Choice Based Credit System (CBCS)
(Effective from Academic Year 2021-22)

Syllabus
for
Master of Science in Statistics

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

Master of Science in Statistics

Programme Outcomes (PO):

At the end of the 2-yrs. M.Sc. in Statistics programme, a student will be able to do the following:

1. Develop statistical thinking and problem solving.
2. Understand, formulate and use statistical models arising in various fields of study.
3. To learn from data and apply knowledge of Statistics in various fields of study.
4. Acquire knowledge and understanding in advanced areas of statistics, chosen by the student from the courses available.
5. Handle and analyse large databases with computer skills and use their results and interpretations to make practical suggestions for improvement.

Programme Specific Outcomes (PSO):

On successful completion of the course a student will be able to:

1. Gain sound knowledge in theoretical and practical aspects of Statistics.
2. To apply statistical tools at a number of data generating fields in real life problems.
3. To handle large data sets and carry out data analysis using software and programming languages such as R, Python and SAS.
4. Gain statistical skills, including problem-solving, project work and presentation so as to enable students to take prominent roles in a wide spectrum of employment and research.
5. Describe complex statistical ideas to non-statisticians.
6. Get a wide range of job opportunities in industry as well as in the government sector.

Outline of Syllabus for M. Sc. in Statistics:

Semester-I					
Srl.	Code	Type	Paper Name	Credit	Marks
1	STAPGCCT01	Core (T)	Linear Algebra and Linear Models	4	50
2	STAPGCCT02	Core (T)	Probability Theory	4	50
3	STAPGCCT03	Core (T)	Statistical Inference-I	4	50
4	STAPGCCT04	Core (T)	Mathematical Analysis and Statistical Simulation	4	50
5	STAPGCCP01	Core (P)	Programming in R and Python	4	50
6	PGAUC01	Compulsory	Elementary Arabic and Islamic Studies	0	50
Total				20	250
Semester-II					
Srl.	Code	Type	Paper Name	Credit	Marks
1	STAPGCCT05	Core (T)	Regression Analysis	4	50
2	STAPGCCT06	Core (T)	Multivariate Analysis	4	50
3	STAPGCCT07	Core (T)	Statistical Inference-II	4	50
4	STAPGCCT08	Core (T)	Time Series Analysis	4	50
5	STAPGCCP02	Core (P)	Programming in SAS and Statistical Lab-I	4	50
6	PGAEC01	Compulsory	Disaster Management/Human Rights & Value Education/Yoga & Life Skills	0	50
Total				20	250

Semester-III

Srl.	Code	Type	Paper Name	Credit	Marks
1	STAPGCCT09	Core (T)	Stochastic Processes	4	50
2	STAPGCCT10	Core (T)	Design of Experiments	4	50
3	STAPGDET01	DE (T)	Any one from: 1. Survival Analysis 2. Statistical Learning with Big Data-I 3. Reliability Theory 4. Actuarial Statistics	4	50
4	STAPGDET02	DE (T)	Any one from: 1. Statistical Genetics and Ecology 2. Advanced Data Analytic Techniques 3. Demography 4. Operations Research	4	50
5	...PGGEC01	GE	To be chosen from other PG programme	4	50
			Total	20	250

Semester-IV

Srl.	Code	Type	Paper Name	Credit	Marks
1	STAPGCCT11	Core (T)	Sample Survey	4	50
2	STAPGDET03	DE (T)	Any one from: 1. Clinical Trials and Epidemiology 2. Statistical Learning with Big Data-II 3. Statistical Process and Quality Control	4	50
3	STAPGDEP01	DE (P)	Any one from: 1. Statistical Lab-II 2. Statistical Lab-III 3. Statistical Lab-IV	4	50
4	...PGGEC02	GEC	To be chosen from other PG programme	4	50
5	STAPGPRJ01	PRJ	Project and Dissertation	4	50
			Total	20	250

Prerequisites for Statistical Lab-II: 1.Survival Analysis, 2. Statistical Genetics and Ecology 3. Clinical Trials and Epidemiology

Prerequisites for Statistical Lab-III: 1. Statistical Learning with Big Data-I, 2. Advanced Data Analytic Techniques, 3. Statistical Learning with Big Data-II

Prerequisites for Statistical Lab-IV: 1. Reliability Theory or Actuarial Statistics 2. Statistical Process and Quality Control

Generic Elective Courses:

1. STAPGGEC01 Statistical Methods using R
2. STAPGGEC02 Survey Sampling and Experimental Designs, or Probability and Stochastic Processes

Detailed Syllabus:

Semester-I

STAPGCCT01 Linear Algebra and 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.

Teaching Plan:

Week 1: Concept of Vector spaces and Subspaces with examples, Direct sum of subspaces.

Week 2 : Algebra of subspaces viz. sum, intersection, union etc, Linear combinations, Spanning sets, Linear spans

Week 3: Linear dependence and independence in vector spaces, Row and Column space of a matrix, Basis and Dimensions.

Week 4: Orthogonality, Orthonormal sets and Bases, Gram Schmidt Orthogonalization Process.

Week 5: Eigenvalues and eigenvectors.

Week 6: 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.

Week 7: Algebraic and geometric multiplicity of characteristic roots, Diagonalization of matrices, Factorization of a matrix, Eigenvalues and eigenvectors for solution of Differential equations

Week 8-9: 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.

Week 10-11: Real quadratic form, Linear transformation of quadratic forms, Index and signature, Singular value decomposition. Optimum values of a quadratic form, Vector and matrix differentiation.

Week 12-13: 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. Linear models for correlated errors.

Week 14: ANOVA: fixed, random and mixed effects model, ANCOVA, Multiple comparison, S-method and T-method of multiple comparison

STAPGCCT02 Probability Theory

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.

Teaching Plan:

Week 1-2: Classes of sets, fields, σ -fields, minimal σ -field, Borel σ – field in R^K, sequence of sets, limsup and liminf of a sequence of sets.

Week 3-4: Measure, Probability measure, properties of a measure, Caratheodory extension theorem (statement only), Lebesgue and Lebesgue-Stieltjes measures on R^K.

Week 5-6: 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.

Week 7-8: Integration of complex-valued functions, characteristic functions. Inversion and Continuity theorems.

Week 9: Radon-Nikodym theorem (statement and use), Product measure and Fubini's theorem (statement and use). Borel –Cantelli lemma.

Week 10-12: Sequence of random variables and modes of convergence (convergence in distribution, in probability, almost surely, and quadratic mean) and their interrelations.

Week 13-14: Scheffe's theorem, Slutsky's theorem. Laws of large numbers. Central Limit Theorems. Asymptotic normality.

STAPGCCT03 Statistical Inference-I

Course Objectives: This course introduces students to the basic theory behind the development and assessment of statistical analysis techniques in the areas of point estimation and hypothesis testing.

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

1. Explain in detail the notion of a parametric model and point estimation of the parameters of those models.

2. Understand the different principles of data reduction.

3. Demonstrate the plausibility of pre-specified ideas about the parameters of the model by examining the area of hypothesis testing.

