effect, relative and absolute measures of effect. Communicable and non-communicable diseases.

Unit-4: [L10]
Disease registries, international classification of diseases. Measures of disease frequency: Prevalence, Incidence, Risk, Odds of disease, Incidence time, Relationship between prevalence, rate and risk, measures of disease occurrence, direct and indirect method of standardization, cumulative rate, cumulative risk, proportional incidence.

Reference Books:

1. Gore, A.P. and Paranjpe S.A. (2001): A course on Mathematical and Statistical Ecology- Kluwer Academic Publishers.
2. Pielou, E.C. (1977) An Introduction to Mathematical Ecology-Wiley
3. Hilborn and Mangel (1997): The Ecological Detective: Confronting Models with Data- Princeton University Press.
4. Henry M. and Stevens H. (2009): A primer of ecology with R- Springer.
5. Deepti Shyam Sunder (2019) : Fundamentals of Epidemiology and Biostatistics, CBS Publishers & Distributors.
6. Alan J. Silman : Epidemiological Studies, Cambridge University Press
7. K. Park (2013): Parks's Textbook of Preventive and Social Medicine, Banarasi das Bhanot Publishers, Jabalpur.

STAUGSEC2303 Statistical Computations using Python

Course Objectives:

In this course students will

1. Learn and practice about Writing and compiling codes in Python, and writing outputs in word.
2. Handle datasets in Python.
3. Use loops, controls, and functions in Python.
4. Learn and practice statistical computations in Python.
5. Applications of python in the context of machine learning.
6. Big data using Python.

Learning Outcomes:

At the end of the course a student will be able to

1. Write and execute programs in Python.
2. Perform various statistical computations using Python.
3. Testing of hypothesis using Python.
4. Simulation of standard distributions using Python.
5. Use some Machine learning techniques in Python.
6. Use Python for Big data analysis.

Detailed Syllabus:

Unit-I **[L-10]**

Introduction to Python: Basic Idea; Simple Syntax; Basic Operations; Different Libraries and exploring various functions such as math; Function; Loop; Array.

Exploring arrays with NumPy: Creation of Various data structures such as Lists, Dicts, Series, Array, matrix, sequence and various operations on these.

Unit-II **[L-15]**

Empowering data analysis with pandas, Reading and writing data from various formats, Data handling and management, Statistical computations using library such as statistics

Chart and Diagrams using library such as Matplotlib; Random Number Generation from a known and unknown distribution; Simulation; Application in various Statistical field; Idea of Parallel Computing and/or Efficient Programming.

Unit-III **[L-20]**

Execution of data mining and machine learning methods in Python. Leveraging Python in the World of Big Data

References:

1. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, Sebastian Raschka, Vahid Mirjalili; Packt Publishing (2019)
2. Mastering Python for Data Science, Samir Madhavan; Packt Publishing (2015)
3. Introduction to Machine Learning with Python_ A Guide for Data Scientists, Andreas C. Mueller, Sarah Guido; O'Reilly Media (2016)

Semester-IV

STAUGMCC2407/ STAUGMIN2303 Sample Survey and Time Series

Course Objective	In this course a student will learn 1. Population, Sample and their definitions and examples 2. Need for Sampling 3. Planning and execution of sample surveys. 4. Judgment and probability sampling schemes. 5. Simple Random Sampling with and without replacement. 6. The details of time series and their components. 7. The concept of estimation of different components of time series data. 8. The basics of Box-Jenkins approach.
Learning Outcomes	At the end of the course a student will be able to 1. Draw random numbers. 2. Organize a sample survey. 3. Determination of sample size. 4. Prepare questionnaire. 5. Determine the proper sampling method. 6. Estimate the various components of time series. 7. Apply the concept of stationarity to the analysis of time series data in various contexts.

Unit 1	**10L**
Concept of population and sample, complete enumeration versus sampling, sampling, and non-sampling errors. Types of sampling: non-probability and probability sampling, basic principle of sample survey, simple random sampling with and without replacement, definition, and procedure of selecting a sample, estimates of: population mean, total and proportion, variances of these estimates, estimates of their variances and sample size determination.	
Unit 2	**10L**
Stratified random sampling: Technique, estimates of population mean and total, variances of these estimates, proportional and optimum allocations, and their comparison with SRS. Practical difficulties in allocation, estimation of gain in precision, post stratification and its performance. Systematic Sampling:	

Technique, estimates of population mean and total, variances of these estimates (N=nxk). Comparison of systematic sampling with SRS and stratified sampling in the presence of linear trend and corrections.	
Unit 3	**10L**
Stochastic Process: Introduction and Stationary Process. Introduction to time series data, application of time series from various fields. Modelling time series as deterministic function plus IID errors: Components of a time series (trend, cyclical and seasonal patterns, random error) Decomposition of time series. Estimation of trend: free hand curve method, method of moving averages, fitting various mathematical curves and growth curves. Effect of elimination of trend on other components of the time series.	
Unit 4	15L
Estimation of seasonal component by Method of simple averages, Notions of multiplicative models: ratio to Trend. Introduction to stochastic modelling: Concept of stationarity. Illustration of how a stationary time series may show temporal patterns. Stationarity in mean.	

PRACTICAL/LAB WORK

Submit a Research Report based on empirical study on some real-life situation. The student will individually/group collect, analyze, interpret the data, and prepare a report.

References:

1. Lilien, Gary L. and Philip Kotler, 1983. Marketing Decision Making; A Model Building Approach, Harper & Row, New York.
2. Shenoy, GVS, et al., (1983). Quantitative Techniques for Managerial Decision Making, Wiley Eastern
3. C. Chatfield : The Analysis of Time Series – An Introduction
4. G.E.P. Box ,G.M. Jenkins & G.C.Reinsel : Time Series Analysis – Forecasting & Control
5. P. J. Brockwell & R.A. Davis : Introduction to Time Series Analysis and Forecasting
6. A.Pankratz : Forecasting with Univariate Box-Jenkins Model
7. G. Janacek and L. Swift : Time Series –Forecasting, Simulation, Applications

STAUGMCC2408 Linear Models:

Course Objectives: The main objective of this paper is to

1. Understand multidimensional space which is essential for various courses offered in this programme.
2. Learn about vector spaces, linear span, basis, dimension, linear dependence etc.
3. Learn about eigenvalues and eigenvectors, and diagonalization of a matrix.
4. Learn about quadratic forms, differentiation of vectors and matrices and optimum values of quadratic forms.
5. Learn about various generalized inverses and their properties.
6. Learn about estimability of linear parametric functions.
7. Learn how to test Linear hypothesis.

Learning Outcomes: After completing this course students will be able to

1. Check the linear dependence and independence of a set of vectors
2. Find basis and dimension, orthonormal basis etc.
3. Find Eigenvalues and Eigenvectors.
4. Find generalized inverses.
5. Find optimum values of linear form, quadratic forms, bilinear form and their examples in Statistics.
6. Check estimability, find BLUE.
7. Test linear hypothesis.

Detailed Syllabus:

Vector spaces and Subspaces with examples, Direct sum and Algebra of subspaces viz. sum, intersection, union etc, Linear combinations, Spanning sets, Linear spans, Linear dependence and independence in vector spaces, Row and Column space of a matrix, Basis and Dimensions. Orthogonality, Orthonormal sets and Bases, Gram Schmidt Orthogonalization Process. (12)

Eigenvalues and eigenvectors, Spectral decomposition of a symmetric matrix (Full rank and non-full rank cases), Example of spectral decomposition, Spectral decomposition of asymmetric matrix, Cayley Hamilton theorem, Algebraic and geometric multiplicity of characteristic roots, Diagonalization of matrices, Factorization of a matrix (12)

Generalized inverse of a matrix, Different classes of generalized inverse, Properties of g-inverse, Reflexive g-inverse, Minimum norm g-inverse, Least squares g-inverse, Moore-Penrose (MP) g-inverse and its properties, Real quadratic form, Linear transformation of quadratic forms, Index and signature, Singular value decomposition. Optimum values of a quadratic form, Vector and matrix differentiation.

(14)

Gauss-Markov model: Estimation space and error space, estimable function, BLUE and related results, Least Square estimation, Gauss- Markov Theorem. Sum of squares due to a test of linear functions. Description of F test for a general linear hypothesis (proof is not required). Linear models for correlated errors. (8)

ANOVA: fixed, random and mixed effects model, ANCOVA, Multiple comparison, S-method and T-method of multiple comparison (4)

References:

1. R. B. Bapat . Linear Algebra and Linear Models (3rd Ed.), Springer
2. D. A. Harville (2008). Matrix Algebra From a Statistician's Perspective, 2nd Ed. Springer
3. Biswas, S. (1997). A Text Book of Matrix Algebra, 2 nd ed., New Age International Publishers.
4. Golub, G.H. and Van Loan, C.F. (1989). Matrix Computations, 2nd ed., John Hopkins University Press, Baltimore-London.
5. Hadley, G. (2002). Linear Algebra. Narosa Publishing House (Reprint).
6. Robinson, D.J.S. (1991). A Course in Linear Algebra with Applications, World Scientific, Singapore.
7. Rao, C.R. (1973). Linear Statistical Inferences and its Applications, 2nd ed., John Wiley & Sons.
8. Searle, S.R. (1982). Matrix Algebra useful for Statistics, John Wiley & Sons.
9. Strang, G. (1980). Linear Algebra and its Application, 2nd ed., Academic Press, London New York.

STAUGMCC2409 Regression Analysis

Course Objectives: The objective of this course is to provide the student with various techniques involved in regression analysis including developing linear regression model, variable selection, handling multicollinearity, dealing with categorical response variables. The departures from the Gauss-Markov set-up will be discussed in this course.

Learning Outcomes:

After successful completion of this course, student will be able to:

1. Fit and interpret linear models for real life problems.
2. Perform the test for significance of model parameters.
3. Select suitable variables for the linear model.
4. Check and handle the multicollinearity.
5. Fit model when response variable is categorical.

