Department of Statistics
Assam University, Silchar

M.A/M.Sc Statistics Syllabus (CBCS)
Applicable from 2021-22 Session

Statistics Syllabus CBCS Pattern (New Course)
Master in Science (M.Sc)
Department of Statistics, Assam University, Silchar

Course No.	Title
Semester I	
101	Real Analysis
102	Linear Algebra
103	Probability Theory
104	Distribution Theory
105	Statistical Computing
Semester II	
201	Linear Models and Regression Analysis
202	Survey Sampling
203	Fundamentals of Data Collection and Analysis (CBCS paper open to PG students of other departments)
204	Statistical Inference I (CBCS paper open to all PG students)
205	Statistical Computing
Semester III	
301	Design and Analysis of Experiments
302	Statistical Inference II
303	Stochastic Process
304	Applied Statistics I
305	Statistical Computation in SPSS (Practical Paper)
Semester IV	
401	Multivariate Analysis
402	Applied Statistics II
403	(i) Industrial Statistics and Optimization Techniques I or (ii) Actuarial Statistics I
404	(i) Industrial Statistics and Optimization Techniques II or (ii) Actuarial Statistics II
405	Project Work and Practical

Note:

1. Each of the papers shall be of equal credit.
2. Students shall be exposed both to statistical theory and statistical computing.
3. Any one of the two modules in paper 403 and 404 will be offered to the students as specialization, i.e. either 403(i) and 404(i) or 403(ii) and 404(ii).
4. The Project Work shall be an application of statistical theory to practice. Students are supposed to identify a practical problem, collect necessary data, analyse them using statistical methods and report the findings.

Program Objectives:

1. Provide students with a firm background in Mathematical tools that requires to pursue higher studies in Statistics.
2. Offer the students with courses that shall build their fundamental understanding in Statistics- which includes Probability, Probability Distributions, and Statistical Inference etc.
3. To provide training on the theoretical concepts of model building in Applied Statistics and their practical applications.
4. Deliver training to the students on important subfields of Applied Statistics like Designs of Experiment, Demography, Survey Sampling, Time Series Analysis etc.
5. Improve the students with advanced knowledge of mathematical statistics
6. Arrange for necessary training is Statistical packages, spreadsheets for Statistical computing as well as designing computer program for statistical computing.
7. Develop and implement data analysis strategies based on theoretical principles, and detailed knowledge of the underlying data through projects and computer programs
8. Develop the ability to analyse data sets in the context of real world problems and interpret results using statistical tools, writing statistical reports and apply software for computation.

Program Outcomes:

After successful completion of the course students will be able to:

(i) Gain sound knowledge of both theoretical and practical aspects of Statistics for its proper use and applications

(ii) Explain the complex Statistical techniques and ideas to non-statisticians

(iii) Handle and analyse large datasets in the context of real world problems, use and interpret the results for practical purpose.

(iv) Get wide range of job opportunities in Government, Private sectors including small and big industries as data Consultant, Statistician, Decision making, Planning etc.

(v) Have expertise in Statistical thinking to create new theories, formulas, techniques and applications.

(vi) Have sound knowledge to pursue research works in Statistics and relevant discipline.

First Semester
Paper 101
Real Analysis

Course Outcome: Understand the basic concepts of real analysis including completeness of set, suprimum, infimum, sequence and series and their convergence, real-valued function, continuous function, uniform continuity and convergence, power series, Fourier series , unconstrained and constrained optimization problems with several variables, double and multiple integrals, uniform convergence in improper integrals and differentiation under the sign of integral.

Learning Outcome: These concepts of real analysis are necessary in the study of probability theory, distribution theory, statistical inference, stochastic process, Bayesian inference, linear and non-linear models, optimization techniques, multivariate statistics and other topics of statistics. On studying this paper the students are expected to develop a mathematical foundation that is necessary for understanding various advanced topics of statistics.

Unit-I: Introduction to Real Analysis

- Review of sets and functions, countable and uncountable sets.
- Introduction to real numbers: bounded and unbounded sets, suprimum and infimum of a set, completeness property of the set of real numbers.
- Open and closed sets, limit points and Bolzano-Weirestrass theorem (for sets), compact sets and Heine-Borel theorem.

(9 lectures)

Unit-II: Sequence and Series

- Sequences: Covergent sequences, bounded sequences, Cauchy's general principle of convergence.
- Infinite series, different tests of convergence of series (without proof).
- Continuous functions and their properties, uniform continuity. (9 lectures)

Unit-III: Mean Value Theorems

- Differentiability of functions.
- Mean value theorems and their applications, Taylor's theorem and maxima-minima of functions of one real variable.
- Uniform convergence, power series and Fourier series. (9 lectures)

Unit-IV: Maxima-Minima

- Functions of several variables: continuity and differentiability.

- Maxima-minima of functions of several variables, constrained maxima-minima of functions.

(9 lectures)

Unit-V: Multiple and Improper Integrals

- Double and multiple integrals, evaluation of multiple integrals by repeated integration, change of variables in multiple integration.
- Uniform convergence in improper integrals, differentiation under the sign of integral.
- Integral transform: Fourier transform. (9 lectures)

References

1. T.M. Apostol (1974), **Mathematical Analysis**, 2nd edition, *Narosa Book Distributors Pvt. Ltd.*
2. R. G. Bartle and D. R. Sherbert (2000) **Introduction to Real Analysis**, 4th edition, *Wiley India Pvt. Ltd.*
3. S. C. Malik and S. Arora (2017) **Mathematical Analysis**, 5th edition, *New Age International Publishers.*
4. A. Kumar and S. Kumaresan, (2014) **A Basic Course in Real Analysis**, 1st edition, *CRC Press.*
5. Royden HL and Fitzpatrick PM (2015) **Real Analysis**, 4th Edition, Pearson.

First Semester
Paper 102
Linear Algebra

Course Outcome: The paper helps in understanding the basic concepts of linear algebra including matrices, vector space, linear dependence, basis and dimension of vector space, norms and orthogonality, orthogonal projections, Gram-Schmidt orthogonalization process, Eigen values and Eigen vectors, generalized inverse, quadratic forms and solution of system of linear equations.

Learning Outcome: The concepts of linear algebra shall help the student in the study of multivariate analysis, time series, linear models and regression, design of experiments, stochastic process, bivariate and multivariate distributions, statistical inference, stochastic process, Bayesian inference, linear and non-linear models, optimization techniques, and several other topics of statistics they shall encounter in other courses of the same program. This, paper lays the foundation of the student in further understanding of the subject.

Unit I: Matrices, System of linear equations and Determinants

- Review of matrices, vectors and their operations, Elementary operations on matrices
- Echelon and canonical forms, Rank of a matrix
- System of linear equations, their solution by Gaussian elimination and Gauss-Jordan elimination
- Determinant of a matrix, Inverse of a matrix. (9 lectures)

Unit II: Vector Space

- Vector spaces, Subspaces
- Linear combination of vectors, linear span
- Linear dependence and independence of vectors, Basis and dimension of vector spaces. (9 lectures)

Unit III: Linear transformation of a matrix

- Linear transformations, Rank and nullity of a linear transformation, rank-nullity theorem
- Matrix representation of linear transformations

- Inner product spaces, norms and orthogonality, orthogonal projections
- Gram-Schmidt orthogonalization process.

(9 lectures)

Unit IV: Eigenvalues and eigenvectors, Matrix polynomials

- Eigenvalues and eigenvectors of a matrix, eigenspaces and multiplicities
- Diagonalizable matrices, Similarity with triangular matrices, Matrix polynomials
- Cayley-Hamilton theorem, Spectral decomposition of real symmetric matrices.

(9 lectures)

Unit V: Generalized Inverse and Quadratic forms

- Generalized inverse of a matrix, Generalized inverse and linear systems
- Moore and Penrose inverse, Quadratic forms, Matrix in quadratic forms, Positive and nonnegative definite matrices, minors
- Simultaneous diagonalization and the generalized eigenvalue problem. (9 lectures)

References

1. Kunze Ray, Hoffman Kenneth (2008) Linear Algebra, 2nd Edition, PHI Learning Pvt. Ltd.
2. Seymour Lipschutz, Marc Lipson (2005) Linear Algebra 3rd Edition, Tata McGraw Hill.
3. Seymour Lipschutz, Marc Lipson (2017) Schaum's Outlines of Linear Algebra, Third Edition, McGrow Hills Inc.
4. Biswas, S. (2012). Textbook of Matrix Algebra, PHI, New Delhi.
5. Bhimasanaram P., A. Ramachandra Rao (2010) Linear Algebra (Texts and Readings in Mathematics) 2nd Revised edition, Hindustan Book Agency.
6. Banerjee, S., Roy, A. (2014). Linear Algebra and Matrix Analysis for Statistics, First Edition, CRC Press.
7. Dasgupta A. (2014) Abstract and Linear Algebra, Ashoke Prakashan, Kolkata.
8. Khanna, V. K. and Bhambri, S. K. (2016). A Course in Abstract Algebra, Fifth Edition, Vikas Publishing House.