4. Understand Neyman-Pearson fundamental lemma, UMP test, Interval estimation and confidence interval.

5. Apply large sample tests to real life datasets.

Detailed Syllabus:

Properties of estimator, mean square error and minimum MSE estimator, unbiasedness and minimum variance unbiased estimator, Rao-Cramer lower bound of variance, statement of Bhattacharya's bound (6)

Data reduction, sufficiency, factorization theorem and its illustration, concept of minimal sufficiency, Exponential family (4)

Completeness, bounded completeness, Rao-Blackwell and Lehmann-Scheffe theorems (3)

Methods of estimation: method of moments, method of maximum likelihood (3)

Review of testing of hypothesis, Neymann Pearson Lemma, Heuristic approach of derivation of tests from Binomial, Poisson, Univariate and Bivariate normal distributions (6)

Randomized and non randomized tests, Neyman- Pearsonian theory of testing of hypothesis, Neyman-Pearson fundamental lemma, Generalised NP lemma MP, UMP, and LMP tests, unbiasedness, UMPU test (8)

Families of distributions with monotone likelihood ratio property, exponential family of distributions (5)

Test for composite hypothesis: similar test and test with Neyman structure, case involving nuisance parameter (5)

Likelihood ratio test for standard univariate continuous distributions. (5)

Large sample tests using variance stabilizing transformations, Pearsonian chi-square (5)

References:

1. E. L. Lehman : Testing of Statistical Hypotheses
2. G. Casella & R. L. Berger: Statistical Inference
3. C.R. Rao: Linear Statistical Inference and its Applications
4. Gibbon : Non parametric Inference
5. T. S. Ferguson : Mathematical Statistics
6. B. K. Ghosh : Sequential Tests of Statistical Hypotheses
7. D. A. S. Fraser : Nonparametric methods in Statistics
8. J. O. Berger : Statistical Decision Theory and Bayesian Analysis
9. A. Wald : Sequential Analysis

Teaching Plan:

Week 1-2: Properties of estimator, mean square error and minimum MSE estimator, unbiasedness and minimum variance unbiased estimator, Rao-Cramer lower bound of variance, statement of Bhattacharya's bound

Week 3-4: Data reduction, sufficiency, factorization theorem and its illustration, concept of minimal sufficiency, Exponential family. Completeness, bounded completeness, Rao-Blackwell and Lehmann-Scheffe theorems

Week 5: Methods of estimation: method of moments, method of maximum likelihood

Week 6-7: Review of testing of hypothesis, Neymann Pearson Lemma, Heuristic approach of derivation of tests from Binomial, Poisson, Univariate and Bivariate normal distributions

Week 8-10: Randomized and non randomized tests, Neyman- Pearsonian theory of testing of hypothesis, Neyman- Pearson fundamental lemma, Generalised NP lemma MP, UMP, and LMP tests, unbiasedness, UMPU test. Families of distributions with monotone likelihood ratio property, exponential family of distributions

Week 11: Test for composite hypothesis : similar test and test with Neyman structure, case involving nuisance parameter.

Week 12: Likelihood ratio test for standard univariate continuous distributions.

Week 13-14: Large sample tests using variance stabilizing transformations, Pearsonian chi-square

STAPGCCT04 Mathematical Analysis and Statistical Simulation

Course Objectives: The aim of this course is

1. To learn some important topics of Mathematical analysis.
2. To know the idea behind statistical simulations.
3. To generate random numbers from any distribution using different simulation algorithms.

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

1. Gain clear understanding of Riemann integral
2. Understand and apply multiple and repeated integrals
3. Solve problems related to convergence of sequence and series of functions and power series.
4. Generate both discrete and continuous random numbers
5. Apply different simulation techniques.

Detailed Syllabus:

Review of Sequence and Series, Functions of bounded variation, Riemann integration and Riemann-Stieltjes integration, Statement of the standard property and problem based on them, Multiple integrals, repeated integrals, Change of variables in multiple integration.

(12)

Differentiation under integral sign, Leibnitz rule, Dirichlet integral, Liouville's extension, Uniform convergence of sequence of functions and series of functions, Cauchy's criteria and Weierstrass M-test, Power Series. (13)

Simulation and its uses, Definition of System, Types of Systems, Simulation Experiments and Field Experiments, Random Number Generators from Uniform and other Continuous and Discrete Distributions, Tests of Randomness and Goodness of Fit. (12)

Simulation of random variables from multivariate distributions and stochastic processes, Monte-Carlo methods. Computer Intensive Inference Methods - Jack-Knife, Bootstrap, cross validation, Monte Carlo methods and permutation tests, Importance sampling, Metropolis Hastings Algorithm. (13)

References:

1. R. G. Bartle (1976). Elements of Analysis, John Wiley & Sons.
2. W. Rudin (1985). Principles of Mathematical Analysis, McGraw Hill.
3. K. A. Rose (2004). Elementary Analysis: The Theory of Calculus, Springer (SIE).
4. S. M. Ross. Simulation, 5th ed., Academic press.
5. R. Y. Rubinstein. Simulation and the Monte Carlo Method, 3rd ed., John Wiley & Sons.
6. D. E. Knuth. The Art of Computer Programming, Vol. 1,2/Semi numerical Algorithms, Pearson Education (Asia).

Teaching Plan:

Week 1-3: Review of Sequence and Series, Functions of bounded variation, Riemann integration and Riemann-Stieltjes integration, Statement of the standard property and problem based on them, Multiple integrals, repeated integrals, Change of variables in multiple integration.

Week 4-7: Differentiation under integral sign, Leibnitz rule, Dirichlet integral, Liouville's extension, Uniform convergence of sequence of functions and series of functions, Cauchy's criteria and Weierstrass M-test, Power Series.

Week 8-9: Simulation and its uses, Definition of System, Types of Systems, Simulation Experiments and Field Experiments, Random Number Generators from Uniform and other Continuous and Discrete Distributions, Tests of Randomness and Goodness of Fit.

Week 10-11: Simulation of random variables from multivariate distributions and stochastic processes, Monte-Carlo methods.

Week 12-14: Computer Intensive Inference Methods - Jack-Knife, Bootstrap, cross validation, Monte Carlo methods and permutation tests, Importance sampling, Metropolis Hastings Algorithm.

STAPGCCP01 Programming in R and Python

Course Objectives:

In this course students will

1. Learn and practice about Writing and compiling codes in R, writing outputs in word.
2. Handle datasets in R
3. Use loops and controls in R, functions in R
4. Learn and practice statistical computations in R
5. Learn about Python programming and write program in Python

Learning Outcomes:

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

1. Write and execute programs in R.
2. Perform various statistical computations using R.
3. Testing of hypothesis using R
4. Simulation of standard distributions using R
5. Write programs and some basic computations using Python

Detailed Syllabus:

Introduction to R; R help; help.search(), R mailing list, contributed documentation on CRAN. Data types in R : numeric/character/logical; real/integer/complex, strings and the paste command, matrices, data frames, lists, Creation of new variables, Creation of patterned variables, Saving workspace/history. Writing programs in R markdown. (8)

Importing and exporting datasets, subsetting datasets, various datasets in R (4)

Graphs in R: the plot command, histogram, bar plot, box plot, points, 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. (6)

Functions and loops in R, Programming in R. (6)

Basic statistical tests using R: 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. (6)

Vector and Matrix operations such as addition, subtraction, multiplication, matrix inverse, Linear equations and eigenvalues, matrix decomposition - LU, QR and SVD. (4)

Linear models: the lm function; ANOVA/ANCOVA/regression, models, the summary function, goodness of fit measures, predicted values and residuals. (4)

Simulation: Random no. generation and Simulations: runif, rnorm, rchisq, rt, rbinom, sample etc.; set.seed, Monte Carlo techniques. (4)

Python Programming: Basic Idea; Simple Syntax; Basic Operations; Different Libraries; Function; Loop; Array

Data handling and management; Chart and Diagrams; Random Number Generation from a known and unknown distribution; Simulation; Application in various Statistical field; Idea of Parallel Computing and/or Efficient Programming (8)

References:

1. Introductory Statistics with R, Peter Dalgaard, Springer, (2008), 2nd Edition
2. An introduction to R: Longhow Lam
3. An Introduction to Statistical Learning with applications in R, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani; Springer, (2013)
4. Extending the Linear Model with R; Julian J. Faraway (2006)
5. Practical Regression and Anova using R, Julian J. Faraway (2002)
6. Data Analysis & Graphics using R, An example based approach, John Maindonald and W. John Braun; 3rd Edition, (2010), Cambridge University Press.
7. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, Sebastian Raschka, Vahid Mirjalili; Packt Publishing (2019)
8. Mastering Python for Data Science, Samir Madhavan; Packt Publishing (2015)
9. Introduction to Machine Learning with Python_ A Guide for Data Scientists, Andreas C. Mueller, Sarah Guido; O'Reilly Media (2016)

Teaching Plan:

Week 1-2: Introduction to R; R help; help.search(), R mailing list, contributed documentation on CRAN. Data types in R : numeric/character/logical; real/integer/complex, strings and the paste command, matrices, data frames, lists, Creation of new variables, Creation of patterned variables, Saving workspace/history. Writing programs in R markdown.

Week 3 : Importing and exporting datasets, subsetting a dataset

Week 4-5: Graphs in R: the plot command, histogram, bar plot, box plot, points, 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.

Week 6-7: Functions and loops in R, Writing own functions.