Detailed Syllabus:

General theory of regression, simple and multiple regression, fitting of polynomial regression by orthogonal methods, examination of regression equation. (10)

Detection of outliers and influential observations: residuals and leverages, DFBETA, DFFIT, Cook's Distance and COVRATIO. (4)

Building a regression model: Transformations – Box-Cox and Box-Tidwell models, Stepwise regression, Model selection (adjusted R^2, cross validation and C_p criteria, AIC, PRESS). Model selection problems. Concept of best subset regression (7)

Multicollinearity – detection and remedial measures. Ridge regression and Lasso. (10)

Departures from the Gauss-Markov set-up: Heteroscedasticity and Autocorrelation – detection and remedies. (8)

Logistic regression, Dummy variables, piecewise regression, splines and scatter plot smoothing. (8)

Checking for normality: Q-Q plots, Normal Probability plot, Shapiro-Wilks test. (3)

References:

1. N.R. Draper & H. Smith : Applied Regression Analysis
2. D.W. Belsley, E. Kuh & R.E. Welsch: Regression Diagnostics – identifying Influential data & sources of collinearity
3. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani: An Introduction to Statistical Learning with applications in R, Springer, 2013.
4. J. Rousseeuw & A.M. Leroy : Robust Regression & Outlier Detection
5. R.D. Cook & S. Weisberg : Residual and its Influence in Regression
6. J. Johnston : Econometric Methods (3rd ed.)
7. G.G. Judge, W.E. Griffith, R.C. Hill,
8. W. Lutkepohl & T.C. Lee : The Theory and Practice of Econometrics (2nd ed.)
9. T.P. Ryan : Modern Regression Methods (2nd ed.)
10. J.O. Rawlings, S.G. Pantula & D.A. Dickey : Applied Regression Analysis: A Research Tool
11. S. Chatterjee & A.S. Hadi : Regression Analysis by Example

Semester-V

STAUGMCC3510 Multivariate Analysis-I

Course Objective	1. Know the basics of multivariate data and multiple regressions. 2. Introduction of multivariate normal and multinomial distributions and their properties. 3. Idea of Principal Components Analysis and Factor Analysis.
Learning Outcomes	1. Explain the key concepts multiple regression and related terms associated with it. 2. Derive some properties of multivariate normal and multinomial distributions. 3. Test for Multiple and partial Correlation. 4. Application of Principal Components Analysis and Factor Analysis.

Theory	
Unit 1	**20L**
Multivariate Data: multiple regression, multiple and partial correlation coefficients. Random Vector: Probability mass/density functions, Distribution function, mean vector & Dispersion matrix and its properties, Marginal & Conditional distributions. Multiple and partial correlation coefficient.	
Unit 2	**15L**
Multivariate Normal distribution and its properties. Multinomial Distribution and its properties. Tests for Multiple and partial correlation coefficients.	
Unit 3	**10L**
Applications of Multivariate Analysis: Principal Components Analysis and Factor Analysis (Application Oriented discussion, derivations not required)	
PRACTICAL/LAB. WORK:	30
1. Test for Multiple Correlation. 2. Test for Partial Correlation. 3. Multivariate Normal Distribution. 4. Principal Components Analysis. 5. Factor Analysis.	

References:

1. Anderson, T.W. (2003): An Introduction to Multivariate Statistical Analysis, 3rdEdn., John Wiley
2. Muirhead, R.J. (1982): Aspects of Multivariate Statistical Theory, John Wiley.
3. Kshirsagar, A.M. (1972): Multivariate Analysis, 1stEdn. Marcel Dekker.
4. Johnson, R.A. And Wichern, D.W. (2007): Applied Multivariate Analysis, 6thEdn., Pearson & Prentice Hall.
5. Mukhopadhyay, P.: Mathematical Statistics.
6. Goon, A.M., Gupta, M.K. and Dasgupta, B. (2002): Fundamentals of Statistics, Vol. I, 8th Edn. The World Press, Kolkata.
7. Rohatgi, V. K. and Saleh, A.K. Md. E. (2009): An Introduction to Probability and Statistics. 2ndEdn. (Reprint) John Wiley and Sons.

STAUGMCC3511 Inference-II

Course Objective	1. Learn parametric, sequential estimation (point, as well as, interval) and testing (simple, as well as, composite hypotheses) procedures. 2. Understand UMPU tests, SPRT, OC and ASN.
Learning Outcomes	1. Perform sequential estimation techniques and testing procedures to deal with real life problems. 2. Understand UMPU tests, SPRT, OC and ASN.

Theory	
Unit 1	**25L**
Review and problems: Consistency and efficiency. Factorization theorem. Complete statistic, Minimum variance unbiased estimator (MVUE), Rao-Blackwell and Lehmann-Scheffe theorems and their applications. Cramer-Rao inequality and MVB estimators. Confidence intervals, Confidence set, shortest length confidence interval, Concepts of Uniformly Most Accurate (UMA) confidence sets, relationship with tests of hypotheses. Delta Method, Derivation and uses of large sample standard error of sample moments, Standard deviation, Coefficient of Variation, b_1 & b_2 measures, Correlation coefficient. Asymptotic distribution of sample quantiles. Transformations of Statistics to stabilize variance: derivation and uses of Sin^{-1}, square root. Uses of logarithmic and z-transformations. Large sample tests for binomial proportions, Poisson means (single and two independent samples cases) and correlation coefficients. Large Sample distribution of Pearsonian χ^2 –statistic and its uses.	
Unit 2	**20L**
Sequential procedures, Wald's SPRT and its properties, fundamental identity, OC and ASN functions, optimality of SPRT An introduction of inference procedure based on random effects	
Unit 3: Invariant tests, Invariance and Sufficiency, Canonical reduction of linear models	
PRACTICAL/LAB. WORK:	30L
1.	

References:

1. Anderson, T.W. (2003): An Introduction to Multivariate Statistical Analysis, 3rdEdn., John Wiley
2. Muirhead, R.J. (1982): Aspects of Multivariate Statistical Theory, John Wiley.
3. Kshirsagar, A.M. (1972): Multivariate Analysis, 1stEdn. Marcel Dekker.

4. Johnson, R.A. And Wichern, D.W. (2007): Applied Multivariate Analysis, 6th Ed., Pearson & Prentice Hall.
5. Mukhopadhyay, P.: Mathematical Statistics.
6. Goon, A.M., Gupta, M.K. and Dasgupta, B. (2002): Fundamentals of Statistics, Vol. I, 8th Ed. The World Press, Kolkata.
7. Gibbons, J. D. and Chakraborty, S (2003): Nonparametric Statistical Inference. 4th Edition. Marcel Dekker, CRC.
8. Rohatgi, V. K. and Saleh, A.K. Md. E. (2009): An Introduction to Probability and Statistics. 2ndEdn. (Reprint) John Wiley and Sons.

STAUGMCC3512 Mathematical Analysis

Theory	
Unit 1	**15L**
Representation of real numbers as points on a line, Algebraic, Field Structure, Order Structure and Completeness properties of **R** (Concepts only), Archemedian Property, Bounded and unbounded sets, neighborhood of a point, Supremum and infimum, Topological properties of real line. Functions, Countable, Uncountable sets and Uncountability of **R**.	
Unit 2	**15L**
Sequences and their convergence, Subsequences, monotonic sequences, bounded sequences, squeeze theorem Limits of some special sequences such as r^n, $\left(1+\frac{1}{n}\right)^n$ & $n^{\frac{1}{n}}$, Concept of limsup and liminf.	
Unit 3	**15L**
Infinite series, positive termed series and their convergence, Comparison test, ratio test and root test. Absolute convergence of series, Leibnitz's test for the convergence of alternating series, Conditional convergence, Rearrangement and Riemann's Theorem (Statement only).	
Unit 4	15L
Continuous functions, uniform continuity, uniform convergence of sequences and series of functions, term by term differentiation and integration, applications to power series.	

References:

1. R.G. Bartle and D. R. Sherbert, Introduction to Real Analysis, 3rd Ed., John Wiley and Sons
2. (Asia) Pvt. Ltd., Singapore, 2002.
3. Gerald G. Bilodeau , Paul R. Thie, G.E. Keough, An Introduction to Analysis, 2nd Ed., Jones & Bartlett, 2010.
4. Brian S. Thomson, Andrew. M. Bruckner and Judith B. Bruckner, Elementary Real Analysis, Prentice Hall, 2001.
5. S.K. Berberian, A First Course in Real Analysis, Springer Verlag, New York, 1994.
6. S. Shirali, H. L. Vasudeva, An Introduction to Mathematical Analysis, Alpha Science International Ltd, 2013.

Semester-VI

STAUGMCC3613 Survival Analysis

Course Objective	1. The objectives of this course are to study the different models from Survival Analysis, to provide the construction of parametric and non-parametric estimators of survival distributions, and probability density functions based on incomplete data. 2. The models with right-censored, truncated and interval censored data will be considered. 3. The properties and constructions of several goodness-of-fit tests based on the modified empirical processes are studied. 4. Introduction to Competing Risk Theory.
Learning Outcomes	1. Describe many practical models by means of the counting processes based on the censored and truncated observation. 2. Ability to recognize the difference between parametric and non-parametric survival models. 3. Able to estimate survival function, cumulative hazard rate function using the so-called Kaplan-Meier estimator and Nelson-Aalen estimators, respectively. 4. Understanding of the Cox proportional hazard model and its connection to the log linear model. 5. Investigate the constructions of confidence intervals and the asymptotic properties of goodness-of-fit tests for one-sample and two-sample problems under random censorship. 6. Test hypotheses when modeling the survival time distributions using R-programming codes.