First Semester
Paper 103
Probability Theory

Course Outcome: This paper shall give the students knowledge about the basic concepts of probability measure, measurable function, integrability of functions, probability space, theorems of probability, functions of random variables, distribution function, convergence of sequence of random variables, generating function and their convergence, law of large numbers, central limit theorems and urn models.

Learning Outcome: Knowledge of probability is the basis of understanding the other domains of statistics. The knowledge of other subfields of statistics that depends on the concept of probability includes stochastic process, statistical inference, time series, regression analysis, multivariate analysis and other areas of statistical modeling. In addition to all these the students learn how to quantify the chance of an outcome in the midst of uncertainty.

Unit I: Measure Theory

- Classes of Sets, Fields, sigma-fields, minimal sigma-field, Borel field
- Sequence of sets, limits of a sequence of sets
- Measure, Probability measure, Integration with respect to measure. (9 lectures)

Unit II: Probability: Basic Concepts, Conditional Probability

- Markov's, Holder's, Minkowski's and Jenson's Inequalities
- Probability space, Basic terminologies and theorems on probability, theorems on compound probability
- Independence of events, conditional probability
- Bayes' Theorem and its applications, Theorem of total probability. (9 lectures)

Unit III: Random Variables and Probability Functions

- Random Variable, Probability distribution (discrete and continuous), Distribution function
- Bi-dimensional and Multi-dimensional random variables
- Marginal and conditional distributions, Mathematical expectations. (9 lectures)

Unit IV: Generating Functions, Characteristic Functions

- Generating functions, probability generating function, moment generating function characteristic functions
- Uniqueness theorem, inversion theorem
- Joint characteristic function. (9 lectures)

Unit V: The Laws of Large Numbers, Inequalities, The Central limit Theorem

- Law of large numbers, Chebychev's and Khinchin's weak law of large numbers, Strong law of large numbers,
- Kolmogorov's theorem, Central limit theorem, De-Moivre's Laplace central limit theorem, Liapounov's central limit theorem
- Statement of Lindeberg- Feller's central limit theorem. (9 lectures)

References

1. Ash, Robert B. (2000). Probability and Measure Theory, Second Edition, Academic Press, New York.
2. Feller, W. (1985). Introduction to Probability Theory and its Applications, Wiley Eastern, New Delhi
3. Bhatt, B.R. (1999). Modern Probability Theory, 3rd Edition, New Age International Publishers.
4. A. K. Md. Ehsanes Saleh and Vijay K. Rohatgi (2010) An Introduction to Probability and Statistics, Wiley India Pvt. Ltd.
5. Das K. K. and Bhattacharjee D. (2008). An Introduction to Probability Theory, Asian Books, New Delhi.
6. Capinski, M. and Zastawniah (2001). Probability through problems, Springer.

First Semester
Paper 104
Distribution Theory

Course Outcome: The course shall expose the students to different aspects of probability distributions both discrete and continuous, their properties and uses. Some theories and problems related to common families of distributions like the exponential family of distributions, Pitman family of distributions etc. shall also be covered. Sampling distributions and Bivariate distributions are also taught to the students in this course. Advanced topics of distribution theory like mixture distributions, truncated distributions, compound distributions and non-central distributions figure in this course.

Learning Outcome: After successful completion of this course, the students will be able to understand the genesis and basic concepts of various univariate and bivariate discrete and continuous distributions, sampling distributions, compound distribution, zero-modified distributions, extreme value distributions, and distribution order statistics. Apply the univariate and bivariate distributions to model data from real life. Use order statistics in product failure, droughts, floods and other extreme occurrences. Finally, students will use these concepts in other courses of statistics- as distribution theory is the basis of advanced learning in statistics. Almost every topic of statistics need the concepts of distribution theory.

Unit 1: Univariate discrete distributions

- Derivation, properties and applications of binomial, multinomial, Poisson, negative binomial, geometric, hyper geometric distribution, Power series distribution
- Exponential families of distributions, Pitman family of distributions.(9 lectures)

Unit II: Univariate continuous distributions

- Derivation, properties and applications of normal, exponential, gamma, beta, Weibull, Cauchy and lognormal distributions. (9 lectures)

Unit III: Bivariate distributions and Sampling Distributions

- Derivation, properties and applications of bivariate normal distribution
- Derivation, properties and applications of Chi-square, t and F-distributions and their interrelationship. (9 lectures)

Unit IV: Non-central distributions, compound distribution distributions and Truncation

- Derivation, Properties and applications of non-central chi-square, t and F distributions
- Compound distribution- Neyman's Type A distribution, Poly-Eggenberger distribution, Inverse Polya-Eggenberger distribution

- Truncation of basic discrete and continuous distributions with their properties.

(9 lectures)

Unit V: Order Statistics, Mixture distribution and Extreme value distributions

- Distributions of rth order statistics, joint density of two order statistics, some special joint distributions resulting from order statistics
- Distribution of range and other systematic statistics, moments of order statistics, recurrence relations and identities for moments of order statistics from an arbitrary distributions
- Mixture distribution- finite mixture, zero-modified distributions, mixture of binomial distribution, extreme value distributions. (9 lectures)

References

1. Johnson, N.L., Kemp, A.W., and Kotz, S.. (2005): Univariate Discrete Distributions, 5th edition, Wiley Interscience, John Wiley & Sons.
2. Johnson, N.L., Kotz, S., and Balakrishnan, N. (2002): Continuous Univariate Distributions, Vol. 1, John Wiley
3. Johnson, N.L., Kotz, S., and Balakrishnan, N. (2002): Continuous Univariate Distributions, Vol. 2, John Wiley
4. Hogg, R. V., Craig, A., and Mckean, J.W. (2019): Introduction to Mathematical Statistics, 8th edition , Pearson
5. Freund, J.E. (1998): Mathematical Statistics, Prentice Hall of India, New Delhi
6. Biswas, S. (2002): Topics in Statistical Methodology, New Age International Publishers, New Delhi
7. Wilks, S.S. (2007): Mathematical Statistics, Buck Press
8. Das, K.K. and Bhattacharjee, D. (2008): A treatise on Statistical Inference and Distributions, Asian Books, New Delhi

First Semester
Paper 105
Statistical Computing

Course Outcome: The paper aims at enriching the computing power of students using spreadsheets and packages like Microsoft Excel and R. The paper will help the students to understand data entry, process of computation, relevant formula and data handling in R and SPSS. how to go around with the computing part of some of the theoretical aspects using statistical packages.

Learning Outcome: On studying this paper the students shall find themselves capable of handling live data. After completing this paper students are expected to have computational expertise on topics of linear algebra, distribution theory and other basic topics of statistics. *The paper will help the students to understand how to compute the theoretical aspects that they learn in the class. The paper teaches the students data handling and performing Basic Data analysis in Excel and R.*

Unit-I: Fundamentals of Excel Operations

- Workbooks and Worksheets, Excel file operations, Printing contents of worksheet.
- Data Entry, Text Entry, Text Editing, Formatting numbers in cells, Cell Merging.
- Working with formula in Excel, use of $, Coping and pasting formula in Excel, Mathematical and statistical functions in Excel
- Basic Graphical tools in Excel: Bar Chart, Pie Chart, Frequency Polygon, Histogram, error bar plot, Ogive, Line diagram, Scatter diagram, Radar Plot, Doughnut plot, Bubble Plot (9 lectures)

Unit-II: Use of Data Analysis Toolpak

- Loading of Data Analysis Toolpak in Excel.
- Computation of Analysis of Variance, Correlation, Covariance, Descriptive Statistics, Exponential Smoothing
- Random number generation from different distributions, Regression, Sampling, t-test, Z-test
- Optimization using Solver (9 lectures)

Unit-III: Matrix Operations and Distribution Fitting in Excel

- Computation of Determinants of matrix, Inverses of a matrix, Rank of a matrix, Solutions of matrix equations, Characteristic roots and vectors of a matrix
- Distribution Fitting and checking the goodness of fit: Binomial, Poisson, Normal, Log-normal, and Beta distributions (9 lectures)

Unit-IV: R Programming Basics

- Starting and exiting R, getting help, setting, listing and deleting variables, creating vector, defining variables, creating sequences, performing vector arithmetic.
- Entering data from key-board, reading tabular data files, reading from and writing to CSV-files, reading data from HTML tables, Initializing a matrix in R, performing Matrix operations. (9 lectures)

Unit-V: Statistical Computing in R

- Statistical Graphics in R: Basic Graphs- Bar, Pie, line, histograms, Box-plot, Scatter plot, QQ Plot.
- Descriptive statistics involving one and two variables and common tests of significance including one-way and two-way ANOVA (9 lectures)

References

1. Gardener M. (2010) Beginning R : The Statistical Programming Language, Wiley India Pvt. Ltd., New Delhi.
2. Teetor, Paul (2011) R Cookbook, O'Really.
3. Bhattacharjee, D. (2010). Practical Statistics using Microsoft Excel, Asian Books, New Delhi.
4. Levine, D. M., Stephan, D., Krehbiel, T.C. and Berenson, M.L. (2006). Statistics for Managers Using Microsoft Excel, Prentice Hall of India Pvt. Ltd., New Delhi.
5. Albright, S.C., Winston, W. L. and Zappe C. J. (2009) Decision Making Using Microsoft Excel, Cengage Learning, New Delhi.
6. Braun, W. J. and Murdoch, D. J. (2007) A First Course on Statistical Programming with R, Cambridge University Press.
7. Quirk, Thomas J., Rhiney, Eric. (2020) Excel 2019 for Advertising Statistics. A Guide to Solving Practical Problems, Springer.