Week 8-9: Basic statistics using R: 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.

Week 10: Vector and Matrix operations such as addition, subtraction, multiplication, matrix inverse, Linear equations and eigenvalues, matrix decomposition - LU, QR and SVD.

Week 11: Linear models: the lm function; ANOVA/ANCOVA/regression, models, the summary function, goodness of fit measures, predicted values and residuals.

Week 12: Simulation: Random no. generation and Simulations: runif, rnorm, rchisq, rt, rbinom, sample etc.; set.seed, Monte Carlo techniques.

Week 13: Python Programming: Basic Idea; Simple Syntax; Basic Operations; Different Libraries; Function; Loop; Array

Week 14: Data handling and management; Chart and Diagrams; Random Number Generation from a known and unknown distribution; Simulation; Application in various Statistical field; Idea of Parallel Computing and/or Efficient Programming

Semester-II

STAPGCCT05 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

Teaching Plan:

Week 1-3: General theory of regression, simple and multiple regression, fitting of polynomial regression by orthogonal methods, examination of regression equation.

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

Week 5-6: 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

Week 7-9: Multicollinearity – detection and remedial measures. Ridge regression and Lasso.

Week 10-11: Departures from the Gauss-Markov set-up: Heteroscedasticity and Autocorrelation – detection and remedies.

Week 12-13: Logistic regression, Dummy variables, piecewise regression, splines and scatter plot smoothing.

Week 14: Checking for normality: Q-Q plots, Normal Probability plot, Shapiro-Wilks test.

STAPGCCT06 Multivariate Analysis

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 mean 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 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

Teaching Plan:

Week 1-2: Multivariate normal distribution and its properties. Sampling from Multivariate normal distribution – independence of sample mean vector and variance-covariance matrix. Wishart distribution. Distribution of quadratic forms – Cochran's theorem.

Week 3-4: 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.

Week 5-6: Multivariate linear model: estimation of parameters, tests of linear hypotheses

Week 7-8: Multivariate Analysis of variance of one and two way classified data, simultaneous confidence intervals, Multivariate Analysis of Covariance.

Week 9-10: Principal Component Analysis: Population and sample Principal components and their uses. Plotting techniques.

Week 11-12: Factor Analysis: The orthogonal factor model, Estimation of factor loading, Estimation of Factor scores, Interpretation of Factor Analysis.

Week 13: Cluster Analysis

Week 14: Classification and discrimination procedures for discrimination between two multivariate normal populations- sample discriminant function.

STAPGCCT11 Statistical Inference-II

Course Objectives:

1. To learn various decision rules theories and its applications of decision making as individuals, in groups, and in organizations.
2. To learn parametric, non-parametric and sequential estimation (point, as well as, interval) and testing (simple, as well as, composite hypotheses) procedures.

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:

1. Perform sequential estimation techniques and testing procedures to deal with real life problems.
2. Understand UMPU tests, SPRT, OC and ASN.
3. Implement nonparametric statistical tests.
4. Understand decision problem, loss function, risk function and decision rules.
5. Treat "evidence" as the value of observations and prescribe methods to deal rationally with it.
6. Equip students with skills to carry out and interpret posterior and preposterior data based modeling and analyses.
7. Compute probability that the theory in question could produce the observed data.

Detailed Syllabus:

Theory of interval estimation, UMA, UMAU confidence intervals, shortest expected length confidence interval (4)

Sequential procedures, Wald's SPRT and its properties, fundamental identity, OC and ASN functions, optimality of SPRT (4)

An introduction of inference procedure based on random effects (3)

Nonparametric Methods: Sign test, Mann-Whitney test, Run test, Test of randomness, Confidence limits for Quantiles based on Sign test statistic. (15)

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. (24)

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.

Teaching Plan:

Week 1: Theory of interval estimation, UMA, UMAU confidence intervals, shortest expected length confidence interval

Week 2: Sequential procedures, Wald's SPRT and its properties, fundamental identity, OC and ASN functions, optimality of SPRT

Week 3: An introduction of inference procedure based on random effects

Week 4-8: Nonparametric Methods: Sign test, Mann-Whitney test, Run test, Test of randomness, Confidence limits for Quantiles based on Sign test statistic.

Week 9-14: 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.

STAPGCCT08 Time Series Analysis

Course Objectives:

1. To learn and develop a scientific view to understand the time series data and its analysis.
2. To learn stationary and non-stationary, and seasonal and non-seasonal time series models.
3. Learn to estimate model parameters and compare different models developed for the same dataset in terms of their estimation and prediction accuracy.

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

1. Understand the concepts of time series and their application to health, climate, finance and other areas.
2. To learn and compute ACVF and ACF.
3. Remove trend and seasonality using different methods to convert the time series into stationary.
4. Apply auto regressive, moving average, ARMA, ARIMA models, Box-Jenkins approach to forecast time-series data empirically.
5. Check and validate models with its residual analysis and diagnostic checking.

6. Understand Correlogram and Periodogram analysis and different Smoothing methods in R.

Detailed Syllabus:

Time-series as discrete parameter stochastic process, autocovariance and autocorrelation functions and their properties. (6)

Exploratory time Series analysis, tests for trend and seasonality, exponential and moving average smoothing. Holt and Winters smoothing, forecasting based on smoothing. (6)

Detailed study of the stationary processes: (1) moving average (MA), (2) auto regressive (AR), (3) ARMA and (4) AR integrated MA (ARIMA) models. Box-Jenkins models, choice of AR and MA periods. (18)

Linear predictor operator and its application. Durbin Levinson algorithm and Innovation algorithm.Discussion (without proof) of estimation of mean, autocovariance and autocorrelation functions under large sample theory, estimation of ARIMA model parameters. Wold Decomposition Theorem. (10)

Spectral analysis of weakly stationary process, periodogram and correlogram analysis, computations based on Fourier transform. (10)

References:

1. C. Chatfield : The Analysis of Time Series – An Introduction
2. G.E.P. Box ,G.M. Jenkins & G.C.Reinsel : Time Series Analysis – Forecasting & Control
3. P. J. Brockwell & R.A. Davis : Introduction to Time Series Analysis and Forecasting
4. A.Pankratz : Forecasting with Univariate Box-Jenkins Model
5. G. Janacek and L. Swift : Time Series –Forecasting, Simulation, Applications

Teaching Plan:

Week 1-2: Time-series as discrete parameter stochastic process, autocovariance and autocorrelation functions and their properties.

Week 3-4 : Exploratory time Series analysis, tests for trend and seasonality, exponential and moving average smoothing. Holt and Winters smoothing, forecasting based on smoothing.

Week 5-7: Detailed study of the stationary processes: (1) moving average (MA), (2) auto regressive (AR), (3) ARMA and (4) AR integrated MA (ARIMA) models. Box-Jenkins models, choice of AR and MA periods.

Week 8-11:Linear predictor operator and its application. Durbin Levinson algorithm and Innovation algorithm.Discussion (without proof) of estimation of mean, autocovariance and autocorrelation functions under large sample theory, estimation of ARIMA model parameters. Wold Decomposition Theorem.

Week 12-14: Spectral analysis of weakly stationary process, periodogram and correlogram analysis, computations based on Fourier transform.

STAPGCCP02 Programming in SAS and Statistical Lab-I

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 (4)

Control and Loops: Do-loops, IF-THEN-ELSE etc. Functions, Arrays. (4)

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. (4)

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

Graphical Representation: Proc GPLOT etc. (5)

Simulation: generating random variables, simulating standard univariate and multivariate distributions. (4)

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. (4)

Practicals to be performed using SAS/R/Python:

Regression analysis:

1. Problems based on simple and multiple linear regression models to real as well as simulated data sets.
2. Practical problems related to Multicollinearity and autocorrelation.
3. Problems based on model selection.
4. Computation of multiple and partial correlation and checking residual diagnostic to validate the model.
5. Problems based on application of logistic regression.

Multivariate Analysis:

1. Practical based on Multivariate linear model, estimation of parameters and test of linear hypotheses
2. Practical based on Principal Component Analysis
3. Practical based on Cluster Analysis
4. Practical based on Classification and discrimination procedures

Statistical Inference-II:

1. Practical problems based on finding bayes estimator
2. Practical problems based on Sign test, Mann-Whitney test, Run test, Test of randomness

Time Series Analysis

1. Converting to time series data and plotting of the data.

2. Problem based on different smoothing techniques and their forecasting.
3. Testing stationarity of a dataset
4. Fitting of ARIMA/sARIMA models and their forecasting methods using innovation algorithms.
5. Spectral analysis of the time series data.