Theory	
Unit 1	**15L**
Survival Analysis: Functions of survival times, survival distributions and their applications- exponential, gamma, Weibull, Rayleigh, lognormal distributions, and distribution having bath-tub shaped hazard function. Mean Residual Time.	
Unit 2	**15L**
Censoring Schemes: Type I, Type II and progressive or random censoring with biological examples. Estimation of mean survival time and variance of the estimator for Type I and Type II censored data with numerical examples.	
Unit 3	**15L**
Non-parametric methods: Actuarial and Kaplan-Meier methods for estimating survival function and variance of the Estimator.	
Unit 4	**15L**
Competing Risk Theory: Indices for measurement of probability of death under competing risks and their interrelations. Estimation of probabilities of death using maximum likelihood principle and modified minimum Chi-square methods.	
PRACTICAL/LAB. WORK:	
1. To estimate survival function 2. To determine death density function and hazard function 3. To identify type of censoring and to estimate survival time for type I censored data 4. To identify type of censoring and to estimate survival time for type II censored data	

5. To identify type of censoring and to estimate survival time for progressively type I censored data
6. Estimation of mean survival time and variance of the estimator for type I censored data
7. Estimation of mean survival time and variance of the estimator for type II censored data
8. Estimation of mean survival time and variance of the estimator for progressively type I censored data
9. To estimate the survival function and variance of the estimator using Non-parametric methods with Actuarial methods
10. To estimate the survival function and variance of the estimator using Non-parametric methods with Kaplan-Meier method
11. To estimate Crude probability of death
12. To estimate Net-type I probability of death
13. To estimate Net-type II probability of death
14. To estimate partially crude probability of death
15. To estimate gene frequencies

References:

1. Lee, E.T. and Wang, J.W. (2003): Statistical Methods for Survival data Analysis, 3rd Edition, John Wiley and Sons.
2. Kleinbaum, D.G. (1996): Survival Analysis, Springer. Chiang, C.L. (1968): Introduction to Stochastic Processes in Bio Statistics, John Wiley and Sons.
3. Indrayan, A. (2008): Medical Biostatistics, 2nd Edition Chapman and Hall/CRC.

STAUGMCC3614 Inference-III

Course Objectives:

1. Introduction of some non-parametric tests.
2. To learn various decision rules theories and its applications of decision making as individuals, in groups, and in organizations.
3. Also understanding of the fundamentals of Bayesian inference including concept of subjectivity and priors by examining some simple Bayesian models.

Learning Outcomes: After successful completion of this course, student will be able to:

- Application of some nonparametric tests.
- Understand decision problem, loss function, risk function and decision rules.
- Treat “evidence” as the value of observations and prescribe methods to deal rationally with it.

- Equip students with skills to carry out and interpret posterior and preposterior data based modeling and analyses.
- Compute probability that the theory in question could produce the observed data.

Detailed Syllabus:

Nonparametric Tests: Introduction and Concept, Test for randomness based on total number of runs, Empirical distribution function, One Sample Tests: Kolmogrov- Smirnov, Sign, Signed rank. Wilcoxon-Mann-Whitney test. Kruskal-Wallis test. Efficiency Calculation

Bayesian Analysis: Overview and comparison of the three paradigms, classical statistics, data analysis and Bayesian analysis. Relative advantages and disadvantages. Choice of subjective priors conjugate priors. 2-persons game, Loss functions - squared error, absolute error and 0 - 1; reach function. Bayesian estimation of parameters.

Practical/Lab work:

1. Test for randomness based on total number of runs.
2. Kolmogorov -Smirnov test for one sample.
3. Sign test.
4. Signed rank test.
5. Wilcoxon-Mann-Whitney test.
6. Kruskal-Wallis test.

References:

1. J. Aitchison and I.R. Dunsmore: Statistical Prediction Analysis, Cambridge University Press.
2. G. E. P. Box and G. C. Tiao: Bayesian Inference in Statistical Analysis, Addison & Wesley.
3. M. H. DeGroot: Optimal Statistical Decisions, McGraw Hill.
4. T. Leonard and J. S. J. Hsu : Bayesian Methods, Cambridge University Press.
5. P. M. Lee: Bayesian Statistics: An Introduction, Arnold Press.
6. C. P. Robert: The Bayesian Choice: A Decision Theoretic Motivation, 2nd ed., Springer Verlag.

STAUGMCC3615 Design of Experiments

Course Objective	1. Define Design of Experiments (DOE) and describe its purpose, importance, and benefits. 2. Define key terms associated with DOE and explain how to conduct a well-designed statistical experiment. 3. Describe the basic designs and their applications.

	4. Define a full factorial experiment and show how to calculate the main and interaction effects. 5. Demonstrate total and partial confounding in a factorial design. 6. Explain the role of split plot design.
Learning Outcomes	1. Explain the key concepts of DOE, and why it is used. 2. Calculate treatment effects of a basic and factorial design. 3. Analyze complete and partial confounding in a factorial design. 4. Analysis of Split Plot designs. 5. Analysis of Groups of experiments in RBD and LSD.

Theory	
Unit 1	**6L**
Experimental designs: Role, historical perspective, terminology: Treatments, Experimental units & Blocks, Experimental error, Basic principles of Design of Experiments (Fisher). Uniformity trials, fertility contour maps, choice of size and shape of plots and blocks in Agricultural experiments. Uses in Industrial Experiments.	
Unit 2	**25L**
Basic designs: Completely Randomized Design (CRD), Randomized Block Design (RBD), Latin Square Design (LSD) – layout, model and statistical analysis, relative efficiency. Analysis with one missing observation in RBD and LSD.	
Unit 3	**20L**
Factorial experiments: advantages, notations and concepts. 2^n experiments: design and analysis. Total and Partial confounding for 2^n (n≤5). Factorial experiments in a single replicate.	
Unit 4	**9L**
Split Plot Design RBD and Strip arrangements, Groups of experiments with RBD and LSD.	
PRACTICAL/LAB. WORK:	
1. Analysis of a CRD. 2. Analysis of an RBD. 3. Analysis of an LSD. 4. Analysis of an RBD with one missing observation. 5. Analysis of an LSD with one missing observation. 6. Analysis of **2^2** and**2^3** factorial in CRD and RBD. 7. Analysis of a completely confounded two- level factorial design in 2 blocks. 8. Analysis of a completely confounded two- level factorial design in 4 blocks. 9. Analysis of a partially confounded two- level factorial design. 10. Analysis of a single replicate of a 2^n design. 11. Analysis of Split Plot and Strip Plot designs. 12. Analysis of Groups of experiments in RBD and LSD	

References:

1. Cochran, W.G. and Cox, G.M. (1959): Experimental Design. Asia Publishing House.
2. Das, M.N. and Giri, N.C. (1986): Design and Analysis of Experiments. Wiley Eastern Ltd.
3. Goon, A.M., Gupta, M.K. and Dasgupta, B. (2005): Fundamentals of Statistics. Vol. II, 8thEdn. World Press, Kolkata.
4. Kempthorne, O. (1965): The Design and Analysis of Experiments. John Wiley.

5. Montgomery, D. C. (2008): Design and Analysis of Experiments, John Wiley.
6. Wu, C. F. J. And Hamada, M. (2009). Experiments, Analysis, and Parameter Design Optimization (Second edition), John Wiley.
7. Dean, A.M. and Voss, D. (1999): Design and Analysis of Experiments. Springer Texts in Statistics

STAUGMDS3601 Statistical Quality Control / Econometrics / Data Structure and Database Management

Statistical Quality Control

Course Objective	1. Know the definition and dimension of quality. 2. Introduction of product control and process control, Statistical Process Control. 3. Know the different control charts and their comparison. 4. Introduction of product control and different sampling plan. 5. Basics of six-sigma methods.
Learning Outcomes	1. Explain the seven tools of SPC. 2. Construction and interpretation of statistical control charts X-bar & R-chart, X-bar & s-chart, np-chart, p-chart, c-chart, u-chart. 3. To estimate of process capability. 4. Apply different sampling plan related to product control and their characteristics. 5. Know the complete overview of six-sigma method.

Theory	
Unit 1	**10L**
Quality: Definition, dimensions of quality, Difference between product control and process control, Statistical Process Control - Seven tools of SPC, chance and assignable Causes of quality variation.	
Unit 2	**15L**
Statistical Control Charts - Construction and Statistical basis of 3-σ Control charts, Rational Sub-grouping, Control charts for variables: X-bar & R-chart, X-bar & s-chart. Control charts for attributes: np chart, p-chart, c-chart and u-chart. Comparison between control charts for variables and control charts for attributes. Analysis of patterns on control chart, Estimation of process capability.	
Unit 3	**13L**
Definitions related to product control, Acceptance sampling plan, Principle of acceptance sampling plans, Single sampling plan - their OC, AQL, LTPD, AOQ, AOQL, ASN, ATI functions with graphical interpretation, Double sampling plan - their OC, AQL, LTPD, AOQ, AOQL, ASN, ATI functions with graphical interpretation, use and interpretation of Dodge and Roming sampling inspection plan tables.	
Unit 4	**7L**
Introduction to Six-Sigma: Overview of Six Sigma, Lean Manufacturing and Total Quality Management (TQM), Introduction to ISO quality standards: ISO 9001, ISO 14001, BIS.	
PRACTICAL/LAB. WORK:	30L

1. Construction and interpretation of statistical control charts
    a. X-bar & R-chart
    b. X-bar & s-chart
    c. np-chart
    d. p-chart
    e. c-chart
    f. u-chart
2. Single sample inspection plan: Construction and interpretation of OC, AQL, LTPD, ASN, ATI, AOQ, AOQL curves

3. Calculation of process capability

References:

1. Montogomery, D. C. (2009): Introduction to Statistical Quality Control, 6th Edition, Wiley India Pvt. Ltd.
2. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. II, 8th Edn. The World Press, Kolkata.
3. Mukhopadhyay, P (2011): Applied Statistics, 2nd edition revised reprint, Books and Allied (P) Ltd.
4. Montogomery, D. C. and Runger, G.C. (2008): Applied Statistics and Probability for Engineers, 3rd Edition reprint, Wiley India Pvt. Ltd.
5. Ehrlich,B.Harris(2002):Transactional Six Sigma and Lean Servicing, 2nd Edition St. Lucie Press.
6. Hoyle, David (1995): ISO Quality Systems Handbook, Heinemann Publication. 2nd Edition, Butterworth.

Econometrics

Course Objective	1. Introduction to econometric models. 2. Basics of estimation under linear restrictions. 3. Concepts of multicolinearity. 4. Defining autocorrelation and generalized least square estimation. 5. Details of Heteroscedastic disturbances. 6. Introduction to Instrumental variable method.
Learning Outcomes	1. Problems related to consequences of Multicollinearity. 2. Diagnostics of Multicollinearity. 3. Solving the problems related to consequences of Autocorrelation (AR(I)). 4. Diagnostics of Autocorrelation. 5. Solving the problems related to consequences Heteroscedasticity and its diagnostics. 6. Estimation of problems of General linear model under Heteroscedastic distance terms. 7. Solving the problems on Autoregressive models and Instrumental variable.