Second Semester
Paper 201
Linear Models and Regression

Course Outcome: The course is designed to teach the students about the theory of fitting simple linear regression and the properties of the estimated parameters of a regression model. The concept of simple linear regression is then extended to curvilinear as well as to multiple regression. Other aspects of regression like residual analysis, autocorrelation, multicollinearity and the issues arising out of them and their solutions shall also be discussed. Overview of ridge regression, logistic regression and generalized linear models are also provided.

Learning Outcome: After successful completion of this course student will be able to understand the concepts of linear models and regression including simple linear regression, multiple regression, inverse regression, non-linear regression, polynomial regression, logistic regression, non-linear growth models. Students will apply the maximum likelihood estimation for estimating parameters of these models and testing of hypothesis of parameters or functions of parameters. Students are expected to choose an appropriate linear or non-linear models in a given research setting and interpret the model and report the findings scientifically. Finally, students are expected to suggest appropriate regression models for given datasets to predict the behaviour of complex systems or analyze experimental, financial and biological data.

Unit 1: Simple linear regression

- Regression lines and properties, estimation and tests of hypothesis associated with the parameters of regression
- Properties of least square regression
- Confidence intervals and bands for slope and intercept
- Evaluating the goodness of fit, residual analysis, effect of outliers, transformation of variables. Interclass correlation and correlation ratio. (9 lectures)

Unit II: Multiple linear regression

- Estimation and tests of hypotheses associated with parameters of multiple regression
- Properties of least squares, confidence intervals for mean, regression coefficients and prediction in multiple regression, co -linearity in multiple regression
- Analysis of regression residuals, check for normality of the error terms in multiple regression
- Inverse regression, two-phase linear regression, inclusion of qualitative variable as repressors, multiple and partial correlations. (9 lectures)

Unit III: Auto-correlation, Multicollinearity and ridge regression

- Problem of correlated errors
- Detection and removal of auto-correlation, variance-covariance of least square estimators, estimation of error variance with auto-correlation
- Detection and correction of multicollinearity, ridge regression. (9 lectures)

Unit IV: Non-linear regression

- Fit of polynomial regression in one variable and several variables
- Use of orthogonal polynomials
- Gompertz and logistic non-linear growth models. (9 lectures)

Unit V: Logistic Regression and generalized Linear models

- Logit transformation, maximum likelihood estimation
- Tests of hypothesis: Wald test, likelihood ratio test, score test, test for overall regression
- Multiple logistic regression-Forward and Backward methods, Interpretation of parameters. Basic concept of generalized linear models. (9 lectures)

References

1. Rencher, A.C, and Schaalje, G. B. (2008): Linear models in Statistics, 2nd edition, Wiley Interscience.
2. Montgomery, D.C., Peck, E.A and Vining, G.G. (2015): Introduction to Linear Regression Analysis, John wiley & sons, New York.
3. Seber, G.A.F and Wild, C.J. (2005) Nonlinear Regression, John Wiley.
4. Seber, G.A.F and Lee, A. J. (2003): Linear Regression Analysis, John Wiley.
5. Chatterjee, S and Hadi, A.S. (2013): Regression Analysis by Example, 5th edition, John Wiley & Sons, New York.
6. Draper, N. R and Smith, H. (1998): Applied Regression Analysis, 3rd edition, John Wiley & Sons, New York.
7. Rao, C.R., Toutenburg, H., Shalabh and heumann, C. (2008): Linear models and Generalizations- Least squares and Alternatives, Springer.
8. Monahan, J.F. (2008): A Primer on Linear Models, CRC Press.
9. Khuri, A. I. (2010): Linear Model Methodology, CRC Press.

Second Semester
Paper 202
Survey Sampling

Course Outcome: Students shall understand the basic concepts of various probability sampling including simple random sampling, stratified random sampling, probability proportional to size sampling, cluster sampling, multi-stage sampling, systematic sampling and non-probability sampling including distance sampling, snowball sampling, network sampling and importance sampling. Students are taught how to estimate population mean and population total using ratio and regression methods of estimation. Students are exposed to sampling and non-sampling errors, modeling observational error, and randomized response technique to get information for sensitive issues.

Learning Outcome: After studying this paper students can be equipped to perform different types of sampling procedure in real life situation. Students shall acquire the theoretical as well as practical knowledge of field study to analyze the data, interpret the results and draw valuable conclusions coming from various aspects.

Unit I: Basic Sampling Techniques

- Basic ideas and distinctive features of sampling techniques.
- Review of important results in simple random sampling, stratified random sampling, estimation with different type of allocation of strata. (9 lectures)

Unit II: Probability Proportional to Size Sampling

- Sampling with varying probabilities (unequal probability sampling): pps with replacement/without replacement methods [including Lahiri's scheme]
- Hurwitz and Des Raj estimators for a general sample size
- Murthy's estimator. (9 lectures)

Unit III: Ratio and Regression Method of Estimation

- Ratio method of estimation (Hartley Ross and Jacknife estimators)
- Regression method of estimation including its optimum property, combined ratio estimator. (9 lectures)

Unit IV: Various Sampling Techniques

- Double (Two-phase) sampling with special reference to the selection with unequal probabilities
- Cluster sampling with clusters of equal and unequal sizes,
- Two-stage sampling (with varying sizes of first-stage units
- Multi-stage sampling, systematic sampling and its application. (9 lectures)

Unit V: Errors in Survey, Randomized Response Techniques

- Errors in survey, Non sampling errors
- Randomized response technique (Wamer's model and Simmons model)
- Basics of distance sampling, snowball sampling and network sampling. (9 lectures)

References

1. Cochran, W.G. (1997). Sampling Techniques, Wiley Eastern, New Delhi.
2. Mukhopadhyay, P.(1998). Theory and Methods of Survey Sampling, Prentice Hall of India, New Delhi.
3. Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). Sampling Theory of Surveys with Applications, Iowa State University Press and IARS.
4. Murthy, M.N. (1977). Sampling Theory and Methods, Statistical Publishing Society, Kolkata.
5. Chaudhury, A. and Mukherjee, R. (1988). Randomized Response: Theory and Techniques, Marcel Decker, New York.

Second Semester
Paper 203
Fundamentals of Data Collection and Analysis
(CBCS Paper open to PG students of other departments)

Course Outcome: This is a CBCS open paper for PG students of other departments. This paper deals with fundamentals of data collection and analysis which consists of Measures of Central Tendency, Measures of Dispersion, Primary and Secondary data, Scaling Techniques, Correlation, Measures of disease frequency, Tests of Significance, Data Management, Data Coding, Data Cleaning, Data Processing etc.

Learning Outcome: From this paper students shall acquire both theoretical and practical knowledge of basic statistics. With successful completion of this course the students can handle the statistical issues like analysing, presenting and interpretation of real life data that they might be encountering in their core program.

Unit I: Measures of Central Tendency and Dispersion

- Measures of Central Tendency (including windsorized mean and trimmed mean)
- Measures of Dispersion (including relative measures)
- Outliers, normalization and standardization of data
- Weighted (Iyengar-Sudarshan method) and unweighted composite index.