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

Teaching Plan:

Week 1: 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

Week 2: Control and Loops: Do-loops, IF-THEN-ELSE etc. Functions, Arrays.

Data 3: Data Handling: Importing and Exporting data sets, Sorting data sets, creating new variable, subsetting data sets using DROP, KEEP, IF-THEN etc., Merging data sets, appending data sets, uses of CLASS, BY etc.

Week 4-5: Descriptive Statistics: Proc statements MEANS, UNIVARIATE, SORT, FREQ, CORR, TABULATE, Graphical Representation: Proc GPLOT etc.

Week 6: Simulation: generating random variables, simulating standard univariate and multivariate distributions.

Week 7: 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.

Week 8-9: Practicals on Regression

Week 10-11: Practicals on Multivariate

Week 12: Practicals on Bayesian statistics and Nonparametric methods.

Week 13-14 Practicals on Time Series

Semester-III

STAPGCCT09 Stochastic Processes

Course Objectives:

In this course students will

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:

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 (20)

Branching process. Galton-Watson branching process, estimation of probability of extinction. (5)

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. (10)

Brownian Motion: Limit of Random Walk, Its Defining Characteristics and Peculiarities. Its Variations: Standard Brownian Motion, Brownian Bridge, Brownian Motion Reflected at Origin, Geometric Brownian Motion, Brownian Motion with Drift. Reflection Principle. Some Applications. Stochastic Calculus, Stochastic Differential Equations (15)

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

Teaching Plan:

Week 1-2: Introduction to Stochastic Processes, Classification according to state and time with examples, Discrete time Markov chain, Transition probability matrix.

Week 3: Classification and finding of states according to transient and recurrent and related results.

Week 4: Periodicity of a Markov chain, Restricted random walk.

Week 5: Stationary distribution of discrete time Markov chain.

Week 6: Branching process. Galton-Watson branching process, estimation of probability of extinction.

Week 7-9: 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.

Week 10-11: Brownian Motion: Limit of Random Walk, Its Defining Characteristics and Peculiarities. Its Variations: Standard Brownian Motion, Brownian Bridge, Brownian Motion Reflected at Origin.

Week 12-14: Geometric Brownian Motion, Brownian Motion with Drift. Reflection Principle. Some Applications. Stochastic Calculus, Stochastic Differential Equations.

STAPGCCT10 Design of Experiments

Course Objectives:

1. To learn the basic principles in the design of simple experiments.
2. To learn different tests for comparing pairs of treatment means, ANCOVA, factorial experiments, fractional factorial experiments, confounding, BIBD, PBIBD with solving real life examples.
3. To learn the applications of different designs in agriculture.

Learning Outcomes:

1. Compare the pairs of treatment means using different methods when the null hypothesis is rejected in ANOVA.
2. Analyze the data using split plot, strip plot and general factorial experiments.
3. Construct fractional factorial experiments and apply confounding in real life problems.
4. Understand the analysis of BIBD, PBIBD, Quasi-Latin square, Youden square and cross over design and their applications in agriculture, business and industries.

Detailed Syllabus:

Basic principles of design, elimination of heterogeneity in one and two directions , Missing plot techniques. (10)

Block Designs: Connectedness, Orthogonality and balancing; intra block analysis of BIBD, resolvable and affine resolvable designs. (6)

Intrablock analysis of BIB, Recovery of inter-block information in BIB designs; Row column and Youden Square designs, Elementary ideas of Lattice and PBIB design (group divisible only) (7)

Construction of mutually orthogonal Latin Squares (MOLS); Construction of BIBD through MOLS and other ways. (5)

Factorial experiment, Confounding and balancing in symmetric factorial experiments, Split plot and Strip plot techniques. (12)

Response Surface Designs: First-order response surface designs and orthogonal designs (7)

First-order response surface designs and its extension with correlated errors. (3)

References:

1. M.C. Chakraborty : Mathematics of Design and Analysis of Experiments
2. M. Das and N. Giri (1979): Design and Analysis of Experiments, Wiley Eastern.
3. A. Dey : Theory of Block Designs
4. D. Raghavarao : Constructions & Combinatorial Problems in Design of Experiments
5. D.Raghavarao & L.V.Padgett : Block Design: Analysis, Combinatorics and Applications

6. R.C. Bose : Mathematical Theory of Symmetric Factorial Design (Sankhya – Vol. 8)
7. D. G. Kabe and A. K. Gupta: Experimental Designs: Exercises and Solutions
8. G. Casella : Statistical Design
9. T. P. Ryan : Modern Experimental Design
10. C. F. J. Wu & M. S. Hamada : Experiments: Planning, Analysis and Optimization (2nd edition)
11. D.C. Montgomery : Design and Analysis of Experiments

Teaching Plan:

Week 1-3: Basic principles of design, elimination of heterogeneity in one and two directions, Missing plot techniques.

Week 4-5: Block Designs: Connectedness, Orthogonality and balancing; intra block analysis of BIBD, resolvable and affine resolvable designs.

Week 6-7: Intrablock analysis of BIB, Recovery of inter-block information in BIB designs; Row column and Youden Square designs, Elementary ideas of Lattice and PBIB design (group divisible only)

Week 8: Construction of mutually orthogonal Latin Squares (MOLS); Construction of BIBD through MOLS and other ways.

Week 9-11: Factorial experiment, Confounding and balancing in symmetric factorial experiments, Split plot and Strip plot techniques.

Week 12-13: Response Surface Designs: First-order response surface designs and orthogonal design

Week 14: First-order response surface designs and its extension with correlated errors.

STAPGDET01

Any one from:

1. Survival Analysis
2. Statistical Learning with Big Data-I
3. Reliability Theory
4. Actuarial Statistics

STAPGDET02

Any one from:

1. Statistical Genetics and Ecology
2. Advanced Data Analytic Techniques
3. Demography
4. Operations Research

STAPGGEC01

This is a generic elective course. To be chosen from another PG Programme.

Semester-IV

STAPGCCT11 Sample Survey

Course Objectives: The main objective of this course is

1. To learn scientific views to conduct the survey in the proper way to collect the data about specific perspectives.
2. To Learn a variety of probability and non-probability sampling methods for selecting a sample from a population.
3. To learn techniques and methodologies in survey sampling along with applications in real life problems.

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

1. Understand the basic principles underlying survey design and estimation.
2. Apply the different sampling methods for designing and selecting a sample from a population.
3. Implement Cluster sampling, Ratio and Regression estimation in real life problems.
4. Learn about various approaches (design based and model-based) to estimate admissible parameters under equal and unequal probability sampling design.
5. Apply various sampling methods; systematic, stratified and cluster sampling.
6. Understand the cluster and two stage sampling with varying sizes of clusters/first stage units.
7. Learn about the randomized response techniques.

Detailed Syllabus:

Probability sampling from a finite population-–notions of sampling design, sampling scheme, inclusion probabilities; some problems of sampling design construction based on inclusion probabilities (6)

Simple random sampling with and without replacement, Systematic sampling, Unequal probability sampling with and without replacement. Related estimators of population total / mean, their variances and variance estimators –Mean per distinct unit in simple random with replacement sampling, Des Raj and Murthy’s estimator in unequal probability sampling without replacement. (14)

Stratified sampling – Allocation problem and construction of strata (optimal, proportional and equal allocation) (4)

Estimation based on auxiliary data (involving one or more auxiliary variables) under design-based and model based approaches. Ratio, Product, Difference and Regression estimators, Unbiased Ratio estimators – Probability proportional to aggregate size sampling (8)

Sampling and subsampling of clusters, Two-stage sampling with equal/unequal number of second stage units and simple random sampling without replacement / unequal probability sampling with replacement at first stage, Ratio estimation in two-stage sampling. Two-way stratification, post-stratification, controlled sampling, (8)

Double sampling for stratification, Double sampling ratio and regression estimators, Sampling on successive occasions. randomized response techniques for one qualitative sensitive characteristic.