Theory	
Unit 1	**15L**

Introduction: Objective behind building econometric models, nature of econometrics, model building, role of econometrics, structural and reduced forms. Estimation under linear restrictions. Dummy variables, Qualitative data.	
Unit 2	**15L**
Multicollinearity: Introduction and concepts, detection of multicollinearity, consequences and solutions of multicollinearity.	
Unit 3	**15L**
Autocorrelation: Concept, consequences of auto correlated disturbances, detection and solution of autocorrelation. Generalized least squares estimation.	
Unit 4	**15L**
Heteroscedastic disturbances: Concepts and efficiency of Aitken estimator with OLS estimator under heteroscedasticity. Consequences of heteroscedasticity. Tests and solutions of heteroscedasticity. Errors in variables: Correlation between error and regressors. Instrumental variable method (Single-equation model with one explanatory variable)	
PRACTICAL/LAB. WORK:	
1. Problems related to consequences of Multicollinearity. 2. Diagnostics of Multicollinearity. 3. Problems related to consequences of Autocorrelation. 4. Diagnostics of Autocorrelation. 5. Problems related to consequences Heteroscedasticity. 6. Diagnostics of Heteroscedasticity. 7. Estimation of problems of General linear model under Heteroscedastic distance terms. 8. Problems on Autoregressive models. 9. Problems on Instrumental variable.	

References:

1. Gujarati, D. and Sangeetha, S. (2007): Basic Econometrics, 4th Edition McGraw Hill Companies.
2. Johnston, J. (1972): Econometric Methods, 2nd Edition, McGraw Hill International.
3. Koutsoyiannis, A. (2004): Theory of Econometrics, 2nd Edition, , Palgrave Macmillan Limited.
4. Maddala, G.S. and Lahiri, K. (2009): Introduction to Econometrics, 4th Edition, John Wiley & Sons.

Data Structure and Database Management

Course Objective	1. Learn various data structures and abstract data types including lists, stacks, queues, trees, and graphs. 2. Understand various sorting and searching algorithms. 3. Learn a substantial, complex data structure. 4. To explain basic database concepts, applications, data models, schemas and instances. 5. To demonstrate the use of constraints and relational algebra operations. 6. Describe the basics of SQL and construct queries using SQL. 7. To emphasize the importance of normalization in databases. 8. To facilitate students in Database design. 9. To familiarize issues of concurrency control and transaction management.

Learning Outcomes	1. Construct and analysis various data structures and abstract data types including lists, stacks, queues, trees, and graphs. 2. Implement various sorting and searching algorithms. 3. Build a substantial, complex data structure. 4. Apply the basic concepts of Database Systems and Applications. 5. Use the basics of SQL and construct queries using SQL in database creation and interaction. 6. Design a commercial relational database system (Oracle, MySQL) by writing SQL using the system. 7. Analyze and Select storage and recovery techniques of database system.

Unit 1	**25L**
Tools for Analysis of Algorithms (Asymptotics, Recurrence Relations). Algorithms on arrays and matrices. Data Structures (Linked Lists and their variants, Stacks, Queues, Trees, Heaps and some variants) and applications. Sorting, Searching and Selection (Binary Search, Insertion Sort, Merge Sort, Quick Sort, Radix Sort, Counting Sort, Heap Sort etc.. Median finding using Quick-Select, Median of Medians). Basic Graph Algorithms (BFS, DFS, strong components etc.). Dijkstra's Shortest Paths algorithm, Bellman Ford algorithm, All pairs shortest path problem - Floyd Warshall's algorithm.	
Unit 2	**10L**
Introduction: Overview of Database Management System, Introduction to Database Languages, advantages of DBMS over file processing systems. Relational Database Management System: The Relational Model, Introduction to SQL: Basic Data Types, Working with relations of RDBMS: Creating relations e.g. Bank, College Database (create table statement).	
Unit 3	**10L**
Modifying relations (alter table statement), Integrity constraints over the relation like Primary Key, Foreign key, NOT NULL to the tables, advantages and disadvantages of relational Database System. Database Structure: Introduction, Levels of abstraction in DBMS, View of data, Role of Database users and administrators, Database Structure: DDL, DML, Data Manager (Database Control System).Types of Data Models Hierarchical databases, Network databases, Relational databases, Object oriented databases.	

Lab: Based on the lectures of the course. [30L]

References:

1. Gruber, M(1990): Understanding SQL, BPB publication.
2. Silberschatz, A, Korth, H and Sudarshan, S(2011) “Database System and Concepts”, 6th Edition McGrawHill.
3. Desai, B. (1991): Introduction to Database Management system, Galgotia Publications.

Semester-VII

STAUGMCC4716 Time Series and Sample Survey-II

Course Objective	1. Basics of stationary process. 2. Know the details of time series and their components. 3. Know the concept of estimation of different components of time series data. 4. Know the basics of Box-Jenkins approach. 5. Understand the details of Moving-average (MA) process and Autoregressive (AR) process of orders one and two.
Learning Outcomes	1. Understand and apply the concept of stationarity to the analysis of time series data in various contexts (such as actuarial studies, climatology, economics, finance, geography, meteorology, political science, and sociology). 2. Run and interpret time-series models and regression models for time series. 3. Using Yule-Walker equation and Least squares to fit AR (1) and AR (2) models to real life data. 4. Use the Box-Jenkins approach to model and forecast time-series data empirically. 5. Forecasting by exponential smoothing.

Theory	
Unit 2	**13L**

Review of SRS and stratified sampling
Systematic Sampling: Technique, estimates of population mean and total, variances of these estimates (N=nxk). Comparison of systematic sampling with SRS and stratified sampling in the presence of linear trend and corrections.

Unit 3 **12L**

Introduction to Ratio and regression methods of estimation, first approximation to the population mean and total (for SRS of large size), variances of these estimates and estimates of these variances, variances in terms of correlation coefficient for regression method of estimation and their comparison with SRS. Cluster sampling (equal clusters only) estimation of population mean and its variance, comparison (with and without randomly formed clusters). Relative efficiency of cluster sampling with SRS in terms of intra class correlation. Concept of sub sampling.

Unit 4 **20L**

Review of time series
Box-Jenkins modelling: Moving-average (MA) process and Autoregressive (AR) process of orders one and two. ACF and its graphical use in guessing the order of MA processes. Estimation of the parameters of AR (1) and AR (2) using least square and Yule-Walker equations. Forecasting: Exponential smoothing methods.

PRACTICAL/LAB. WORK:

1. Plotting a real life time series, and detecting various features (trend, periodic behaviours etc). Suggested data sets:
    a) Sun spot data
    b) Dollar-Rupee exchange rates
    c) Stock market data
2. Fitting and plotting of mathematical curves:
3. modified exponential curve
4. Gompertz curve
5. Fitting of trend by Moving Average Method.
6. Plotting detrended series.
7. Measurement of Seasonal indices Ratio-to-Moving Average method.
8. Plotting ACF of a given time series.
9. Using Yule-Walker equation and Least squares to fit AR (1) and AR (2) models to real life data
10. Forecasting by exponential smoothing.
11. Calculation of price and quantity index numbers.
12. Construction of Consumer and wholesale price index numbers.

References:

1. Chatfield C. (1980): The Analysis of Time Series –An Introduction, Chapman & Hall.
2. Kendall M.G. (1976): Time Series, Charles Griffin.
3. Brockwell and Davis (2010): Introduction to Time Series and Forecasting (Springer Texts in Statistics) ,2nd Edition.
4. Goon, A.M., Gupta, M.K. and Dasgupta, B. (2002): Fundamentals of Statistics, Vol. II, 8th Edn. The World Press, Kolkata.
5. Mudgett B.D. (1951): Index Numbers, John Wiley.

6. Allen R.G.D. (1975): Index Numbers in Theory and Practice, Macmillan.
7. Nagar A.L. & Das R. K. (1976): Basic Statistics.

STAUGMCC4717 Clinical Trials and Epidemiology

Course Objectives:

1. To learn and develop a scientific view to study the statistical challenges of clinical comparison of two or more treatments in human subjects.
2. Learn about the use of the cross-over design and its limitations.
3. To learn different methods of carrying out and analysing epidemiological studies.
4. To study pertinent issues such as appropriate design, data quality, analysis, and interpretation and presentation of results in environmental studies.

Learning Outcomes:

1. Understand the need and ethics of clinical trials.
2. Apply various designs of clinical trials to the data.
3. Describe optimal cross-over designs experiment with a continuous normally distributed outcome.
4. Understand designs based on clinical endpoints, drug interaction study.
5. Understand the basic epidemiology and carry out and analyse epidemiological studies.
6. To construct appropriate design, data quality, analysis, and interpretation and presentation of results in environmental studies.

Detailed Syllabus:

Introduction to clinical trials: the need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I – IV trials, multicenter trials. (4)

Data management: data definitions, case report forms, database design, data collection systems for good clinical practice. (6)

Design of clinical trials: parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, review of factorial designs, objectives and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials, design and monitoring of phase III trials with sequential stopping (7)

Reporting and analysis: analysis of categorical outcomes from Phase I – III trials, analysis of survival data from clinical trials. (6)

Introduction to Meta-analysis of clinical trials: Ideas of Meta Analysis, Fixed Effects Model, Random Effects Model, Analysis of Bias, Small sample effects (4)

Introduction to Epidemiology, Principles of Epidemiologic investigations, Different Epidemiologic measures (risk, relative risk, odds, odds ratio, incidence, prevalence), Confounding and interaction (Mantel–Haenszel methods, estimation, and tests) (12)

Design and Analysis of Epidemiologic studies, Epidemiological studies for certain diseases; Some modelling approaches for identifying the risk factors (11)

References:

1. S. Piantadosi (1997): Clinical Trials: A Methodologic Perspective. Wiley and Sons.
2. C. Jennison and B. W. Turnbull (1999): Group Sequential Methods with Applications to Clinical Trials, CRC Press.
3. L. M. Friedman, C. Furburg, D. L. Demets (1998): Fundamentals of Clinical Trials Springer Verlag.
4. J. L. Fleiss (1989): The Design and Analysis of Clinical Experiments. Wiley and Son.
5. E. Marubeni and M. G. Valsecchi (1994): Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley and Sons.
6. K.J. Rothman & S. Greenland : Modern Epidemiology
7. S.Selvin : Statistical Analysis of Epidemiologic Data
8. D. McNeil : Epidemiological Research Methods
9. D.C. Thomas : Statistical Methods in Genetic Epidemiology
10. J.F. Jekel, J.G. Elmore & D.L. Katz : Epidemiology, Biostatistics and Preventive Medicine

STAUGMDS4702 (WR & WOR)

Statistical Learning with Big Data-I

Course Objectives:

Statistical learning refers to a set of tools for modeling and understanding complex datasets. It is a recently developed area in statistics and blends with parallel developments in computer science and, in particular, machine learning. The objectives of this course are as follows:

1. To understand the concept of Big data, data Mining for enterprise data management and as a cutting edge technology tool.