(9 lectures)

Unit II: Sampling and Scaling

- Primary data, method of data collection, secondary data and its sources
- Random and non-random sampling techniques and their types
- Determination of sample size, scaling techniques- Common comparative and non-comparative scales. (9 lectures)

Unit III: Probability and Statistical Tools involving Two Variables

- Correlation (Karl Pearson's, Spearman's), scatter plot
- Lines of regression, residual analysis, r^2 for model fit, outlier detection
- Association of attributes, Coefficient of association, Coefficient of colligation. Probability (Basic laws)
- Simple problems of Binomial and Poisson distribution. (9 lectures)

Unit IV: Fundamentals of Bio-Statistics and Normal Distribution

- Measures of disease frequency, prevalence and incidence, crude, specific and standardized rates, risk ratio, odds ratio, confounding, Mantel-Haenszel odds ratio, infant and maternal mortality rates.
- The normal probability distribution and computation of probabilities using the standard normal table. (9 lectures)

Unit V: Data Handling and Hypothesis Testing

- Data management, data coding, data cleaning, data processing, rectangular data file, missing data analysis
- Tests of significance- Z-test, t-test, F-test and Chi-square test. (9 lectures)

References

1. Madsen B. (2008) Statistics for Non-statistician, Springer
2. Urdan, TC. (2005). Statistics in Plain English, Lawrence Erlbaum Associates Publishers
3. Gun AM., Gupta MK. And Dasgupta B. (2001) Fundamentals of Statistics, Vol. 1, World Press.
4. Boslaugh S and Watters PA (2008) Statistics in a Nutshell, O'Really
5. Naresh K. Malhotra and Satyabhusan Dash (2010) Marketing Research: An Applied Orientation, Fifth Edition, Pearson Education.
6. Rumsey D. (2007) Intermediate Statistics for Dummies, Wiley Publishing, Inc.
7. Iyengar, N. S. and Sudarshan, P., 1982, A method of classifying regions from multivariate data. *Economic and Political Weekly*. Dec. 18: 2048-2052.
8. James G, Witten D, Hastie, T and Tibshirani, R (2017) An Introduction to Statistical Learning, Springer, Singapore.

Second Semester
Paper 204
Statistical Inference-I

Course Outcomes: Students shall be exposed to the knowledge of properties of estimators, methods of estimation of parameters, methods of obtaining minimum variance unbiased estimators. Students shall learn about point estimation, hypothesis testing and interval estimation under a variety of discrete and continuous probability models.

Learning Outcome: After successful completion of this course the students will be able to understand the basic concepts of statistical inference including point estimation, interval estimation and testing of hypothesis and apply these techniques to deal with real life problems. Further, students can evaluate the properties of these estimators and tests for both finite sample sizes and asymptotically as the sample size tends to infinity. Finally students can identify and select optimal estimators and tests for a given research or applied problems.

Unit 1: Point Estimation

- Properties of estimators-Unbiasedness, consistency, efficiency, sufficiency
- Neyman factorization criterion, Completeness, CAN and BAN estimators
- Cramer-Huzurbazar theorem, exponential families of distributions and its properties, Non-regular families admitting complete sufficient statistics

(9 lectures)

Unit II: Minimum Variance Unbiased Estimators (MVUE)

- Rao-Cramer's inequality, Chapman-Robin's inequality
- Bhattacharya's Bound, Mean square error
- Best linear unbiased estimator (BLUE) , Rao-Blackwell theorem, Lehmann-Scheffe theorem (9 lectures)

Unit III: Methods of Estimation

- Method of maximum likelihood, asymptotic properties of MLE, MLE in censored and truncated distributions
- Newton-Raphson method and method of scoring, Fisher information for one and several parameter models
- Method of moments, method of minimum chi-squares, modified minimum chi-square, Pitman method of estimation for location and scale (9 lectures)

Unit IV: Tests of Hypothesis

- Neyman-Pearson lemma, most powerful (MP) and uniformly most powerful (UMP) tests, UMP tests for simple null hypothesis against one-sided alternatives and for one sided null against one-sided alternatives in one parameter exponential family
- extension of these results to Pitman family when only upper or lower ends depends on the parameters. (9 lectures)

Unit V: Interval Estimation

- Pivotal quantity method and general method of constructing confidence interval, large sample confidence interval, construction of shortest expected length confidence interval
- Uniformly most accurate one-sided confidence interval and its relation to UMP tests for one-sided null against one-sided alternative hypothesis. (9 lectures)

References

1. Hogg, R. V., Craig, A., and Mckean, J.W. (2019): Introduction to Mathematical Statistics, 8th edition , Pearson
2. Kale, B. K. (1999): A first course on Parametric Inference, Narosa Publishing House
3. Bartoszynski, R and Bugaj, M.N. (2007): Probability and Statistical Inference, John Wiley & Sons
4. Lehmann, E. L. (1986): Theory of Point Estimation, John Wiley & Sons
5. Lehmann, E.L. (1986): Testing Statistical Hypothesis, John Wiley & Sons
6. Dudewicz, E. J and Mishra, S.N. (1988): Modern mathematical Statistics, John Wiley & Sons
7. Rao, C. R. (1973): Linear Statistical Inference and Its Applications, 2nd edition, Wiley Eastern Limited, New Delhi
8. Wilks, S.S. (2007): Mathematical Statistics, Buck Press
9. Rohtagi, V.K and Saleh, A. K, Md, E. (2005): An Introduction to Probability and Statistics, 2nd Edition, John Wiley & Sons
10. Zacks, S. (1971): Theory of Statistical Inference, , John Wiley & Sons
11. Mishra, A. (2020): Theory of Statistical Estimation, Notion Press, Chennai
12. Mishra, A. (2020): Theory of Statistical Hypothesis Testing, Notion Press, Chennai.
13. Das, K.K. and Bhattacharjee, D. (2008): A Treatise on Statistical Inference and Distributions, Asian Books, New Delhi.

Second Semester
Paper 205
Statistical Computing

Course Outcome: The main purpose of this paper is to introduce the students of object oriented programming concepts in C language and also provide them with an overview in Python Programming. C is a powerful general-purpose programming language, which can be used to develop software like operating systems, databases, compilers and so on.

Learning Outcome: The learning outcome of this paper is to enrich the computing power of students and provide them some exposure to the programming skills in C. Students will be able to develop their programming logic which will help them to make their own programs, applications in C language. The students can learn some basics of C language and can switch over to any other programming language in future. With the sudden development of Data Science as a new discipline, where Python has become very popular students are given some exposure to Python programming. This shall help them in case they take up Data Science as a career in future.

Unit-I: Fundamentals of C-Programming

- Programming, High Level Languages, Operating Systems, Language interpreters
- Writing small programs in C and running them, taking values from variables
- Data types in C and constants (9 lectures)

Unit-II: Operators in C and Looping

- Arithmetic and Relational operators in C
- Combining operators with assignments
- For statement, nested For-loops, While-do loops, use of Continue and Break statements

(9 lectures)

Unit-III: Decision Making and Working with Arrays

- If-statement, If-Else and Else-if Construct, switch statement and conditional operators
- Library functions in C
- Defining an Array, Initializing an Array, working with two-dimensional array, Arrays of Characters (9 lectures)

Unit-IV: Functions and Pointers

- Defining a function, Argument and Local Variables, Returning Function Results
- Defining Pointer Variables, Use of Pointer in expressions, Working with pointers
- Programs involving character strings (9 lectures)

Unit V: Python Programming

- Installing Python for Windows, Writing simple programs in Python, Defining variables, accepting user input.
- Understanding arithmetic and logical operations in Python, Data types in Python.
- Writing and Manipulating lists, Branching with if, Looping while-true, looping over times, Breaking out of loops (9 lectures)

List of Practicals

Using the knowledge of C/Python students are expected to program the following:

1. Factorial of a positive Integer
2. Ordering of a given set of observations
3. Finding maximum and minimum of a given set of observations.
4. Mean, variance and quantiles for ungrouped and grouped data.
5. Correlation coefficients for ungrouped data.
6. Fitting of exponential curve and straight line to the given data.
7. Fitting of Binomial and Poisson distributions.
8. Drawing of random samples from Binomial, Poisson, Normal and Gamma distributions.
9. Fitting of standard distributions and tests for goodness of fit.
10. Tests of significance based on Chi-square, t and F statistics and related confidence intervals.
11. Large sample tests.
12. Construction of difference table, Forward, Backward and Central difference interpolation formulae. Divided difference table. Newton's divided difference and Lagrange's interpolation formula.
13. Numerical Integration: Trapezoidal Rule, Simpson's 1/3rd rule and Weddle' formula for numerical integration.
14. Operations of Matrix algebra and computation of the inverse of a matrix.

References

1. Kochan SG. (2005) Programming in C (Third Edition), Sams Publishing, Indiana, USA.
2. Kanitkar Y. (2013) Let us C, Thirteenth Edition, BPB Publications, New Delhi
3. Balagurusamy E. (2012) Programming in ANSI C, Tata McGraw Hill.
4. Gottfried B., (1996) Schaum's Outline of Programming with C, McGraw Hill.
5. McGrath M. (2013) Python in Easy Steps, Mc Graw Hill Education (Indian edition).

6. Guttag, J V (2013) Introduction to Computation and Programming Using Python, Revised and Expanded Edition, MIT Press, Cambridge. (Marketed in India by PHI)
7. Kernighan BW and Ritchie D. (2015) The C Programming Language, Second Edition, Pearson.

Third Semester
Paper 301
Design and Analysis of Experiments

Course Outcome: Understand the meaning of randomization, replication, local control and contrast and make use of analysis of variance-one way, two-way with equal and unequal number of observations per cell along with analysis of covariance. Identify the common and important types of experimental designs with respective advantages and disadvantages in terms of power, cost and time. Apply the basics of statistical inference including estimation, testing of hypothesis and confidence interval in the designs. Understand and use factorial experiments, fractional factorial experiment and confounding in experiments. Differentiate between connectedness, orthogonality and balanced designs.