References: (10)

1. W. G. Cochran : Sampling Techniques, 3rd ed.
2. Des Raj & Chandak : Sampling Theory
3. A.S. Hedayat & B. K. Sinha : Design and inference in finite population sampling
4. P. Mukhopadhyay : Theory and Methods of Survey Sampling
5. M. N. Murthy : Sampling Theory and methods
6. S. Sampath: Sampling Theory and Methods

Teaching Plan:

Week 1-2: Probability sampling from a finite population-–notions of sampling design, sampling scheme, inclusion probabilities; some problems of sampling design construction based on inclusion probabilities

Week 3-4 : Simple random sampling and systematic sampling with related estimators of population total / mean, their variances and variance estimators

Week 5-6: Unequal probability sampling with and without replacement. Related estimators of population total / mean, their variances and variance estimators – Des Raj and Murthy's estimator in unequal probability sampling without replacement.

Week 7: Stratified sampling – Allocation problem and construction of strata (optimal, proportional and equal allocation)

Week 8-9: Estimation based on auxiliary data (involving one or more auxiliary variables) under design-based and model based approaches. Ratio, Product, Difference and Regression estimators, Unbiased Ratio estimators – Probability proportional to aggregate size sampling

Week 10-11: Sampling and subsampling of clusters, Two-stage sampling with equal/unequal number of second stage units and simple random sampling without replacement / unequal probability sampling with replacement at first stage, Ratio estimation in two-stage sampling. Two-way stratification, post-stratification, controlled sampling,

Week 12-14: Double sampling for stratification, Double sampling ratio and regression estimators, Sampling on successive occasions. Randomized response techniques for one qualitative sensitive characteristic.

STAPGDET03

Any one from:

1. Clinical Trials and Epidemiology
2. Statistical Learning with Big Data-II
3. Statistical Process and Quality Control

STAPGDEP01

Any one from:

1. Statistical Lab-II

This is practical based on

1. STAPGCCT09 Stochastic Processes
2. STAPGCCT10 Design of Experiments
3. STAPGCCT11 Sample Survey
4. STAPGDET01 Survival Analysis
5. STAPGDET02 Statistical Genetics and Ecology
6. STAPGDET03 Clinical Trials and Epidemiology

2. Statistical Lab-III

This is practical based on

1. STAPGCCT09 Stochastic Processes
2. STAPGCCT10 Design of Experiments
3. STAPGCCT11 Sample Survey

4. STAPGDET01 Statistical Learning with Big Data-I
5. STAPGDET02 Advanced Data Analytic Techniques
6. STAPGDET03 Statistical Learning with Big Data-I

3. Statistical Lab-IV

This is practical based on

1. STAPGCCT09 Stochastic Processes
2. STAPGCCT10 Design of Experiments
3. STAPGCCT11 Sample Survey
4. STAPGDET01 Reliability Theory / Actuarial Statistics
5. STAPGDET03 Statistical Process and Quality Control

STAPGGEC02

This is a generic elective course. To be chosen from another PG Programme.

STAPGPRJ01 Project and Dissertation

This should be some innovation in Statistics/application of Statistics in view of the developments in Statistics based on the knowledge gained during the program. The Project Work will be spread over the 2nd year (Sem-III and Sem-IV) of the program. The topic of their project work/dissertation will be decided at the beginning of the 3rd Semester by the Head of the Department in consultation with the supervisors. A project report is to be submitted by a student. The project report will include: a) Review of the relevant literature, b) Objectives of the study, c) Materials and Methods, d) Results/Observations (supported by figures/tables etc as required), e) Discussion of the Results/Observations, f) Summary and g) References. The project work will be evaluated by the teachers of the department by a presentation and viva-voce examination. Out of total 50 marks assigned to the project, 30 marks will be assigned on the evaluation of the project work separately by the examiners and 20 marks will be assigned on the oral presentation and viva – voce.

Steps in project work:

1. Conceptual phase-formulation of the research problem, literature review, developing the hypothesis.
2. Empirical phase- preparing the research design, determination of sample size, collection of data.
3. Analytical phase- analysis of data, hypothesis testing, generalization and interpretations, writing up, conclusions.

Department Specific Electives:

Survival Analysis

Course Objectives:

1. To learn several tools and techniques for complete/incomplete biological experimentations.
2. To learn various methods used for modelling and evaluating survival data, also called time-to-event data.

Learning Outcomes:

1. Recognize the difference between parametric and non-parametric (Kaplan Meier method) survival models,
2. Estimate survival probabilities both with and without the presence of covariates (Cox PH model).
3. Acquire the practical exposures to handle time to event data.
4. Handle real life data from clinical trials.

Detailed Syllabus:

Introduction. Basic functions and Models. various censoring mechanisms and likelihood functions based on that. (8)

Parametric univariate estimation: Standard models – exponential, Weibull, log-logistic, log-normal and Gamma. (6)

Nonparametric univariate estimation: Actuarial, Kaplan-Meier and Nelson-Aalen estimators. (8)

Tests of equality of survival functions: Gehan's and Mantel-Haenszel tests. (4)

Semiparametric regression models: Cox proportional hazard model – estimation, tests, diagnostics. (8)

Additive Models. Accelerated Models (4)

Competing risk theory, Multivariate survival models; Random effects models for survival data analysis. (8)

Frailty Models. (4)

References:

1. J.P. Klein & M. L. Moeschberger: Survival Analysis: Techniques for Censored and Truncated Data
2. P.J. Smith : Analysis of Failure and Survival Data

3. J.D. Kalbfleisch & R.L. Prentice : The Statistical Analysis of Failure Time Data, 2nd ed.
4. R.G. Miller : Survival Analysis
5. D.J. Kleinbaum & M. Klein : Survival Analysis – A Self-Learning Text

Teaching Plan:

Week 1-2: Introduction. Basic functions and Models. various censoring mechanisms and likelihood in those cases.

Week 3-5: Parametric univariate estimation: Standard models – exponential, Weibull, log-logistic, log-normal and Gamma.

Week 6-7: Nonparametric univariate estimation: Actuarial, Kaplan-Meier and Nelson-Aalen estimators.

Week 8: Tests of equality of survival functions: Gehan's and Mantel-Haenszel tests.

Week 9-10: Semiparametric regression models: Cox proportional hazard model – estimation, tests, diagnostics.

Week 11: Additive Models. Accelerated Models

Week 12-13: Competing risk theory, Multivariate survival models; Random effects models for survival data analysis.

Week 14: Frailty Models.

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

Teaching Plan:

Week 1-2: 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.

Week 3-4: 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.

Week 5: 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.

Week 6-7: 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.

Week 8-9: 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.

Week 10: Resampling Methods: Cross-Validation, Validation Set Approach, Leave-One-Out Cross-Validation, k-Fold Cross-Validation

Week 11-14: 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

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

Teaching Plan:

Week 1 - 2: Reliability concepts and measures, components and systems, coherent systems, reliability of coherent systems.

Week 3 - 5 : 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.

Week 6 - 8: 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.

Week 9 - 10: Reliability estimation based on failure times from variously censored life-tests data for parametric families.

Week 11 - 12: Kaplan – Meier estimation of reliability curve, Greenwood formula, non-parametric methods for comparison of several reliability curves, Log rank tests.

Week 13-14: Regression models in reliability, Cox PH and Accelerated failure time models; Estimation of parameters and diagnostics.

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.

Teaching Plan:

Week 1: Review of decision theory and actuarial applications

Week 2-4: 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.

Week 5-6: Risk models: models for claim number and claim amount in short-term contracts, moments, compound distributions, moments of insurers and reinsurer's share of aggregate claims.

Week 7: Review of Bayesian statistics/estimation and application to credibility theory.

Week 8-9: Experience rating: Rating methods in insurance and banking, claim probability calculation, stationary distribution of proportion of policyholders in various levels of discount.

Week 10-11: 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.

Week 12: Review of generalized linear model, residuals and diagnostics, goodness-of-fit, applications.

Week 13-14: 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.

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)

Introduction to ecology and evolution, population dynamics: 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.

Teaching Plan:

Week 1-4: 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.