2. To enable identifying data sources, processing, and imparting knowledge tools to analyze sets of data to gain useful business understanding.
3. To learn statistical learning which has become a very hot field in many scientific areas as well as marketing, finance, and other business disciplines.
4. To learn methods such as the ridge regression, lasso and sparse regression, classification.
5. To learn about cross-validation techniques and bootstrap.

Learning Outcomes:

1. Understand the Big data and data mining techniques.
2. Understand supervised learning techniques for univariate and multivariate data.
3. Apply classification and regression methods to real life problems in various fields.
4. Apply cross-validation techniques and bootstrap.

Detailed Syllabus:

Big data analysis- introduction, Big Data landscape, examples of real world big data problems, sources of Big Data. V's of Big Data and impacts on data collection, monitoring, storage, analysis and reporting. 5-step process to structure Big-data analysis. What are and what are not big data problems, recast big data problems as data science questions. explanation of the architectural components and programming models used for scalable big data analysis. (6)

Basic data mining tasks, Introduction to databases, including simple relational databases, data warehouses and introduction to online analytical data processing. Association rules and prediction, data attributes, applications to electronic commerce. (6)

Statistical Learning: Supervised and unsupervised learning, parametric and non-parametric methods of statistical learning. Regression and classification problem, Trade-Off Between Prediction Accuracy and Model Interpretability. (4)

Linear Regression: Review of Simple and Multiple linear regression, Qualitative Predictors, Potential Problems: Non-linearity of the response-predictor relationships, Correlation of error terms, Non- constant variance of error terms, Outliers, High-leverage points. Collinearity. K-Nearest Neighbor Regression, Comparison of Linear Regression with K-Nearest Neighbor. (8)

Classification: Logistic Regression, Estimation of the Regression Coefficients, Making Predictions, Multiple Logistic Regression, Logistic Regression for more than 2 Response Classes, Linear Discriminant Analysis, Application of Bayes Theorem for Classification, Linear Discriminant Analysis for $p = 1$, Linear Discriminant Analysis for $p>1$, Quadratic Discriminant Analysis, Comparison of Classification Methods. (8)

Resampling Methods: Cross-Validation, Validation Set Approach, Leave-One-Out Cross-Validation, k-Fold Cross-Validation (4)

Linear Model Selection and Regularization: Review of Subset Selection, Best Subset Selection, Stepwise Selection, Choosing the Optimal Model. Shrinkage Methods: Ridge Regression, Lasso, Selecting the Tuning Parameter. Dimension Reduction Methods: Principal Components Regression, Partial Least Squares. Considerations in High Dimensions: High-Dimensional Data, what goes Wrong in High Dimensions? Regression in High Dimensions, Interpreting Results in High Dimensions (14)

References:

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani: An Introduction to Statistical Learning with applications in R, Springer, 2013.
2. Berson, A. and Smith, S.J. (1997) Data Warehousing, Data Mining, and OLAP, McGraw-Hill.
3. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) Classification and Regression Trees, Wadsworth and Brooks/Cole.
4. Han,J. and Kamber.M. (2000) Data Mining; Concepts and Techniques, Morgan Kaufmann.
5. Mitchell, T.M. (1997) Machine Learning, McGraw-Hill.
6. Ripley, B.D. (1996) Pattern Recognition and Neural Networks, Cambridge University Press.
7. Sebastian Raschka, Vahid Mirjalili (2019); Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2; Packt Publishing
8. Samir Madhavan (2015), Mastering Python for Data Science; Packt Publishing
9. Andreas C. Mueller, Sarah Guido (2016), Introduction to Machine Learning with Python_ A Guide for Data Scientists; O'Reilly Media

Reliability Theory

Course Objectives:

1. To learn the reliability theory and the analysis of time to failure data.
2. To fit the censored and uncensored data using different parametric models.

Learning Outcomes:

1. Understand the elements of reliability, lifetime models, hazard function and its applications.
2. Understand the concept of censoring, life distributions and ageing classes.

Detailed Syllabus:

Reliability concepts and measures, components and systems, coherent systems, reliability of coherent systems. (6)

Life-distributions, reliability function, hazard rate, Mean residual life, common univariate life distributions – exponential, Weibull, gamma, lognormal, Rayleigh, piecewise exponential etc. Bivariate exponential. Reliability and expected survivability of series, parallel, mixed, maintained and non-maintained systems with and without redundancy, preventive maintenance policy. (15)

Notions of ageing – IFR, IFRA, NBU, DMRL and NBUE classes and their duals, preservation of such classes under reliability operations, Loss of memory property, Partial ordering of life distributions. (10)

Reliability estimation based on failure times from variously censored life-tests data for parametric families. (4)

Kaplan – Meier estimation of reliability curve, Greenwood formula, non-parametric methods for comparison of several reliability curves, Log rank tests. (5)

Regression models in reliability, Cox PH and Accelerated failure time models; Competing Risk Model; Estimation of parameters and diagnostics. (10)

References:

1. J.D. Kalbfleisch & R.L. Prentice : The Statistical Analysis of Failure Time Data, 2nd ed.
2. P.J. Smith : Analysis of Failure and Survival Data
3. R.E. Barlow and F. Proschan: Statistical Theory of Reliability and Life Testing
4. J.F. Lawless : Statistical Models and Methods for Lifetime Data
5. Nelson : Statistical models for failure time data

Advanced Probability

Course Objectives: The aim of the course is to focus on applications of measure theory in probability, understanding different modes of convergence and knowledge of Weak Law of Large Numbers, Strong Law of Large Numbers and the Central Limit Theorem with their applications.

Learning Outcomes: After successful completion of this course, student will be able to:

1. Understand the concepts of random variables, sigma-fields generated by random variables, probability distributions and independence of random variables related to measurable functions.

2. Gain the ability to understand the concepts of measurable functions and standard results from different theorems.

3. Acquire a good idea about the sequence of random variables and different modes of convergence.

4. Learn the concepts of weak and strong laws of large numbers and central limit theorem.

Detailed Syllabus:

Classes of sets, fields, sigma fields, minimal sigma field, Borel sigma field, sequence of sets, limsup and liminf of a sequence of sets. Measure, properties of a measure. Caratheodory extension theorem (statement only), Lebesgue and Lebesgue-Stieltjes measure. (16)

Measurable functions, sequence of measurable functions, integration of a measurable function with respect to a measure, monotone convergence theorem, Fatou's lemma, dominated convergence theorem. Integration of complex-valued functions, characteristic functions. Inversion and Continuity theorems. (16)

Radon-Nikodym theorem (statement and use), Product measure and Fubini's theorem (statement and use). Borel –Cantelli lemma. (4)

Sequence of random variables and modes of convergence (convergence in distribution, in probability, almost surely, and quadratic mean) and their interrelations. Scheffe's theorem, Slutsky's theorem. Laws of large numbers. Central Limit Theorems. Asymptotic normality. (14)

References:

1. A. K. Basu : Measure theory and Probability
2. P. Billingsley : Probability and Measure
3. J. F. C. Kingman & S. J. Taylor : Introduction to Measure and Probability
4. K. L. Chung: A Course in Probability Theory, 2nd Edition, Academic Press, New York.
5. B.R. Bhat: Modern Probability Theory, 3rd Edition, New Age International Publishers.

STAUGMDS4703 (WR & WOR)

Advanced Data Analytic Techniques

Course Objectives: The main aim to solve a wide range of business problems which require modelling, simulation and predictive analytical approaches.

Learning Outcomes: After successful completion of this course, student will be able to:

1. Apply different resampling techniques.
2. Handle the situation when there is missing data.
3. Analyze longitudinal data.
4. Apply the concepts of Generalized Linear Models in real life problems.

Detailed Syllabus:

Review of Resampling Techniques: Permutation tests. Introduction to Jackknife and Bootstrap-methods for estimating bias, standard error and distribution function based on iid random variables, Standard examples. Bootstrap confidence intervals (6)

Missing data analysis: Informative or non-informative missingness; MCAR, MAR and MNAR. Complete case / Available case estimation, Mean imputation, Hot and cold deck imputation; MICE. EM & MCEM algorithms and data augmentation techniques. (12)

Longitudinal data analysis: Longitudinal regression: Cohort vs longitudinal effect, Bias and efficiency. Robust estimation -Weighted least-squares; Robust standard error estimation. Parametric estimation: ML and REML. Marginal, subject specific and transition models for continuous, binary and count outcomes. Concept of GEE. (12)

Family of Generalized Linear Models: Exponential family of distributions, Formal structure for the class of GLMs, Link functions, Likelihood equations for GLMs, Important distributions for GLMs, A class of link functions-the power function, Overdispersion, Quasi likelihood. (14)
Models for Proportions: Binomial GLMs, Models for counts: Poisson GLMs (6)

References:

1. J.J. Faraway : Linear Models with R
2. J.J. Faraway : Extending the Linear Model with R
3. D. Ruppert et al. : Semiparametric Regression
4. R.J.A. Little & D.B.Rubin : Statistical Analysis with Missing Data
5. C.K. Enders : Applied Missing Data Analysis
6. M.A. Tanner : Tools for Statistical Inference
7. G.J. McLachlan & T. Krishnan : The EM Algorithm and Extensions
8. B. Efron & R.J. Tibshirani : An introduction to bootstrap
9. B.Efron : The jackknife, the bootstrap, and other resampling plans
10. B. Efron : Bootstrap methods – another look at jackknife
11. J. Shao & D. Tu : The Jackknife and Bootstrap

12. P.J. Diggle et. al. : Analysis of Longitudinal Data (2nd ed).

Statistical Genetics and Ecology

Course Objectives:

1. To learn the fundamental concepts of genetics and to apply statistical methods.
2. Understand the parametric growth models and single species growth models.
3. Study growth models in a random environment and testing the Goodness-of-fit in Growth curves.
4. Study stochastic differential equation model and Ito calculus.