Learning Outcome: Finally, students are expected to choose an appropriate design in a given research setting and interpret the model and report the findings scientifically. On studying this paper the students are expected to suggest appropriate experimental designs for agricultural and livestock experiments so as to minimize the experimental error.

Unit I: Linear Model, ANOVA, ANOCOVA

- Review of linear estimation
- Linear model, Fixed effect, Random effect and Mixed effect model, ANOVA: one way, two way (with m>1) observations per cell
- Analysis of Co-variance. (9 lectures)

Unit II: Application of Analysis of Variance

- Orthogonal Contrasts
- Application of ANOVA in the study of relationship between two variables
- Linearity of regression, polynomial regression, homogeneity of group regression coefficients, multiple linear regression model
- Non parametric ANOVA. (9 lectures)

Unit III: Designs of Experiment

- Designs of Experiment: Basic Principles, basic designs and their analysis, relative efficiency
- Missing plot analysis, Graeco-Latin square design, Quasi-Latin Square design.

(9 lectures)

Unit IV: Factorial Experiments

- Factorial experiments: 2^n-factorial experiment (n=2 & 3) factorial experiments in randomized blocks.
- Confounding in factorial experiments and 3^n (n=2, 3) factorial experiments, Double confounding in 2^n experiment
- Symmetrical factorial experiments (s^m, where s is a prime or a prime power).

(9 lectures)

Unit V: Incomplete Block Designs

- Incomplete Block Designs, Concepts and Connectedness
- Orthogonality and Balance. Balance Incomplete Block Design (BIBD), Analysis with Intrablock and Inter block information
- Resolvable and affine resolvable designs. Split-plot and Strip-plot Design.

(9 lectures)

References

1. Cochran, W.G. and Cox, G.M. (1959). Exponential Designs, Asia Publishing House, Singapore.
2. Das, M.N. and Giri, N.C. (1986). Design and Analysis of Experiments, Wiley Eastern Limited.
3. Joshi, D.D. (1987). Linear Estimation and Design of Experiments, Wiley Eastern, New Delhi.
4. Montgomery, D.C. (2005). Design and Analysis of Experiments, Sixth Edition, John Wiley and Sons.
5. Goon, A. M., Gupta, M. K. and Das Gupta, B (1985): An Outline of Statistical Theory, Vol.-2, the World Press Ltd., Calcutta,
6. Dey, Alok (1986): Theory of Block Designs, John Wiley and Sons.

Third Semester
Paper 302
302: Statistical Inference-II

Course Outcomes: This is an advanced paper in Statistical Inference. On getting the basics of Statistical inference in the previous semester students shall go ahead to learn topics like Likelihood Ratio Test specific to composite hypothesis testing. Sequential Probability Ratio Test (SPRT) where the sample size of the tests are allowed to vary, Bayesian Inference, Decision theory and Non-parametric tests.

Learning Outcomes: After successful completion of this course student will be able to apply various parametric, nonparametric, sequential and Bayesian testing procedures to deal with real life problems. Students are expected to identify appropriate statistical inference techniques from a given research or applied problems and perform correct statistical analysis using different inferential techniques and tests.

Unit 1: The Likelihood Ratio Test (LRT) and Monotone Likelihood Ratio Test

- One-sided and two-tailed and two-tailed likelihood ratio tests for mean and variance of normal population, consistency and unbiasedness of LRT
- Asymptotic distribution of LRT and applications
- Monotone Likelihood Ratio Test and applications, Similar regions and similar tests.

(9 lectures)

Unit II: Overview of Bayesian Inference

- Concept of prior and posterior distributions, Types of prior and relevant information
- Bayes' risk and Bayes rules
- Bayesian estimation of parameters and parametric functions under various loss functions (9 lectures)

Unit III: Bayesian testing of Hypothesis

- Lindley's methods, Tests of point null hypothesis
- Jeffrey's test for normal mean
- Bayesian interval estimation-credible intervals and their applications, comparison with classical confidence intervals, Bayesian prediction (9 lectures)

Unit IV: Sequential tests

- Wald's Sequential Probability Ratio Test (SPRT), Properties of SPRT, Efficiency of SPRT

- Fundamental identity, Operating characteristic (OC) function and average sample number(ASN) function, ultimate termination of SPRT, Application of SPRT in decision-making. (9 lectures)

Unit V: Non-parametric methods

- Estimation and confidence interval, U-statistics and their asymptotic properties
- Distribution free confidence intervals for population quantiles and distribution-free tolerance intervals
- Non-parametric tests –single sample and two sample tests including sign test, run tests, Mann-Whitney U tests, Rank order statistics, Wilcoxon signed rank test,, tests of goodness of fit-Chi-square and Kolmogorov-Smirnov tests. (9 lectures)

References

1. Ferguson, T. S. (1967): Mathematical Statistics, Academic Press
2. Gibbons, J.D and Chakraborti, S. (1992): Nonparametric Statistical Inference, Marcel Dekker, New York
3. Kale, B. K. (1999): A first course on Parametric Inference, Narosa Publishing House
4. Lehmann, E.L. (1986): Testing Statistical Hypothesis, John Wiley & Sons
5. Dudewicz, E. J and Mishra, S.N. (1988): Modern mathematical Statistics, John Wiley & Sons
6. Rao, C. R. (1973): Linear Statistical Inference and Its Applications, 2nd edition, Wiley Eastern Limited, New Delhi
7. Rohtagi, V.K and Saleh, A. K, Md, E. (2005): An Introduction to Probability and Statistics, 2nd Edition, John Wiley & Sons
8. Zacks, S. (1971): Theory of Statistical Inference, , John Wiley & Sons
9. Mishra, A. (2020): Theory of Statistical Hypothesis Testing, Notion Press, Chennai.
10. Ghosh, J. K., Delampady, M and Samanta, T. (2006): An Introduction to Bayesian Analysis-Theory and Methods, Springer International Edition
11. Das, K.K. and Bhattacharjee, D. (2008): A treatise on Statistical Inference and Distributions, Asian Books, New Delhi.

Third Semester
Paper 303
STOCHASTIC PROCESS

Course Outcome: The paper introduces the concept of stochastic process, Markov chain, and transition probability matrix along with classification of stochastic process. On studying this paper students can identify the states and stationary distribution of Markov Chain along with distribution of Markov chain at a given time. The paper enables in understanding the concepts and applications of random walk, Gambler's ruin problem, Brownian motion, Wiener Process, Branching process, Stationary process and renewal process.

Learning Outcome: Finally, students are expected to choose appropriate stochastic process model(s) for a given research in applied problem and apply the theory to model real life phenomena and solve several problems concerning random behaviour in different fields of applied science. Thus, this paper shall also provide a basis of further research in model bulding in a probabilistic environment.

Unit I: Introduction to Stochastic Process

- Elements of Stochastic Process, Markov Chains: Definitions, Examples, transition probability matrix, Classification of States of a Markov Chains, Higher transition probabilities in Markov Classification of States and Chains.
- Stability of a Markov System. Limiting behaviour: Irreducible and ergodic chain. Absorption probabilities. (9 lectures)

Unit II: Markov Processes with discrete and continuous state space, Martingales

- Markov Processes with discrete state space: Poisson process: Definitions and Examples, Basic Properties of Poisson Processes
- Some Generalization of the Poisson Process. Simple Birth and Death Processes, General Birth and Death Process, Polya Process.
- Martingales: Definitions and Examples, Supermartingales and Submartingales, Martingale convergence theorems. (9 lectures)

Unit III: Random Walk and Brownian Motion

- Random walk: Introduction, Gambler's ruin, correlated random walk.
- Brownian motion: Definition and Example, Fundamental path properties of Brownian motion, Wiener process. (9 lectures)

Unit IV: Branching Process and Stationary Process

- Branching Process: Discrete time branching processes, Generating Function Relations for Branching Processes
- Extinction probabilities, Examples
- Stationary Processes: Definitions and Examples, Ergodic Theory and Stationary Processes. (9 lectures)

Unit V: Renewal Processes

- Renewal Processes: Definition and Related Concept, Examples of Renewal Processes
- Stopping time: Wald's Equation, Renewal Equations and the Elementary Renewal Theorem, The Renewal Theorem, Delayed and Equilibrium Renewal Processes (9 lectures)

References

1. Bailey, Norman, T.J. (1964). The Elements of Stochastic Processes, John Wiley and Sons.
2. Bartlett, M.S. (1966). An Introduction to Stochastic Processes, Cambridge University Press.
3. Bhat, B.R. (2000). Stochastic Models- Analysis and Applications, New Age International Publishers.
4. Feller, William (1968). An Introduction to Probability Theory and its Applications, Vol. 1 (Third Ed.) John Wiley.
5. Karlin, S. and Taylor, H.M. (1975). A First Course in Stochastic Processes, (Second Ed.), Academic Press.
6. Medhi, J. (1994). Stochastic Processes, Second Edition, Wiley Eastern Ltd.
7. Ross, S.M. (1983). Stochastic Processes. John Wiley and Sons.
8. Mishura, Y. and Shevchenko, G. (2017). Theory and Statistical Applications of Stochastic Processes, Wiley.
9. Jones, P. W. and Smith, P. (2018). Stochastic Processes: An Introduction, Third Edition, CRC Press.