Week 5-7: Description of a DNA sequence. Pair-wise alignment-Needleman-Wunsch algorithm, Discrimination using Markov Chain, Hidden Markov Models and estimation of parameters

Week 8-9: Introduction to ecology and evolution, population dynamics: single species growth models: Exponential, Logistic and Gompertz etc.,

Week 10-12: Growth in stochastic environment, stochastic differential models with application in Biology,

Week 13-14: Goodness-of-fit test for growth curves

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 (2^{nd} ed).

Teaching Plan:

Week 1-2: 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

Week 3-5 : 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.

Week 6-8: 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.

Week 9-12: 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.

Week 13-14: Models for Proportions: Binomial GLMs, Models for counts: Poisson GLMs

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.

Teaching Plan:

Week 1-3: 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.

Week 4-6: 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.

Week 7-8: Measures of Mortality; construction of abridged life tables, infant mortality rate and its adjustments, model life table.

Week 9-10: Stable and quasi-stable populations, intrinsic growth rate. Models of population growth and their filling to population data.

Week 11-14: Internal migration and its measurement, migration models, concept of international migration. Methods for population projection, component method of population projection, Nuptiality and its measurements.

Operations research

Course Objectives: The main objective is to formulate mathematical models based on different industrial problems and apply different tools in operation research for effective decision–making.

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

1. Understand and analyze managerial problems in industry so that they are able to use resources (capitals, materials, staffing, and machines) more effectively.
2. Understand the characteristics of different types of decision-making environments and decision making approaches.

3. Understand the mathematical tools that are needed to solve optimization problems.
4. Analyze the queueing and inventory situations.
5. Understand discrete event simulation and decision analysis with inclusion of modelling based on random events involving uncertainties.
6. Conceptualise optimum event management through Network scheduling.

Detailed Syllabus:

Definition and Scope of Operational Research, Phases in Operational Research, Different types of models and their construction. Transportation problem and assignment problem.

Game Theory – Definition – Saddle Point - Two Person Zero Sum Game - Pure and Mixed Strategies - Algebraic Solution Procedure - Graphical Solution – Principle of Dominance . Replacement Models - Replacement of Items that Deteriorate whose maintenance costs increase with time without change in the money value - Replacement of items that fail suddenly - group replacement policy.

Construction of Network – Rules & Precautions - C.P.M. & P.E.R.T. Networks - Obtaining Critical Path - Time estimates for activities - Probability of completion of project - Determination of floats (total, free, independent & interfering).

Queuing models – specification and effectiveness measures. Steady-state solutions of M/M/1 and M/M/c models with associated distributions of queue-length and waiting time. M/G/1 queue and Pollazcek-Khinchine result.

Analytical structure of inventory problems, EOQ formula of Harris, its sensitivity analysis and extensions allowing quantity discounts and shortages. Multi-item inventory subject to constraints. Models with random demand, the static risk model. P and Q- systems with constant and random lead times.

References:

1. Banks J. (1998). Handbook of Simulation: Principles, Methodology, Advances, Applications and Practice, John Wiley and Sons.
2. Gross, D., Shortle J.F., Thompson J.M. and Harris, C.M. (2008). Fundamentals of
3. Queueing Theory, John Wiley & Sons.
4. Hillier, F.S. and Lieberman, G.J. (2001). Introduction to Operations Research, 7th Ed., Irwin.
5. Hadley, G. and Whitin, T.M. (1963). Analysis of Inventory Systems, Prentice Hall.
6. Ross, S. M. (2013). Simulation, 5th Ed., Academic Press.
7. Taha, H. A. (2016). Operations Research: An Introduction, 10th Ed. Prentice Hall.
8. Winston, W.L. and Goldberg, J.B. (2004). Operations Research: Applications and

9. Algorithms, Thomson Brooks/Cole.

Teaching Plan:

Week 1-3: Definition and Scope of Operational Research, Phases in Operational Research, Different types of models and their construction. Transportation problem and assignment problem.

Week 4-5 : Game Theory – Definition – Saddle Point - Two Person Zero Sum Game - Pure and Mixed Strategies - Algebraic Solution Procedure - Graphical Solution – Principle of Dominance . Week 6-7: Replacement Models - Replacement of Items that Deteriorate whose maintenance costs increase with time without change in the money value - Replacement of items that fail suddenly. group replacement policy.

Week 8-9: Construction of Network – Rules & Precautions - C.P.M. & P.E.R.T. Networks - Obtaining Critical Path - Time estimates for activities - Probability of completion of project - Determination of floats (total, free, independent & interfering).

Week 10-11: Queuing models – specification and effectiveness measures. Steady-state solutions of M/M/1 and M/M/c models with associated distributions of queue-length and waiting time. M/G/1 queue and Pollazcek-Khinchine result.

Week 12-14: Analytical structure of inventory problems, EOQ formula of Harris, its sensitivity analysis and extensions allowing quantity discounts and shortages. Multi-item inventory subject to constraints. Models with random demand, the static risk model. P and Q- systems with constant and random lead times.

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 particular 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

Teaching Plan:

Week 1: 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.

Week 2-3: Data management: data definitions, case report forms, database design, data collection systems for good clinical practice.

Week 4-5: 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

Week 6-7: Reporting and analysis: analysis of categorical outcomes from Phase I – III trials, analysis of survival data from clinical trials.

Week 8: Introduction to Meta-analysis of clinical trials: Ideas of Meta Analysis, Fixed Effects Model, Random Effects Model, Analysis of Bias, Small sample effects

Week 9-11: 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)

Week 12-14: Design and Analysis of Epidemiologic studies, Epidemiological studies for certain particular diseases; Some modelling approaches for identifying the risk factors

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

Teaching Plan:

Week 1-3: 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.

Week 4-6: 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.

Week 7-8: Support Vector Machines: Maximal Margin Classifier, Support Vector Classifiers, Support Vector Machines, SVMs with More than Two Classes, Relationship to Logistic Regression.

Week 9-12: 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.

Week 13-14: Unsupervised Learning: Principal Components Analysis, Clustering Methods, K-Means Clustering, Hierarchical Clustering. Clustering methods from both statistical and data mining viewpoints, vector quantization.

Statistical Process and Quality Control

Course Objectives:

1. To develop a scientific view to analyze the industrial data from a specific perspective.
2. To learn the statistical quality control techniques used in industries such as control charts, acceptance sampling plans etc.
3. To learn some advanced control charts, capability indices and the concept of six-sigma.

Learning Outcomes:

1. Understand the basics of production process monitoring and apply the concept of control charts on it.
2. Apply the acceptance and continuous sampling plans in the production process.
3. Compute capability indices.
4. Know and apply the concept of weighted control charts, six sigma, ISO: 9000 series standards and Taguchi design.

Detailed Syllabus:

Basic concepts of process monitoring and control; process capability and process optimization. (4)

General theory and review of control charts for attribute and variable data; O.C. and A.R.L. of control charts; control by gauging; moving average and exponentially weighted moving average charts; Cu-Sum charts using V-masks and decision intervals; Economic design of X-bar chart. (6)

Acceptance sampling plans for attributes inspection; single and double sampling plans and their properties; plans for inspection by variables for one-sided and two sided specification. (8)

Mil Std. and IS plans; continuous sampling plans of Dodge type and Wald-Wolfiwitz type and then properties. (8)

Sequential sampling plan and its properties; Bayesian sampling plans. (8)

Capability indices Cp, Cpk and Cpm; estimation, confidence intervals and tests of hypotheses relating to capability indices for normally distributed characteristics. (6)

Use of design of experiments in SPC; factorial experiments, fractional factorial designs; construction of such designs and analysis of data. (6)

Multivariate quality control; use of control ellipsoid and of utility functions. (4)

References

1. Montgomery, D.C. (1985): Introduction to Statistical Quality Control; Wiley.
2. Montgomery, D.C. (1985): Design and Analysis of Experiments; Wiley.
3. Ott, E.R. (1975): Process Quality Control; McGraw Hill
4. Phadke, M.S. (1989): Quality Engineering Through Robust Design; Prentice Hall.
5. Wetherill, G.B. (1977): Sampling Inspection and Quality Control; Halsted Press.
6. Wetherill, G.B. and Brown, D.W.: Statistical Process Control: Theory and Practice.

Teaching Plan:

Week 1: Basic concepts of process monitoring and control; process capability and process optimization.