Learning Outcomes:

1. Understand the hidden Markov models and parameter estimation techniques.
2. Understand the standard parametric growth models and single species growth models.
3. To model biological models in both deterministic and stochastic environments.
4. Perform goodness-of-fit tests for growth curves.

Detailed Syllabus:

Mendel's laws, Estimation of allele frequencies, Hardy-Weinberg law, Mating tables, Genotype frequencies with inbreeding, Disequilibrium constant, Inbreeding coefficient, Models of natural selection and mutation, Detection and estimation of linkage (recombination), Linkage analysis: Elston-Stewart algorithm, QTL mapping. (15)

Description of a DNA sequence. Pair-wise alignment-Needleman-Wunsch algorithm, Discrimination using Markov Chain, Hidden Markov Models and estimation of parameters (10)

Review of single species growth models: Exponential, Logistic and Gompertz etc., Growth in stochastic environment, stochastic differential models with application in Biology, Goodness-of-fit test for growth curves (25)

References:

1. D.L. Hartl : A Primer of Population Genetics
2. J. Ott : Analysis of Human genetic Linkage
3. P. Sham : Statistics in Human Genetics
4. R. Durbin, S. Eddyetal : Biological sequence analysis
5. Ben Hui Liu : Statistical Genomics

6. Linda J. S. Allen (2010): An Introduction to Stochastic Processes with Applications to Biology
7. Anil Gore & Sharayu Paranjpe (2001). A Course in Mathematical And Statistical Ecology, Kluwer Academic Publishers.
8. Gardner E.J. & Snustad D.P. Principles of Genetics, John Wiley & Sons Inc.
9. Lange, K (2002). Mathematical and Statistical Methods for Genetic Analysis, Springer.

Demography

Course Objectives:

1. To identify appropriate sources of data and to perform basic demographic analyses using various techniques across populations.
2. To learn the main theories used to understand population studies and societal change.

Learning Outcomes:

1. Understand the interdisciplinary nature of demography, balancing equation, use of Whipple's, Myers and UN indices.
2. Understand the measures of mortality and fertility.
3. Describe the concept of life tables.
4. Apply Quasi, Lotka's stable population models.

Detailed Syllabus:

Sources of demographic data: census and registration, Coverage and content errors in demographic data, Chandrasekharan—Deming formula to check completeness of registration data, adjustment of age data- use of Whipple, Myer and UN indices. population transition theory. (12)

Measures of fertility; stochastic models for reproduction, distributions of time of birth, inter-live birth intervals and of number of births (for both homogeneous and homogeneous groups of women), estimation of parameters; estimation of parity progression from open birth interval data. (10)

Measures of Mortality; construction of abridged life tables, infant mortality rate and its adjustments, model life table. (8)

Stable and quasi-stable populations, intrinsic growth rate. Models of population growth and their filling to population data. (8)

Internal migration and its measurement, migration models, concept of international migration. Methods for population projection, component method of population projection, Nuptiality and its measurements. (12)

References:

1. Kumar, R. (1986): Technical Demography, Wiley Eastern Ltd.
2. Benjamin, B. (1969): Demographic Analysis, George, Allen and Unwin.
3. Chiang, C.L. (1968): Introduction to Stochastic Progression.
4. Cox, P.R. (1970): Demography, Cambridge University Press.
5. Keyfitz, N. (1977): Introduction to the Mathematics of Population-with Revisions, Addison-Wesley, London.
6. Spiegelman, M. (1969): Introduction to Demographic Analysis, Harvard University Press.
7. Wolfenden, H.H. (1954): Population Statistics and Their Compilation, Am Actuarial Society.

STAUGMDS4704 (WOR only)

Statistical Computations using SAS

Course Objectives:

1. Learn and practice about Writing and compiling codes in SAS, writing reports.
2. Handle datasets in SAS.
3. Use loops and controls, functions and various procedures in SAS
4. Learn and practice statistical computations using SAS
5. To expose the students to the usage of various statistical techniques for analysis of data.
6. To provide the students hands-on experience of various statistical techniques and their applications
7. To develop computational skills to implement various statistical techniques taught in this semester.

Learning Outcomes: After successful completion of this course a student can

1. Write and execute programs in SAS
2. Handle datasets in SAS.
3. Perform various statistical computations using SAS.
4. Test hypothesis using SAS
5. Apply various statistical techniques taught in this semester.
6. Apply regression analysis technique in real life problems.
7. Apply multivariate techniques for real data.
8. Analyze Time Series data.

Detailed Syllabus:

Programming in SAS

Introduction to SAS: SAS variables, Libraries ,Windows, Parts of a SAS program, Data sets-Creation, Data step statements like CARDS, INGILE, DATA. Procedures in SAS, printing data sets

Control and Loops: Do-loops, IF-THEN-ELSE etc. Functions, Arrays.

Data Handling: Importing and Exporting data sets, Sorting data sets, creating new variables, subsetting data sets using DROP, KEEP, IF-THEN etc., Merging data sets, appending data sets, uses of CLASS, BY etc.

Descriptive Statistics: Proc statements MEANS, UNIVARIATE, SORT, FREQ, CORR, TABULATE

Graphical Representation: Proc GPLOT etc.

Simulation: generating random variables, simulating standard univariate and multivariate distributions.

Basic statistical tests: one and two sample t tests, Bartlett's test for variance, F test for equality of variances, multi sample means, Nonparametric tests, Chi squared tests, Exact tests and confidence intervals.

References:

1. Geoff Der and Brian S. Everitt : A Handbook of Statistical Analyses using SAS
2. Larry Hatcher: Step-by-Step Basic Statistics Using SAS: Student Guide
3. Larry Hatcher: Step-by-Step Basic Statistics Using SAS: Exercises
4. Robert A. Yaffee and Monnie McGee: Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS
5. Rick Wicklin : Simulating Data with SAS
6. Rick Wicklin : Statistical Programming with SAS/IML Software
7. SAS/ETS 9.2 User's Guide

Industrial Statistics

Unit I: Basic concepts of process monitoring and control. Review of control charts for attributes and variable data. O. C. and ARL of control charts. Cusum & V-masks charts. [15L]

Unit II: Concepts of AQL, LTPD, AOQL average amount of inspection and ASN functions. Acceptance sampling plans for attributes inspection, single, double and sequential sampling plans and their properties. Continuous sampling plans of Dodge type and their properties. [15L]

Unit III: Capability indices Cp, Cp k and Cp m , estimation, confidence intervals and tests of hypothesis relating to capability indices for normally distributed characteristics. [15L]

Unit IV: The weighted control charts: Exponential Weighted Moving Average chart. Multivariate SPC: Multivariate quality control problem, description of Multivariate data, The Hotelling T 2 control chart, Multivariate EWMA control chart, regression adjustment, Latent structure methods. Quality Systems: ISO 9000 standards, QS 9000 standards, concept of six sigma. Total Quality management. Taguchi Design. [15L]

REFERENCES:

1. Montgomery D.C. (1996) Introduction to Statistical Quality Control, Wiley.
2. Wetherill G.B. (1977) Sampling Inspection & Quality Control, Halsted Press.
3. Logothetis N. (1992) Managing Total Quality, Prentice Hall of India.
4. Oakland J.S. (1989) Total Quality Management; Butterworth- Heinemann.
5. Mittog H.J. and Rinne H. (1993) Statistical Methods of Quality Assurance.
6. Guenther W.C (1981) Sampling Inspection in Statistical Quality Control Charter Grifits.
7. Kotz S. (1993) Process capability indices, Chapman and Hall.
8. Abraham Bovas (1998) Quality Improvement through statistical methods
9. Barlow R.E. And Proschan F. (1985) Statistical Theory of methods reliability and Life Testing, Holt Rinehart and Winston.
10. Lawless J.F. (1982) Statistical Models and methods of life Time Data, John Wiley.
11. Bain L.J. And Engelhard (1991) Statistical Analysis of Reliability and Life Testing models Marcel Dekker.
12. Nelson W. (1982) Applied Life Data Analysis, John Wiley.
13. Zacks S. (1992) Introduction to reliability analysis Probability Models and statistical Methods, Springer-verlag.
14. Mahajan M. (2004) Statistical Quality Control.

Decision Theory

Unit I: Decision problem, loss function, risk function, randomized and non- randomized decision rule. Decision principles (Conditional Bayes, Frequentist). Testing and estimation problem as decision problems. Optimal decision rule. [15L]

Unit II: Concept of admissibility and completeness, Bayes rules, minimax decision rule. Admissibility of Bayes rules. Existence of Bayes decision rule. [15L]

Unit III: Definition of non-parametric test, advantages, and disadvantages of nonparametric tests. Single sample problems. (i) Test of randomness (ii) Tests of goodness of fit: Empirical distribution function. Kolmogorov-Smirnov test, comparison of Chi-square and KS test. (iii) Problem of location: Sign test, Wilcoxon's signed rank test, Wilcoxon paired sample signed rank test. [12L]

Unit IV: Two Sample Problems: Different types of alternatives, sign test, Wilcoxon two sample rank sum test, Wald-Wolfiwitz run test, Mann-Whitney-Wilcoxon test, median test. K-S two sample test. One sample U statistic, Kernel and symmetric Kernel, variance of U statistic, two sample U statistics, linear rank statistics and their distribution properties under null hypothesis. [12L]

REFERENCES:

1. Ferguson T. S. (1967) Mathematical Statistics, Academic Press, New York.

2. Fraser, D.A.J. (1957) Non-parametric methods in Statistics, John Wiley.
3. Gibben J.D. (1992) Non-Parametric Statistical inference, Marcel Dekker, Inc., New York.
4. Goon A.M., Gupta M.K., Dasgupta: An Outline of Statistical Inference. The World Press Pvt. Ltd.
5. Berger, J.O. (1980) Statistical Decision Theory: Foundations, Concepts and Methods, Springer-Verlag.
6. Berger, J.O. (1985) Statistical Design Theory and Bayesian Analysis, 2nd ed., Springer-Verlag.
7. Gupta S. S. and Huang, D. (1981) Multiple Statistical Decision Theory, Springer-Verlag, New York.