Third Semester
Paper 304
Applied Statistics I

Course Outcome: Understand the concept of national development in terms of various indices of development including human development index and national income. Determine the levels of mortality and fertility and construct life tables. Projection of population growth and measurement of income inequality using different methods are expected. Understand role, function and activities of various statistical organizations for collecting official statistics. Finally, students are expected to know the collection, analysis, forecasting and interpretation of results of data relating to agricultural, industry, educational and other social statistics. Application of statistical tools in sports data mining.

Learning Outcome: This paper exposes the students to different fields of applied statistics and let them know how the probabilistic and computational models can be applied to some real life domains.

Unit I: Statistics for Nation Building

- Statistics for National Development- Economic development: Growth in per capita income and distributive justice.
- Indices of development, Human Development Index, Composite Index Development, Normalization, Weighting and Aggregation. (9 lectures)

Unit II: Demography I

- Measures of mortality and fertility
- Life table: complete and abridged, King's method, Greville's method, Reed and Merrell method, Chiang's method, uses of life table, graduation of mortality rates. (9 lectures)

Unit III: Demography II

- Population growth in developing and developed countries.
- Population projection using Component method, Leslie marix. Labour force projection.
- Measuring inequality in incomes, Gini coefficient, Theil's measure. Poverty measurement – different issues, measures of incidence and intensity, combined measures e.q., indices due to Kakwani, Sen etc. (9 lectures)

Unit IV: Indian Official Statistics

- Official Statistics- Indian and International statistical systems: Role, function and activities of Central and State statistical organizations.

- Organization of large scale sample surveys.
- Role of National Sample Survey Organization, Sample Registration System (SRS), National Family Health Surveys, cost of living, inflation, educational and other social statistics. (9 lectures)

Unit V: Official Statistics and Sports Analytics

- System of collection of agricultural statistics, Crop forecasting and estimation, productivity, impact of irrigation projects. Statistics related to industries, foreign trade etc.
- Data Mining in Sports with Special reference to Football and Cricket, Performance measurements in Cricket. (9 lectures)

References

1. Basic Statistics Relating to the Indian Economy (CSO) 1990.
2. Guide to Official Statistics (CSO) 1999.
3. Keyfitz, N. (1977). Applied Mathematical Demography, Springer Verlag.
4. Benjamin, B. (1969). Demographic Analysis, George, Allen and Unwin
5. Panse, V.G., Estimation of Crop Yields (FAO).
6. Principles and accommodation of National Population Censuses, UNESCO
7. Statistical System in India (CSO) 1995.
8. Saikia H., Bhattacharjee D. and Mukherjee D. (2020) Cricket Performance Management: Mathematical formulation and Analytics, Springer, Singapore
9. Sen, A. (1997). Poverty and Inequality.
10. UNESCO. Principles for Vital Statistics Systems. Series M -12.
11. Gun AM., Gupta MK. and Dasgupta B. (2001) Fundamentals of Statistics, Vol. 2, World Press.
12. Michela Nardo, Michaela Saisana, Andrea Saltelli & Stefano Tarantola (2005) Tools for Composite Indicators Building, Joint Research Centre, European Commission, Italy.
13. Handbook on Constructing Composite Indicators: Methodology and User Guide, OECD, JRC European Commission.
14. Miller TW (2016) Sports Analytics and Data Science, Pearson Education, Inc., New Jersey.
15. Martin L. (2016) Sports Performance Measurement and Analytics, Pearson Education, Inc., New Jersey.

Third Semester
Paper 305
Statistical Computation in SPSS (Practical Paper)

Course Outcome: The objective of this paper is to introduce the students to Statistical Package for Social Sciences (SPSS) one of the most popular Statistical Packages. The paper shall provide the students' hands on training on the application and the different computational facilities provided in the package. Along with common tools of data analysis like regression, Descriptive Statistics, Tests of Significance both parametric and non-parametric, Graphical tools, some advanced tools like multivariate data analysis, time series analysis, survival analysis and data manipulation tools shall also be taught.

Learning Outcome: Students on studying this paper are expected to get well acquainted with the statistical software and to apply it for a given research problem. Most of the statistical techniques learnt by the students in the previous semesters and the current semester can be computed using SPSS and the students shall be taught how to interpret the findings of the analysis derived in SPSS in a scientific way. This is a course after studying which, students can perform different statistical computations using SPSS and can think of a career on data analysis and even get self-employed as a data analyst.

Unit-I: The SPSS Environment

- Getting Started, the SPSS data Editor, Data View and Variable View
- Arranging the variable view to input primary data, Data Entry, Text Entry
- Coding and Decoding of data, Recoding into same and different variable.
- Basic Graphical tools in SPSS: Histograms, Box Plot, Simple and Clustered Bar Chart, Line Charts, Scatter plot, Probability plots and Q-Q Plot. (9 lectures)

Unit-II: Descriptive Statistics

- Computation of frequencies, cross-tabulation, measures of central tendency, dispersion, skewness and kurtosis, analysis of grouped frequency distributions
- Computation of correlation coefficient (Pearson, Spearman, Kendall, Biserial and Point-biserial correlation), correlation matrix, partial correlation
- Simple Linear Regression, Non-linear regression involving two variables, Multiple Regression, Checking the assumptions of regression, use of dummy variables in regression. (9 lectures)

Unit-III: Parametric and Non-parametric Tests in SPSS

- Performing basic parametric tests in SPSS- Z-test for proportions (one sample and two sample), t-test (single mean, independent samples, paired sample), Levine test, Chi square test for independence of attributes and Goodness-of-fit

- ANOVA-one way, two-way (with single and multiple number of observations per cell), Latin Square Design experiment.
- Basic non-parametric tests: Wilcoxon Signed Rank Test, Mann-Whitney U Test, Wilcoxon Matched Pair Signed Rank Test, Median test, Run test (one sample, two samples), Wald-Wolfowitz Run Test, Kolmogrov-Smirnov (one sample and two sample tests), Kruskal-Wallis One-way ANOVA, Binomial test, McNemar's Test, Cochran's Q Test, Kendall's Coefficient of Concordance. (9 lectures)

Unit-IV: Scaling Techniques, Time Series Analysis and Logistic Regression

- Collection and analysis of Likert Scale type data
- Time Series Analysis: Sequence Charts, Forecasting Models, Fitting ARIMA (p,d,q) models for forecasting, choosing appropriate model
- Fitting binary logistic regression model to appropriate data, interpreting odds-ratio. (9 lectures)

Unit-V: Multivariate Analysis

- Factor Analysis, Using Principle Component Analysis for composite index development.
- Multivariate Analysis of Variance (MANOVA), Cluster analysis- *k*-mean clustering and hierarchical clustering
- Discriminant Analysis as a tool for supervised classification (9 lectures)

References

1. Field A. (2013) Discovering Statistics Using SPSS, Sage Publications.
2. Carver RH. and Nash JG. (2012) Doing Data Analysis in SPSS: Version 18.0, Cengage Learning.
3. Landu, S. and Everitt, BS. (2004) A Handbook of Statistical Analysis in SPSS, Chapman and Hall/CRC Press, London.
4. Cleophas TJ and Zwinderman AH (2012) SPSS for Starters and Second Levelers (Second Edition), Springer, Singapore
5. Ho, Robert (2006) Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS, Chapman and Hall/CRC Press, London.
6. Schmidt W (2019) IBM SPSS: Comprehensive Beginners Guide to Learn Statistics Using IBM SPSS from A to Z, Independently Published

Fourth Semester

Paper 401
Multivariate Analysis

Course Outcome: In multivariate analysis students learn how to deal with the data analysis of several variables simultaneously. Necessary theoretical deductions of different multivariate techniques and deduction of multivariate probability distributions are the learning objectives of this paper. The analysis of real life data using multivariate tools like- factor analysis, discriminant analysis, cluster analysis, principal component analysis are also taught using appropriate statistical package. Given the popularity of Machine Learning some tools of classification like Nearest Neighbourhood, Naïve Bayes Classifier and decision trees are included.

Learning Outcome: After studying this paper, students shall be equipped to perform multivariate data analysis on real life data using statistical packages, interpret the results and in addition shall develop necessary theoretical and mathematical understanding of the multivariate processes. The students shall also learn about some machine learning algorithms.

Unit I : Multivariate Normal and Multinomial Distribution

- Multivariate normal distribution its properties and characterization, Random sampling from a multivariate normal distribution. Maximum likelihood estimators of parameters. Distribution of sample mean vector.
- Multinomial distribution and its properties
- Distribution of Quadratic forms.