Week 2-3: General theory and review of control charts for attribute and variable data; O.C. and A.R.L. of control charts; control by gauging; moving average and exponentially weighted moving average charts; Cu-Sum charts using V-masks and decision intervals; Economic design of X-bar chart.

Week 4-5: Acceptance sampling plans for attributes inspection; single and double sampling plans and their properties; plans for inspection by variables for one-sided and two sided specification.

Week 6-7: Mil Std. and IS plans; continuous sampling plans of Dodge type and Wald-Wolfiwitz type and then properties.

Week 8-9: Sequential sampling plan and its properties; Bayesian sampling plans

Week 10-11: Capability indices Cp, Cpk and Cpm; estimation, confidence intervals and tests of hypotheses relating to capability indices for normally distributed characteristics.

Week 12-13: Use of design of experiments in SPC; factorial experiments, fractional factorial designs; construction of such designs and analysis of data.

Week 14: Multivariate quality control; use of control ellipsoid and of utility functions.

Statistical Lab-II

1. Sample Survey:

1. Selecting a sample from a given sampling design, calculation inclusion probabilities and construction of Horvitz-Thompson estimator.
2. Selecting a sample under equal probability sampling and estimation of parameters.
3. Problem based on PPSWR and PPSWOR.
4. Problem based on stratified sampling and cluster sampling.
5. Problem based on uses of auxiliary variables.
6. Problem based on randomized response technique.

2. Stochastic Processes:

1. Simulation of a Discrete Time Markov chain using transition probability matrix
2. calculation of m-step transition probabilities from 1-step transition probability
3. Estimation of 1-step and m-step transition probabilities from the realization of a Markov chain
4. Finding stationary distribution.

3. Design of Experiments:

1. Problem based on Missing Data Analysis- one and two observations in RBD
2. Problem based on Missing Data Analysis- one and two observations in LSD
3. Problem based on BIBD
4. Problem based on PIBD
5. Problem based on MOLS
6. Problem Based on Split Plot Design

4. Survival Analysis:

1. To identify the type of censoring and to estimate survival time for type I and type II censored data

2. Estimation of Median Survival time from raw data
3. Estimation of mean survival time and variance of the estimator for type I and type II censored data
4. To estimate the survival function and variance of the estimator using Non-parametric methods with Reduced Sample method and actuarial method
5. To estimate the survival function and variance of the estimator using Non-parametric methods with Kaplan-Meier method
6. Estimation of hazard function using Non-parametric method
7. Comparison between two survival functions using Gehan's tests and Mantel-Haenszel tests
8. To estimate Survival function using partial likelihood method
9. To estimate Survival function under frailty model, AFT model, and competing risk model.

5. Statistical Genetics and Ecology

1. Problems on hidden Markov models and parameter estimation techniques.
2. Problems on fitting growth models to real and simulated data under deterministic and stochastic setup.
3. Problems on goodness of fit for growth curves for standard growth curves such as exponential, logistic, Gompertz.

6. Clinical Trials and Epidemiology

1. Designing a clinical trial under parallel design set-up, cross-over design set-up with, cross-sectional design set-up, and longitudinal design set-up with real life background
2. Estimation of bias and random errors for clinical trials
3. Case studies
4. Analysis of categorical outcomes from Phase I – III trials
5. Analysis of survival data from clinical trials.
6. Meta Analysis of summary measures under Fixed Effects Model
7. Meta Analysis of summary measures under Random Effects Model
8. Evaluation of heterogeneity
9. Estimation of small study bias

10. Estimation of risk ratio for epidemiological studies

Statistical Lab-III

1. Sample Survey:

1. Selecting a sample from a given sampling design, calculation inclusion probabilities and construction of Horvitz-Thompson estimator.
2. Selecting a sample under equal probability sampling and estimation of parameters.
3. Problem based on PPSWR and PPSWOR.
4. Problem based on stratified sampling and cluster sampling.
5. Problem based on uses of auxiliary variables.
6. Problem based on randomized response technique.

2. Stochastic Processes:

1. Simulation of a Discrete Time Markov chain using transition probability matrix
2. calculation of m-step transition probabilities from 1-step transition probability
3. Estimation of 1-step and m-step transition probabilities from the realization of a Markov chain
4. Finding stationary distribution.

3. Design of Experiments:

1. Problem based on Missing Data Analysis- one and two observations in RBD
2. Problem based on Missing Data Analysis- one and two observations in LSD
3. Problem based on BIBD
4. Problem based on PIBD
5. Problem based on MOLS
6. Problem Based on Split Plot Design

4. Statistical Learning with Big Data-I:

1. Fitting linear models with real and simulated data and checking for significance of regression etc.
2. Understanding multicollinearity problems with simulated data and to deal with multicollinearity.

3. Problems related to K-Nearest Neighbor Regression, Comparison of Linear Regression with K-Nearest Neighbor.
4. Classification Problems with logistic regression, LDA, QDA and K-Nearest Neighbor and their comparison.
5. Application of cross-validation on Regression and Classification problems.
6. Problems related to shrinkage methods such as Ridge Regression and Lasso.
7. Problems on Principal component regression and partial least squares.

5. Advanced Data Analytic Techniques:

1. Estimation using Jackknife method and Bootstrap method
2. Impute missing observations using EM algorithm and MCEM algorithm
3. Imputation of missing observations under MCAR, MAR and MNAR setup
4. Missing value imputation using Mean Imputation method, and Hot & Cold deck method
5. Estimation of model parameters using REML method for longitudinal data
6. Estimation of subject specific effects for longitudinal data
7. Estimation of regression parameters for models with dichotomous responses
8. Estimation of regression parameters for models with general count responses
9. Identification of overdispersion
10. Estimation using Quasi likelihood function

6. Statistical Learning with Big Data-II:

1. Problems related to polynomial regression, local regression, regression splines, generalized additive models etc.
2. Problems related to tree based methods on regression and classification.
3. Problems on support vector machine (SVM) for classification
4. Problems based on neural networks.
5. Problems related to unsupervised learning methods: principal components analysis for data visualization or data pre-processing, and clustering.

Statistical Lab-IV

1. Sample Survey:

1. Selecting a sample from a given sampling design, calculation inclusion probabilities and construction of Horvitz-Thompson estimator.
2. Selecting a sample under equal probability sampling and estimation of parameters.
3. Problem based on PPSWR and PPSWOR.
4. Problem based on stratified sampling and cluster sampling.
5. Problem based on uses of auxiliary variables.
6. Problem based on randomized response technique.

2. Stochastic Processes:

1. Simulation of a Discrete Time Markov chain using transition probability matrix
2. calculation of m-step transition probabilities from 1-step transition probability
3. Estimation of 1-step and m-step transition probabilities from the realization of a Markov chain
4. Finding stationary distribution.

3. Design of Experiments:

1. Problem based on Missing Data Analysis- one and two observations in RBD
2. Problem based on Missing Data Analysis- one and two observations in LSD
3. Problem based on BIBD
4. Problem based on PIBD
5. Problem based on MOLS
6. Problem Based on Split Plot Design

Practicals from the following papers which were taken by a student as departmental elective.

4. Reliability Theory:

1. Fitting of life-testing data using different probability models (exponential, Weibull, Gamma, log-normal etc.).
2. Reliability estimation and fitting of life-testing data under different censoring schemes (Type-I, Type-II, Hybrid, Progressive).
3. Estimation and plot of Kaplan – Meier reliability curve.
4. Kolmogorov-Smirnov test and comparison of Empirical and Fitted CDF.
5. Estimation and Fitting of competing risk models.
6. Fitting of Accelerated failure time model.

or

4. Actuarial Statistics:

1. Problem based on collective risk model
2. Problem based on individual risk model
3. Problem based on aggregate claims distribution.
4. Problem based on credibility theory.
5. Problems to calculate reserves using run off triangles.

5. Statistical Process and Quality Control:

1. Construction of OC and ARL curves for control charts.
2. Construction of moving average and exponentially weighted moving average control charts.
3. Double sampling inspection plan: construction and interpretation of OC, ASN, AOQ, ATI.
4. Sequential sampling inspection plan: construction and interpretation of OC, ASN, AOQ, ATI.
5. Computation and interpretation of capability indices Cp, Cpk and Cpm.