STAUGPRJ01 Research Work (WR only)

Objective	
The aim of the course is to initiate students to write and present a statistical report, under the supervision of a faculty, on some area of human interest. The project work will provide hands on training to the students to deal with data emanating from some real-life situation and propel them to dwell on some theory or relate it to some theoretical concepts.	

Semester-VIII

STAUGMCC4818 Stochastic Processes

Course Objectives:	1. Learn Stochastic processes and deterministic processes and their distinction and examples. 2. Understand stochastic processes predictive approach. 3. Develop an ability to analyze and apply some basic stochastic processes for solving real life situations.
Learning Outcomes:	At the end of the course a student will be able to 1. Understand the stochastic processes, Markov chains, Transition probability matrix and various types of states. 2. Explain Random walk, Gambler ruins problem and apply Poisson process in real life situations. 3. Formulate and solve problems which involve setting up stochastic models. 4. Explain when stationary distribution exists and able to derive the stationary distribution if exists 5. Understand branching processes with applications. 6. Model continuous time and discrete state space stochastic process and study their behavior. 7. Model Continuous time and continuous state space stochastic processes using Brownian motion process.

Detailed Syllabus:

Unit-1: [20L]

Introduction to stochastic processes. Markov chains with finite and countable state space, classification of states, Chapman - Kolmogorov equations. Calculation of n-step transition probability and its limit. Stationery distribution. Random walk. gambler's ruin problem

Branching process. Galton-Watson branching process, estimation of probability of extinction.

Unit-2: [10L]

Discrete state space continuous time Markov chains. Poisson process, Birth and death process, Applications to queueing problems. Renewal theory: Statement and uses of key renewal theorem.

Unit-3: [15L]

Brownian Motion: Limit of Random Walk, Its Defining Characteristics and Peculiarities. Its Variations: Standard Brownian Motion, Brownian Bridge

References:

1. David F. Anderson: Introduction to Stochastic Processes with Applications in the Biosciences
2. S. M. Ross : Introduction to Probability Models
3. S. M. Ross : Stochastic Process
4. S. Karlin & H.M. Taylor : A First Course in Stochastic Processes
5. J. Medhi : Stochastic Process
6. A.K. Basu : Stochastic Process
7. R.N. Bhattacharyya & E. Waymire : Stochastic Processes and Applications

STAUGMCC4819 Multivariate Analysis-II

Course Objectives:	1. To learn and develop a scientific view to deal with multidimensional datasets and its uses in the analysis of research data. 2. To understand the extensions of univariate techniques to multivariate frameworks and learn to apply dimension reduction techniques used in the data analysis.
Learning Outcomes:	1. Understand multivariate normal distribution and their real-life applications. 2. Understand Wishart distribution, Hotelling T^2 and Mahalanobis's D^2 statistics. 3. Implement dimension reduction techniques using software on real life problems. 4. Demonstrate knowledge and understanding of the basic ideas behind discriminant and clustering analysis techniques with applications.

Detailed Syllabus:

Multivariate normal distribution and its properties. Sampling from Multivariate normal distribution – independence of sample means vector and variance-covariance matrix. Wishart distribution. Distribution of quadratic forms – Cochran's theorem. (8)

Distributions of partial and multiple correlation coefficients and regression coefficients. Hotelling T^2 and Mahalanobis's D^2 application in testing and confidence set construction. (8)

Multivariate linear model: estimation of parameters, tests of linear hypotheses (6)

Multivariate Analysis of variance of one-way and two-way classified data, simultaneous confidence intervals, Multivariate Analysis of Covariance. (8)

Principal Component Analysis: Population and sample Principal components and their uses. Plotting techniques. (6)

Factor Analysis: The orthogonal factor model, Estimation of factor loading, Estimation of Factor scores, Interpretation of Factor Analysis. (6)

Cluster Analysis (4)

Classification and discrimination procedures for discrimination between two multivariate normal populations- sample discriminant function. (4)

References:

1. C.R.Rao : Linear Statistical Inference and its Applications
2. T.W.Anderson : Introduction to Multivariate Analysis
3. A. M. Kshirsagar : Multivariate Analysis
4. S. S. Wilks : Mathematical Statistics
5. G A F Seber: Multivariate Observations
6. M.S.Srivastava & C. G. Khatri : Introduction to Multivariate Statistics
7. R.J.Muirhead : Aspects of Multivariate statistical Theory

STAUGMDS4805 (WR & WOR)

Statistical Learning with Big Data-II

Course Objectives:

This is the continuation of the course **Statistical Learning with Big Data-I.** The objectives of this course are as follows:

1. To focus on non-linear models using polynomial regression and step functions, as well as more sophisticated approaches such as splines, local regression, and generalized additive models.
2. To learn tree-based methods for regression and classification problems
3. To learn the support vector machine (SVM), an approach for classification.
4. To learn the basics of neural networks and deep learning, and some problems, such as convolutional neural networks (CNNs) for image classification, and recurrent neural networks (RNNs) for time series and other sequences.
5. To learn about various unsupervised learning methods: principal components analysis for data visualization or data pre-processing, and clustering.

Learning Outcomes:

1. Fit non-linear models using polynomial regression and step functions, splines, local regression, and generalized additive models.
2. Apply tree-based methods for regression and classification problems
3. Use a support vector machine (SVM) for classification.
4. Fitting a Neural Network for image classification, time series and other sequences.
5. Apply various unsupervised learning methods such as principal components analysis for data visualization or data pre-processing, and clustering.

Detailed Syllabus:

Moving Beyond Linearity in Regression: Polynomial Regression, Step Functions, Basis Functions. Regression Splines: Piecewise Polynomials, Constraints and Splines, The Spline Basis Representation, Choosing the Number and Locations of the Knots, Comparison to Polynomial Regression. Smoothing Splines: An Overview of Smoothing Splines, Choosing the Smoothing Parameter λ. Local Regression. Generalized Additive Models: GAMs for Regression Problems, GAMs for Classification Problems

(10)

Tree-Based Methods: Basics of Decision Trees, Regression Trees , Classification Trees, Trees Versus Linear Models, Advantages and Disadvantages of Trees. Bagging, Random Forests, Boosting, Bagging, Random Forests, Boosting (10)

Support Vector Machines: Maximal Margin Classifier, Support Vector Classifiers, Support Vector Machines, SVMs with More than Two Classes, Relationship to Logistic Regression. (8)

Deep Learning: Single Layer Neural Networks, Multilayer Neural Networks, Convolutional Neural Networks: Convolution Layers, Pooling Layers, Architecture of a Convolutional Neural Network, Data Augmentation, Document Classification. Recurrent Neural Networks and applications in Sequential Models for Document Classification, Time Series Forecasting. Uses of Deep Learning, Fitting a Neural Network. (14)

Unsupervised Learning: Principal Components Analysis, Clustering Methods, K-Means Clustering, Hierarchical Clustering. Clustering methods from both statistical and data mining viewpoints, vector quantization. (8)

References:

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani: An Introduction to Statistical Learning with applications in R, Springer, 2013.
2. T. Hastie, R. Tibshirani & J. Friedman : The Elements of Statistical Learning
3. B.L. Friedman, et al. : Classification and Regression Trees
4. R.A. Johnson & D.W. Wichern : Applied Multivariate Statistical Analysis
5. Mitchell, T.M. (1997) Machine Learning, McGraw-Hill.
6. Sebastian Raschka, Vahid Mirjalili (2019); Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2; Packt Publishing
7. Samir Madhavan (2015), Mastering Python for Data Science; Packt Publishing
8. Andreas C. Mueller, Sarah Guido (2016), Introduction to Machine Learning with Python_ A Guide for Data Scientists; O'Reilly Media

Statistical Methods for Bio-computing

Actuarial Statistics

Course Objectives:

1. To learn the life tables used in insurance products.
2. To learn the concept of interest, different life insurance products, life annuities, net premiums.

3. To motivate students to prepare for exams required for employment in the actuarial science profession.

Learning Outcomes:

1. Understand the utility theory, insurance products and life tables.
2. Understand the concept of interest.
3. Understand the concept of life insurance and the existing insurance products of different insurance companies.
4. Know life annuities, net premium, and net premium reserves.

Detailed Syllabus:

Review of decision theory and actuarial applications. (4)

Loss distributions: modelling of individual and aggregate losses, moments, fitting distributions to claims data, deductibles and retention limits, proportional and excess-of-loss reinsurance, share of claim amounts, parametric estimation with incomplete information. (10)

Risk models: models for claim number and claim amount in short-term contracts, moments, compound distributions, moments of insurer's and reinsurer's share of aggregate claims. (8)

Review of Bayesian statistics/estimation and application to credibility theory. (4)

Experience rating: Rating methods in insurance and banking, claim probability calculation, stationary distribution of proportion of policyholders in various levels of discount. (4)

Delay/run-off triangle: development factor, basic and inflation-adjusted chain-ladder method, alternative methods, average cost per claim and Bornhuetter-Ferguson methods for outstanding claim amounts, statistical models. (8)

Review of generalized linear model, residuals and diagnostics, goodness-of-fit, applications. (4)

Review of time series analysis, filters, random walks, multivariate models, cointegrated time series, non-stationary/non-linear models, application to investment variables, forecasts. Assessment of methods through Monte-Carlo simulations. (8)

References:

1. N. L. Bowers, H. U. Gerber, J. C. Hickman, D. A. Jones and C. J. Nesbitt, Actuarial Mathematics, Society of Actuaries, Itasca, IIIinois, U. S. A. 2 nd d.(1997)
2. Deshmukh S.R. (2009) An Introduction to Actuarial Statistics Using R, Uni. Press.

3. Spurgeon E. T. (1972) Life Contingencies, Cambridge University Press.
4. Neill, A. (1977) Life Contingencies, Heinemann.

STAUGMDS4806 (WOR only)

Advanced Regression Analysis

Course Objective:

In this course a student will:

1. Learn association between nominal-ordinal, nominal-nominal, and ordinal-ordinal type table.
2. Understand the role of generalized linear modelling techniques (GLMs) in modern applied statistics and implement methodology.
3. Learn the techniques and applications of generalized linear model for different data type.
4. Understand the underlying assumptions for GLMs and perform diagnostic checks whilst identifying potential problems.
5. Learn how to fit the model for binary and count type data.
6. Learn the methodology to check for over-dispersion of the model through suitable techniques.