(9 lectures)

Unit II: Wishart Distribution

- Wishart matrix- its distribution, characteristic function and properties, Wilk's Λ Distribution, density function of Wilk's Λ statistic
- Null and non-null distribution of sample correlation coefficient, Distribution of sample regression coefficients.
- Application in testing and interval estimation. (9 lectures)

Unit III: Inferential Issues of Multivariate Data

- Null distribution of Hotelling's T^2 statistic, application in tests on mean vector for one or more multivariate normal populations and also on equality of the components of a mean vector in a multivariate normal population
- Likelihood ratio Teat for mean vectors, Comparing several mean vectors and variance co-variance matrix. (9 lectures)

Unit IV: Analysis of Covariance Structure

- Classification and discrimination procedures for discrimination between two multivariate normal populations – sample discriminant function, tests associated with discriminant functions
- Principal component, Dimension reduction, Canonical variables and canonical correlation, Factor analysis. Multivariate analysis of variance (MANOVA) of one way classified data. (9 lectures)

Unit V: Machine Learning

- Introduction to Machine Learning, Lazy Learning – Classification using Nearest Neighbourhood
- Probabilistic Learning – Classification using Naïve Bayes Classifier, Classification using Decision Trees and Rules. (9 lectures)

References

1. Anderson, T.W. (2003). An Introduction to Multivariate Statistical Analysis, Third Edition, John Wiley & Sons.
2. Arnold, Steven F. (1981). The Theory of Linear Models and Multivariate Analysis, John Wiley& Sons.
3. Giri, N.C. (1977). Multivariate Statistical Inference, Academic Press.
4. Johnson, R.A. and Wichern, D.W. (2007). Applied Multivariate Statistical Analysis, Sixth Edition, Pearson & Prentice- Hall.
5. Kshirsagar, A.M. (1972). Multivariate Analysis, Marcel Dekker.
6. Bhuyan, K. C. (2005) Multivariate Analysis and its Applications, New Central Book Agency (P) Ltd., Kolkata.
7. Singh, B. M. (2002). Multivariate Statistical Analysis: An Introduction to its Theoritical Aspects, South Asian Publishers Pvt. Ltd., New Delhi
8. Lantz Brett (2014) Machine Learning with R (Indian Edition), Originally published by Packt Publishing Ltd. UK.
9. Tabachnick BG and Fidell LS (2017) Using Multivariate Statistics, Sixth Edition, Pearson

Fourth Semester
Paper 402
APPLIED STATISTICS II

Course Outcome: The paper provides the students exposure to different models used in the field of time series analysis, survival analysis and competitive risk theory. On studying this paper students shall acquire knowledge of both theoretical and computational aspects of time series data as well as data coming from the field of medical science. While time series data is generated from various domains like population studies, business, share markets, sports, public health etc.

Learning Outcome: On studying this paper the students shall acquire the theoretical as well as practical knowledge of analyzing the data coming out from share market, business analytics, sports, medical science, public health and so on.

Unit I: Overview of time series

- Time Series: introduction and Overview, Characteristics of time series, Auto covariance and autocorrelation functions and their properties.
- Exploratory time series analysis, tests for trend and seasonality. Exponential and moving average smoothing, Holt and winters smoothing. (9 lectures)

Unit II: Stationary Processes

- Detailed study of the stationary processes: (i) moving average (MA) (ii) auto regressive (AR) (iii) ARMA and (iv) AR integrated MA (ARIMA) models.
- Box-Jenkin's models. (9 lectures)

Unit III: Forecasting and Modelling time series

- Introduction to forecasting, selecting forecasting models, modelling seasonality,
- Forecasting seasonal series, characterizing and modelling cycles conditional forecasting regression analysis and forecasting
- Modelling and Forecasting Financial Time Series, Introduction to ARCH and GARCH models (9 lectures)

Unit IV: Censoring and Survival Analysis

- Censoring. Hazard and survival function. Estimation of the survival function: Nelson-Aalen methods, Kaplan-Meier's method
- Actuarial Estimator. Methods for comparisons of two or more survival curves.

- Proportional risk, Cox regression, non-parametric and parametric methods for analysis of survival data. (9 lectures)

Unit V: Competitive Risk Theory and Clinical Trials

- Measures of competing risks, Crude, Net and Partially crude probabilities (inter-relation and estimation)
- Dependent and independent risk.
- Clinical trials: General concepts, the design and analysis of clinical trials, use of prognostic factors. (9 lectures)

References

1. Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis - Forecasting and Control, Holden-day, San Francisco.
2. Anderson, T. W. (1971). The Statistical Analysis of Time Series, Wiley, N. Y.
3. Montgemory, D. C. and Johnson, L. A. (1977). Forecasting and Time Series Analysis, McGraw Hill.
4. Cox, D.R. and Oakes, D. (1983). Survival analysis, Chapman and Hall.
5. Robert, H. S. and DavidS. S. (2011). Time Series Analysis and Its Applications with R Examples, Springer, Third Edition.
6. Mills, T. C. (2019). Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting, Academic Press, First Edition.
7. Chatfield, C. and Xing, H. (2019). The Analysis of Time Series: An Introduction with R, Chapman and Hall/CRC.

Fourth Semester
Paper 403
403: Industrial Statistics and Optimization Techniques I

Course Outcome: This course shall teach the students to identify and develop operational research models from verbal description of the real life and apply linear programming, non-linear programming, integer programming, quadratic programming, transportation models, assignment models, Network analysis and deterministic and probabilistic inventory models to solve these problems.

Learning Outcome: Most of the issues discussed in this paper comes out of problems faced in different types of business and industries. Understand the characteristics of different types of decision-making environments and decision making approaches in business, industry and real life, formulate them into mathematical models with constraints in the form of equations and inequalities and select appropriate operational research techniques to solve them. Knowledge of such topics increases the ability of a student to translate problems faced in business/industry to mathematical models.

Unit 1: Linear Programming and Integer Programming

- Convex sets, supporting and separating hyper planes, standard linear programming problem, simplex method
- Artificial variable techniques-Big M method and Two-phase method
- Artificial free simplex method, Duality and Dual simplex method. Integer linear programming-Gomory cut method and Branch and Bound method (9 lectures)

Unit II: Network Analysis

- Network Diagram, Time estimates, minimum spanning tree problem, shortest routes problems, critical path method
- Product Evaluation and Review technique, Time-cost minimization, Resource allocation and scheduling. (9 lectures)

Unit III: Deterministic Inventory Models

- Economic order quantity (EOQ) models with constant rate of demand, EOQ models with shortages and without shortages
- EOQ with finite replenishment rate, production inventory models, EOQ models with quantity discount (9 lectures)

Unit IV: Probabilistic Inventory Models

- Lead time analysis, Reorder level, Buffer stock

- Inventory control systems, Probabilistic Inventory models, Probabilistic order level system (POLS), POLS with instantaneous demand. (9 lectures)

Unit V: Transportation models, Assignment models and non-linear programming

- Initial basic feasible solution using North West corner rule, matrix minima method
- Vogel's approximation method and optimal solution of transportation models, Special cases of transportation models
- Assignment models-Hungarian method and special cases of assignment problems. Non-linear programming-Kuhn tucker condition, Quadratic programming-beals and Wolf method. (9 lectures)

References

1. Hillier, F.S and Lieberman, G.J. (2001): Introduction to Operations Research, 7th edition McGraw Hill Inc
2. Hillier, F.S and Lieberman, G.J. (1995): Introduction to Mathematical Programming, 2nd edition McGraw Hill Inc
3. Taha, H. A. (2016): Operations Research-An Introduction, 10th edition, Prentice Hall of India
4. Kanti Swarup, Gupta, P.K and Singh, M. M. (2019): Operations Research, Sultan Chand &Sons, New Delhi
5. Cottle, R.W and Tappa, M.N. (2017): Linear and Non-linear optimization , Springer
6. Winston, W. L and Goldberg, J. B. (2004): Operations Research-Applications and Algorithm, Thomson Brooks/Cole
7. Nita, H. S, Ravi, M and Hardik, S. (2007): Operations Research, Prentice Hall of India private limited.
8. Hadley, G. (1961): Linear Programming, Addiison-Wesley
9. Murthy, K.G. (1983): Linear Programming, 2nd edition, John Wiley

Fourth Semester
Paper 403 (ii)
Actuarial Statistics I

Course Outcome: This paper exposes the students to different aspects of the basics of actuarial science. The paper explains the concept of insurance and its different terminologies like loss function, premium etc. along with some advanced models of demography concerning insurance industry. The entire concept is explained using probability models that can be designed for aggregate values of population survival.

Learning Outcome: On studying this paper students can get the basic knowledge how the insurance industry works, how the policies are designed and the way in which premiums are fixed.