Generic Electives:

STAPGGEC01 Statistical Methods using R

Course Objectives: This course is intended for the PG students from non-Statistics background. The course will involve only the concepts and uses of theories rather than rigorous derivations of the results.

Learning Outcomes: The students will be able to use statistical methods such as descriptive statistics, correlation, regression, and hypothesis formulation and testing in their field of study.

Detailed Syllabus:

Study design. Graphical representation of data. Features of frequency distribution, Measures of central tendency, dispersion, skewness and kurtosis for the study of nature of data. Problems with outliers and extremes. (10)

Idea of correlation and regression for two and three variables; correlation coefficient, correlation ratio, multiple and partial correlations. (8)

Probability. Basic results. Conditional probability and Bayes theorem. Random variables- expectation and variance. Probability models for discrete and continuous variables. Computation of probability in various applied research. (10)

Basics of Statistical inference. Estimation and Hypothesis testing problems in special setups. Applications of statistical inference in applied research. (14)

Introduction to R programming. Statistical Computations using R (8)

References:

1. Goon, A. M., Gupta, M. K. and Dasgupta, B. - Fundamentals of Statistics, Vols 1 & 2
2. Goon, A. M., Gupta, M. K. and Dasgupta, B.- Outlines of Statistical Theory, Vols 1 & 2
3. Mood, A.M., Graybill, F.A. and Boes, D.C. (2007): Introduction to the Theory of
4. Statistics, 3rd Edn. (Reprint), Tata McGraw-Hill Pub. Co. Ltd.
5. Ross, S. (2002): A First Course in Probability, Prentice Hall.
6. Rohatgi, V. K. and Saleh, A.K. Md. E. (2009): An Introduction to Probability and
7. Statistics. 2nd Edn. (Reprint) John Wiley and Sons.
8. Snedecor G.W and Cochran W.G. (1967) Statistical Methods. Iowa State Univ. Press.
9. Casella, G. and Berger R.L. (2002).: Statistical Inference, 2 nd ed. Thomson Learning.
10. A.Agresti (1984): Analysis of Ordinal Categorical Data

Teaching Plan:

Week 1: Study design. Graphical representation of data. Features of frequency distribution

Week 2-3: Measures of central tendency, dispersion, skewness and kurtosis for the study of nature of data. Problems with outliers and extremes.

Week 4-5: Idea of correlation and regression for two and three variables; correlation coefficient, correlation ratio, multiple and partial correlations.

Week 6-7: Probability. Basic results. Conditional probability and Bayes theorem. Random variables-expectation and variance.

Week 8-9: Probability models for discrete and continuous variables. Computation of probability in various applied research.

Week 10-12: Basics of Statistical inference. Estimation and Hypothesis testing problems in special setups. Applications of statistical inference in applied research.

Week 13-14: Introduction to R programming. Statistical Computations using R

STAPGGEC02 Survey Sampling and Experimental Designs

Course Objectives: This course is intended for the PG students from non-Statistics background. The course will involve only the concepts and uses of theories rather than rigorous derivations of the results. In this course students will learn about various sampling techniques and experimental designs.

Learning Outcomes: The students will be able to use statistical methods such as survey methodology and experimental designs in their field of study.

Detailed Syllabus:

Basic concepts of sampling from a finite population; sampling versus complete enumeration; simple random sampling; sample size determination; stratified random sampling; systematic sampling; cluster sampling and multi – stage sampling (all sampling schemes without proof of expressions). (14)

Analysis of variance techniques : One way and two way classified data. (4)

Design of experiments : Randomization, replication, local control; completely randomized design; randomized block design and Latin square design; factorial experiments. (14)

Data analysis : The students will be trained to use Statistical softwares/packages like Excel/SPSS/STATA/R or any other for data analysis. The main focus of the training will also include the use of parametric and non – parametric tests and the interpretation of the results. (18)

References:

1. Cochran, W.G. (1977) : Sampling Techniques, 3 rd Edition, Wiley.
2. Des Raj (2000) : Sample Survey Theory, Narosa Publishing House
3. Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984): Sampling Theory of Surveys with Applications, Iowa State University Press and Indian Society of Agricultural Statistics.
4. Das, M.N. and Giri, N (1986) : Design and Analysis of Experiments, Springer Verlag.
5. Goon, A.M., Gupta, M.K. and Das Gupta, B. (1991) : Fundamentals of Statistics, Vol. II, World Press, Calcutta
6. Gibbons, J.D. (1985) : Non – Parametric Statistical Inference, 2nd Edition, Marcel Dekker, Inc.
7. Rohatgi, V.K. (1988) : An Introduction to Probability and Mathematical Statistics, Wiley Eastern, New Delhi.
8. Siegel, S. : Non – Parametric Statistics for the Behavioural Sciences,
9. Mood, A.M., Greybill, F.A. and Boes, D.C. (1974) : Introduction to the Theory of Statistics, McGraw Hill

Teaching Plan:

Week 1-4: Basic concepts of sampling from a finite population; sampling versus complete enumeration; simple random sampling; sample size determination; stratified random sampling; systematic sampling; cluster sampling and multi – stage sampling (all sampling schemes without proof of expressions).

Week 5: Analysis of variance techniques : One way and two way classified data.

Week 6-9: Design of experiments : Randomization, replication, local control; completely randomized design; randomized block design and Latin square design; factorial experiments.

Week 10-14: Data analysis : The students will be trained to use Statistical softwares/packages like Excel/SPSS/STATA/R or any other for data analysis. The main focus of the training will also include the use of parametric and non – parametric tests and the interpretation of the results.

STAPGGEC02 Probability and Stochastic Processes

Course Objectives: This module aims to provide various ideas about probability theory and stochastic processes.

Learning Outcomes: The outcomes of the course are as follows:

1. Ability to understand and solve problems involving probability.
2. Understand, formulate and analyze stochastic processes.

Detailed Syllabus:

Review of Sample space, discrete probability, independent events, Bayes theorem. Random variables and distribution functions (univariate and multivariate); expectation and moments. Independent random variables, marginal and conditional distributions. Characteristic functions. (15)

Probability inequalities (Tchebyshef, Markov, Jensen). (4)

Modes of convergence, weak and strong laws of large numbers, Central Limit theorems (i.i.d. case). (6)

Introduction to Stochastic Processes, Classification and example, Markov and non-Markov Process, Markov chains with finite and countable state space, Examples, transition probabilities, classification of states, Periodicity of a Markov chain, Random walk on Integers, restricted random walk, Limiting behaviour of n-step transition probabilities, stationary distribution (25)

References:

1. S. M. Ross : A First Course in Probability
2. Chung, K.L. (1983): Elementary Probability Theory with Stochastic Process, Springer / Narosa.
3. Feller, W. (1968): An Introduction to Probability Theory & its Applications, John Wiley.
4. Goon, A.M., Gupta, M.K. & Dasgupta, B. (1994): An Outline of Statistical Theory (Vol-1), World Press.
5. S. M. Ross : Introduction to Probability Models
6. S. M. Ross : Stochastic Process
7. David F. Anderson: Introduction to Stochastic Processes with Applications in the Biosciences
8. S. Karlin & H.M. Taylor : A First Course in Stochastic Processes
9. J. Medhi : Stochastic Process
10. A.K. Basu : Stochastic Process

11. R.N. Bhattacharyya & E. Waymire : Stochastic Processes and Applications

Teaching Plan:

Week 1-2: Review of Sample space, discrete probability, independent events, Bayes theorem. Random variables and distribution functions (univariate and multivariate)

Week 3-5: expectation and moments. Independent random variables, marginal and conditional distributions. Characteristic functions.

Week 6: Probability inequalities (Tchebyshef, Markov, Jensen).

Week 7-8: Modes of convergence, weak and strong laws of large numbers, Central Limit theorems (i.i.d. case).

Week 9: Introduction to Stochastic Processes, Classification and example, Difference between stochastic and deterministic processes, Markov and non-Markov Process, time homo-geneous and time inhomogeneous processes with examples.

Week 10-12: Markov chains with finite and countable state space, Examples, transition probabilities, classification of states, Periodicity of a Markov chain, Random walk on Integers, restricted random walk

Week 13-14: Limiting behaviour of n-step transition probabilities, stationary distribution