Learning Outcomes:

After Completion of this course a student will

1. Find the association between nominal-ordinal, nominal-nominal, and ordinal-ordinal type table.
2. Find the role of generalized linear modelling techniques (GLMs) in modern applied statistics and implement methodology.
3. Apply generalized linear model for different data type.
4. Perform diagnostic checks whilst identifying potential problems.
5. Fit the model for binary and count type data.
6. Check for over-dispersion of the model through suitable techniques.

Detailed Syllabus:

Measures of association for classified nominal and ordinal categorical data. Generalized Linear Models: Introduction, Components of a GLM, Maximum Likelihood estimation, Deviance.

Binary data and Count data: un-grouped and grouped. Models with constant coefficient of variation. Polytomous data.

over dispersion and fitting by quasi-likelihood.

Linear Programming Problems and Game Theory

Course Objective	1. Introduction to Linear programming problem and related terms associated with LPP. 2. Basics of simplex method. 3. Concepts of duality. 4. Introduction of Transportation problem. 5. Details of Game theory.
Learning Outcomes	1. Mathematical formulation of L.P.P and solving the problem using graphical method, Simplex technique and Charne's Big M method involving artificial variables. 2. Identifying Special cases by Graphical and Simplex method and interpretation of degenerate solution, unbounded solution, alternate solution, and infeasible solution. 3. Solve allocation problem using Transportation model. 4. Solve allocation problem using Assignment model. 5. Solve the problems based on mixed game strategy.

Theory	
Unit 1	**10L**
Introduction and Historical Background, Phases of Operations Research, model building, various types of O.R. problems. Linear Programming Problem, Requirements of LPP, Mathematical Formulation of LPP, Graphical Methods to Solve Linear Programming Problems. Convex sets, Extreme point.	
Unit 2	**8L**
Simplex method for solving L.P.P. Charne's M-technique for solving L.P.P. involving artificial variables. Special cases of L.P.P. Concept of Duality in L.P.P: Dual simplex method.	
Unit 3	**12L**
Introduction, Formulation of Transportation Problem (TP). Initial solution by North West corner rule, Least cost method and Vogel's approximation method (VAM), MODI's method to find the optimal solution, special cases of transportation problem. Assignment problem: Hungarian method to find optimal assignment, special cases of assignment problem.	
Unit 4	**15L**
Game theory: Introduction, Competitive Situations, Characteristics of Competitive Games. Rectangular game, Two-Person Zero-Sum game, minimax-maximin principle, solution to rectangular game using graphical method, dominance and modified dominance property to reduce the game matrix and solution to rectangular game with mixed strategy.	
PRACTICAL/LAB. WORK:	30L
1. Mathematical formulation of L.P.P and solving the problem using graphical method, Simplex technique and Charne's Big 2. M method involving artificial variables. 3. Identifying Special cases by Graphical and Simplex method and interpretation a. Degenerate solution b. Unbounded solution c. Alternate solution	

d. Infeasible solution
4. Allocation problem using Transportation model.
5. Allocation problem using Assignment model.
6. Problems based on game matrix.
7. Graphical solution to mx2 / 2xn rectangular game.
8. Mixed strategy.

References:

1. Taha, H. A. (2007): Operations Research: An Introduction, 8 Hall of India.
2. KantiSwarup, Gupta, P.K. and Manmohan (2007): Operations Research, 13th Edition, Sultan Chand and Sons.
3. Hadley, G: (2002) : Linear Programming, Narosa Publications.
4. Hillier, F.A and Lieberman, G.J. (2010): Introduction to Operations Research- Concepts and cases, 9th Edition, Tata McGraw Hill.

Machine Learning

In this course a student will learn:

1. The techniques of linear regression in the context of machine learning.
2. How to select best model.
3. Different linear regression techniques like Ridge regression and Lasso.
4. The classification techniques like logistic regression, Naïve Bays.
5. The applications of tree-based methods and support vector machine.

In this course a student will:

1. Apply linear regression in the context of machine learning.
2. Select best model.
3. Apply different linear regression techniques like Ridge regression and Lasso.
4. Apply classification techniques like logistic regression, Naïve Bays.
5. Apply tree-based methods and support vector machine.

Detailed Syllabus:

Unit-1:
Linear Regression: Basics of linear regression, its assumptions, limitations and industry applications.
Linear Regression Assessment: A car price prediction assignment.

Advanced Regression: Generalized Linear Regression and Regularized Regression techniques like Ridge and Lasso.

Unit-2:
Model Selection: Bias-Variance Tradeoff, Hyperparameter Tuning and Cross-Validation which are necessary to finalize the best ML model.

Unit-3:
Logistic Regression: Univariate Logistic Regression for classification ML. Implementation in Python, evaluation metrics and industry applications are covered.
Naive Bayes: basics of Bayes Theorem, Naive Bayes classifier and implementation in a Spam-Ham classifier.

Unit-4:
Support Vector Machine (Optional): SVM algorithm, its working, kernels and implementation.
Tree Models: Basics of Tree models, their structure, splitting techniques, pruning and ensembles to form Random Forests are covered here.

STAUGMDS4807 (WOR only)

Simulation and Data Analysis

Course Objectives:

In this course students will

1. Learn the methods to simulate data from different probability distribution and to apply these using some computer package like R/Python/SAS.
2. Learn the methodology to fit different probability distributions and their applications.
3. Learn how to solve real world problems which cannot be solved strictly by mathematical approaches.
4. Learn computer intensive methods like Jack-Knife, Bootstrap, cross validation, Monte Carlo methods.

Learning Outcomes:

After completing this course, a student will be able to

1. Simulate data from different probability distribution and apply these using some computer package like R/Python/SAS.
2. Fit different probability distributions and their applications.
3. Solve real world problems which cannot be solved strictly by mathematical approaches.

4. Apply computer intensive methods like Jack-Knife, Bootstrap, cross validation, Monte Carlo methods.

Detailed syllabus:

Simulation of random variables from discrete, continuous, multivariate distributions and stochastic processes, Monte-Carlo methods. Regression analysis, scatter plot, residual analysis. Computer Intensive Inference Methods - Jack-Knife, Bootstrap, cross validation, Monte Carlo methods and permutation tests, Importance sampling, Metropolis Hastings Algorithm.

Financial Statistics

Theory	
Unit 1	**10L**
Probability review: Real valued random variables, expectation and variance, skewness and kurtosis, conditional probabilities and expectations. Discrete Stochastic Processes, Binomial processes, General random walks, Geometric random walks, Binomial models with state dependent increments.	
Unit 2	**10L**
Tools Needed For Option Pricing: Wiener process, stochastic integration, and stochastic differential equations. Introduction to derivatives: Forward contracts, spot price, forward price, future price. Call and put options, zero-coupon bonds and discount bonds.	
Unit 3	**20L**
Pricing Derivatives: Arbitrage relations and perfect financial markets, pricing futures, put-call parity for European options, relationship between strike price and option price. Stochastic Models in Finance: Discrete time process- binomial model with period one.	
Unit 4	**20L**
Stochastic Models in Finance: Continuous time process- geometric Brownian motion. Ito's lemma, Black-Scholes differential equation, Black-Scholes formula for European options, Hedging portfolios: Delta, Gamma and Theta hedging. Binomial Model for European options: Cox-Ross-Rubinstein approach to option pricing. Discrete dividends.	
PRACTICAL/LAB. WORK:	
1. To verify "no arbitrage" principle 2. To verify relationship between spot price, forward price, future price 3. To price future contracts 4. To verify put-call parity for European options 5. To construct binomial trees and to evaluate options using these trees 6. To price options using black – Scholes formula 7. To hedge portfolios using delta and gamma hedging 8. To hedge portfolios theta hedging 9. Pricing of call options using binomial model 10. Computation of dividends on call options as a percentage of stock price. 11. Computation of dividends on call options as a fixed amount of money. 12. Pricing of put options using binomial model	

13. Call-put parity for options following binomial models.

14. Effect of dividends on put options.

References:

1. Bhat,B.R.(2000): Stochastic Models: Analysis and Applications, New Age International Publishers.
2. Taha, H. (1995): Operations Research: An Introduction, Prentice- Hall India.
3. Feller, William (1968): Introduction to probability Theory and Its Applications, Vol I,3rd Edition, Wiley International.
4. Chatfield C. (1980): The Analysis of Time Series –An Introduction, Chapman & Hall.
5. Kendall M.G. (1976): Time Series, Charles Griffin.
6. Brockwell and Davis (2010):Introduction to Time Series and Forecasting (Springer Texts in Statistics) ,2nd Edition.

Course Objective	1. Review of discrete stochastic process. 2. Introduction to wiener process, stochastic integration and stochastic differential equation. 3. Introduction to derivatives and related terms. 4. Details of pricing of derivatives. 5. Uses of some stochastic models in finance.
Learning Outcomes	1. Solving problems associated with wiener process, stochastic integration and stochastic differential equation. 2. Understand relationship between spot price, forward price, future price. 3. Calculate price future contracts. 4. Verify put-call parity for European options. 5. Construction binomial trees and to evaluate options using these trees. 6. Derive price options using black – Scholes formula. 7. To hedge portfolios using delta and gamma hedging. 8. Pricing of call options using binomial model. 9. Computation of dividends on call options as a percentage of stock price. 10. Call-put parity for options following binomial models. 11. Calculate effect of dividends on put options.

Data Mining Techniques

Course Objective: To learn basic data mining techniques and their handling using R/Python.

Learning Outcomes:

1. Understand fundamentals of data mining.
2. Know feature and applications of data mining.
3. Understand data warehousing, OLAP, OLTP, Data visualization.
4. Implement and interpret the results of data scientifically using R/Python.

Detailed Syllabus:

UNIT I: Basic data mining tasks, Introduction to databases, including simple relational databases, data warehouses and introduction to online analytical data processing. Association rules and prediction, data attributes, applications to electronic commerce. (12L+3T)

UNIT II: Unsupervised learning from univariate and multivariate data, Dimension reduction and feature selection. (12L+3T)

UNIT III: Supervised learning from moderate to high dimensional input spaces, artificial neural networks and extensions of regression models, regression trees. (12L+3T)

UNIT IV: Review of classification methods from multivariate analysis, classification and decision trees. Clustering methods from both statistical and data mining viewpoints, vector quantization. (12L+3T)

REFERENCES:

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