Unit I: Utility theory and Force of mortality

- Utility theory, insurance and utility theory, models for individual claims and their sums, survival function,
- Force of mortality, common loss distributions, collective risk models for a single period and for an extended period, ruin theory, applications. (9 lectures)

Unit II: Survival models and Life table

- Survival models
- Life table and its relation with survival function, examples, assumptions for fractional ages, some analytical laws of mortality, select and ultimate tables.

(9 lectures)

Unit III: Multiple life functions and Multiple decrement models

- Multiple life functions, joint life and last survivor status, insurance and annuity benefits through multiple life functions evaluation for special mortality laws.
- Multiple decrement models, deterministic and random survivorship groups.

(9 lectures)

Unit IV: Mortality estimation

- Central rates of multiple decrements, net single premiums. Stationary Population.
- Mortality estimation: exposure to risk, approximation for incomplete data. Parametric, tabular and graphical methods, tests of graduation (9 lectures)

Unit V: Population Projections

- Population Projections, distribution of aggregate claims, compound Poisson distribution and its applications. (9 lectures)

References

1. Bowers et al. (1997). *Actuarial Mathematics*, Second Edition. Society of Actuaries.
2. N.L. Bowers, H.U. Gerber, J.C. Hickman, D.A. Jones and C.J. Nesbitt, (1986). 'Actuarial Mathematics', Society of Actuarias, Ithaca, Illiois, U.S.A. Second Edition (1997).
3. Slud, E.V. (2001). Actuarial Mathematics and Life-Table Statistics, University of Maryland, USA.
4. Trowbridge, C. L. (1989). Fundamental Concepts of Actuarial Science, Actuarial Education and Research Fund, USA.

Fourth Semester
Paper 404 (i)
Industrial Statistics and Optimization Techniques II

Course Outcome: The paper deals with reliability, queuing theory, simulation and statistical quality control. On studying this paper the students shall get the theoretical understanding of how to compute probability of survival of machines, models related to production system in industries, reaching conclusions through models for which live data is not available through simulation and so on. The study of queuing theory shall provide the students with the understanding of the theoretical basis of the mathematical models governing queues, waiting time for getting a service, average number of customers in a queue and so on in both real and virtual queues.

Learning Outcome: Most of the issues discussed in this paper comes out of problems faced in different types of business and industries. Understand the characteristics of different types of decision-making environments and decision making approaches in business, industry and real life, formulate them into mathematical models with constraints in the form of equations and inequalities subject to different constraints and select appropriate technique to solve them.

Unit I: Reliability-I

- Reliability I: Definitions and Relationships between Survival Function, Failure Distribution, Hazard Function, Mean Time to Failure (MTTF)
- Reliability of Systems, Parametric Distributions - Weibull, Gamma, Lognormal and Exponential as life time Distributions. Derivation of reliabilities for these distributions. (9 lectures)

Unit II: Reliability-II

- Reliability II: Concept of Aging, IFR, IFRA classes of Distributions and their Dual, Coherent System as Binary Function: Minimal Cut and Path Sets (vectors)
- Representation of Structure Function of Series, Parallel and k out of n: G Systems of Independent Components. Redundancy: Parallel Redundancy, Standby Redundancy, Imperfect Switching. (9 lectures)

Unit III: Queuing Theory

- Queuing Theory, single-channel queuing model, queuing cost behaviour, multiple channel queuing model
- Steady-state solutions of M/M/1 and M/M/c models with associated distributions of queue length and waiting time. (9 lectures)

Unit IV: Simulation

- Simulation, its applications, Monte Carlo method, generation of random numbers from probability distributions
- Simulation in spreadsheets, variance reduction techniques, regenerative method. (9 lectures)

Unit V: Statistical quality control

- Statistical quality control, process control, product control, control charts for variables and attributes
- Modified control charts, sampling inspection plans (one, two, multiple and sequential), sampling inspection by variables. (9 lectures)

References

1. Kapur and Lamberson (1977) Reliability in Engineering Design, John Wiley & Sons. New York.
2. L.J. Bain and M. Englhardt (1991) Statistical analysis of reliability and life testing models, Marcel Dekker.
3. Zacks, S (1992) Introduction to reliability analysis probability model customers in a queue and so on in both real and virtual queues.
4. Kleinrock L. (1975) Queueing Systems, vol. 1, Theory; John Wiley
5. Saaty T.L. (1961) Elements of Queueing Theory with Applications; McGraw Hill.
6. Grant E.L (1964). Statistical Quality Control, McGraw Hill.
7. Duncan A.J (1974). Quality Control and Industrial Statistics, Taraporewala and Sons.

Fourth Semester
Paper 404 (ii)
Actuarial Statistics II

Course Outcome: With paper number 403 (ii) introducing the concept of insurance and the mathematical models that are used in designing the insurance policies, this paper extends the different concepts of insurance such as the Life insurance, Life annuities, Net premiums and Net premium reserves along with their derivatives. All the concepts of death-level benefit insurance, endowment insurance, differed insurance, varying benefit insurance, varying annuities, recursions, claim amount distributions, stop-loss insurance etc. are taught.

Learning Outcome: The course shall provide the students with knowledge about the basic models related to Actuarial Statistics. The knowledge of this course shall be helpful to them if they want to take up a career in Actuarial Science after completion of the program.

Unit I: Principles of compound interest

- Principles of compound interest.
- Nominal and effective rates of interest and discount, force of interest and discount, compound interest, accumulation factor, continuous compounding. (9 lectures)

Unit II: Life insurance and Endowment insurance

- Life insurance: Insurance payable at the moment of death and at the end of the year of death-level benefit insurance,
- Endowment insurance, differed insurance and varying benefit insurance, recursions, commutation functions. (9 lectures)

Unit III: Life annuities

- Life annuities: Single payment, continuous life annuities, discrete life annuities, life annuities with monthly payments
- Commutation functions, varying annuities, recursions, complete annuities-immediate and apportionable annuities due. (9 lectures)

Unit IV: Net premiums and Apporionable premiums

- Net premiums: Continuous and discrete premiums, true monthly payment premiums

- Apporionable premiums, commutation functions, accumulation type benefits. Payment premiums, apportionable premiums, commutation functions, accumulation type benefits. (9 lectures)

Unit V: Net premium reserves and Credibility Theory

- Net premium reserves: Continuous and discrete net premium reserve, allocations of loss to policy years, commutation functions.
- Claim amount distributions, stop-loss insurance. Credibility Theory: credibility premium, credibility factor, Bayesian and empirical approaches, applications to credibility premiums for standard models. (9 lectures)

References

1. Atkinson, M.E. and Dickson, D.C.M. (2000). An Introduction to Actuarial Studies, Elgar Publishing.
2. Bedford, T. and Cooke, R. (2001). Probabilistic risk analysis, Cambridge.
3. Bowers et al. (1997). Actuarial Mathematics, Second Edition. Society of Actuaries.
4. Philip, M. et. al (1999). Modern Actuarial Theory and Practice, Chapman and Hall.
5. N.L. Bowers, H.U. Gerber, J.C. Hickman, D.A. Jones and C.J. Nesbitt, (1986). Actuarial Mathematics, Society of Actuarias, Ithaca, Illiois, U.S.A. Second Edition (1997)
6. Trowbridge, C. L. (1989). Fundamental Concepts of Actuarial Science, Actuarial Education and Research Fund, USA.

Fourth Semester
Paper 405
Project Work and Practical

Course Outcome: This paper enables to train the students to undertake projects individually and also aims at providing an accessible introduction to various machine learning methods and applications in R and Python. The projects shall enable the students to take up their own statistical study and to understand the application of statistical methods that they learned during the course. The project requires the students to synthesize the topics from the course into some theme of practical use. They are expected to design computer programs or use statistical software for computational purpose. For the project each student shall work under the supervision of a faculty member of the department. In the practical part students shall be provided with hands on training on how to apply packages like R and Python for performing advanced analytical tools of machine learning and multivariate analysis.

Learning Outcome: With this course student shall develop their skill as a researcher in Statistics. They shall learn how to identify a research problem. How to frame the objectives of the research and develop necessary methodology to attain the objectives of the research. Data collection, data handling, data cleaning, data analysis, report writing etc. are also some of the learning objectives of the course. Students shall get some exposure to practical analysis of Big Data, one of the demanding qualification for modern day data analyst.

Project Work

In consultation of the supervisor, students shall decide on a researchable topic for their project. The topic shall be presented before all faculty members for approval with details objective and methodology. Once approved the student shall work on the project. There shall be two mid-term evaluation of the project to appraise the continuous progress. Before the start of the end-semester examination students are required to submit the project report/dissertation in hard copy in duplicate. During the end semester examination students shall present the same, whereby they shall be evaluated by an external examiner.

Practical

Practicals in Machine Learning and Multivariate Data analysis in R and Python

References

1. Brett, L. (2013). Machine Learning with R, PACKT Publishing Limited, Birmingham, UK.
2. Bradley, B. and Brandon, G. (2020). Hands-On Machine Learning with R, Chapman and Hall/CRC.
3. Ghatak, A. (2017). Machine Learning with R, Springer.