INDIAN STATISTICAL INSTITUTE

STUDENTS' BROCHURE

Effective from the Academic Year 2014 -15

POST- GRADUATE DIPLOMA
IN
STATISTICAL METHODS AND ANALYTICS

NORTH-EAST CENTRE
TEZPUR UNIVERSITY CAMPUS, SAIC BUILDING
TEZPUR, ASSAM – 784028

POST-GRADUATE DIPLOMA IN STATISTICAL METHODS AND ANALYTICS

Contents

Post- Graduate Diploma in Statistical Methods and Analytics

1. Curriculum

SEMESTER I
1. Basic Mathematics
2. Probability Theory
3. Statistical Methods
4. Numerical Methods and Optimization
5. Introduction to Packages: R, S and SAS

SEMESTER II
1. Computer Intensive Statistical methods
2. Regression & Time Series
3. Statistical Machine Learning & Statistical Finance
4. Clinical Trials & Actuarial Methods
5. Project

1.1. Brief Syllabi of Courses

Semester I

1. Basic Mathematics
Calculus
Set theory: sets, set operations, functions, equivalence of sets, finite and infinite sets, countable and uncountable sets with examples (2)
Real numbers: field properties and order properties, representation as points on real line, sup and inf, completeness, rationals and irrationals and their properties, intervals (2)
Sequences and Series: limits of sequences, properties, sandwich theorem, bounded and monotone sequences, subsequences, Cauchy criterion (statement only), convergence of series, tests of convergence (Standard tests like comparison, ratio, root tests etc) (6)
Functions: limits and continuity of functions, right and left limits, simple properties (sum, difference, product, composition, etc.), differentiability and simple properties, chain rule, monotonicity and convexity of functions, mean-value theorem (statement only, geometric interpretation of the theorem), maxima-minima, Taylor theorem (statement only) (10)
Integration: Sketch of the idea (without complete details) of Riemann integration, fundamental theorem of calculus (statement only), properties of integral, change of variable (4)
2-variable calculus: continuity, partial derivatives, double integrals, iterated integration, Jacobian rule, differentiation under integration (statement only) (6)

References:
1. Calculus (Vol I & II) - Apostol, T.
2. Introduction to Real Analysis - Bartle, R.G. and Sherbert, D.R.
3. Introduction to Calculus and Analysis (Vol I & II) – Courant, R. and John, F.
4. Principles of Real Analysis - Rudin, W.

Linear Algebra

Introduction to matrices:System of linear equations, matrix representation, basic matrix operations (2)
Vector spaces: Definition and examples, subspaces, linear independence, basis of a vector space (2)
Matrix theory:matrices as linear transformation, elementary operations and elementary matrices, rank, nullity, trace, inverse and determinants of matrices, solutions of system of linear equations (12)
Spectral theory: eigenvalues and eigenvectors of matrices, decomposition of matrices, quadratic forms and definiteness of a matrix (with applications in Statistics) (6)

References:

1. Matrix Theory and Linear Algebra – Hernstein, I.N. and Winter, D.J.
2. Matrix Algebra – Gentle, J. E.
3. Matrix Computations – Golub, G.H. and Van Loan, C.F.
4. Introduction to Linear Algebra – Mirsky, L.

2. Probability Theory

Elementary concepts of probability: experiments, outcomes, sample space, events. (8)
Conditional probability, independence, Bayes theorem. (6)
Random variable, probability distribution and properties; probability mass/density function, cumulative distribution function, expectation, variance, moments. (8)
Binomial, Poisson, Negative Binomial, Hypergeometric, Uniform, Normal and Exponential distributions. (8)
Chebyshev's inequality, weak law of large numbers, central limit theorem (statement). (2)
Distribution of a function of a random variable. (4)
Bivariate distribution; joint, marginal and conditional distributions, moments, covariance, correlation coefficient. (8)
Independent random variables and their sums. Transformation of two random variables. (6)
Sampling distributions: chi-square, t, F. (4)

References:

1. A First Course in Probability - Ross, S.
2. Elementary Probability Theory – Chung, K. L.
3. Introduction to Probability – Roussas, G.
4. Probability- Pitman, J.

3. Statistical Methods

Different types of statistical problems and related data analysis.(2)
Concept of population, sample and statistical inference through examples. (2)
Summarization of univariate data; graphical methods, measures of location, spread, skewness and kurtosis; outliers and robust measures. (14)
Empirical distribution, extension to censored data: Kaplan-Meier estimate. (5)
Analysis of discrete and continuous data, fitting probability distributions, goodness of fit, graphical methods of verifying the fit. (10)
Concept of estimation (point and interval) with examples. Concept of testing of hypotheses; significance level, size, power and p-value. (12)

One and two sample t-tests, paired t-test, nonparametric tests. One and two sample tests for proportions.(12)

References:
1. Introductory Statistics ¬ Ross, S
2. Statistics – Freedman, D., Pisani R. and Purves, R.
3. Applied General Statistics – Croxton, F.E. and Cowden, D. J.
4. Statistics: A Guide to the Unknown –Tanur, J.M. (ed.).
5. Statistics: A New Approach- Wallis, W.A. and Roberts, H.V.

4. Numerical Methods and Optimization

Numerical Methods

Significant digits, round-off errors. Finite computational processes and computational errors. Loss of significant digits. (4)
Solution of nonlinear equation in one variable. Separation of roots and initial approximation. (4)
Improvement of the initial solution using methods of bisection, Regula Falsi and Newton-Raphson.(10)
Fixed point iterative schemes. Errors. Order of convergence and degree of precision.(6)

Optimization

Lagrange method of multipliers, maxima and minima of differentiable functions.(6)
Linear programming: simplex method, dual simplex method, sensitivity.(12)
Unconstrained optimization: Newton, Quasi-Newton method. (8)
Computational methods of optimization. (6)

References:
1. Numerical Analysis for Statisticians– Lange, K.
2. Elementary Numerical Analysis: An Algorithmic Approach- Conte, S.D. and de Boor, C.
3. Operations Research: An Introduction – Taha, H. A.
4. Optimization – Lange, K.

5. Introduction to Packages: R, S and SAS

Introduction to packages: overview of packages, data handling, input-output operations. (10)
Basic programming: data types, arrays, loops etc.; functions and graphics.(10)
Introduction to SAS programming. (10)
Statistical computations - data summary and graphical display of data, basic statistics. (8)
Simulations from probability distributions, comparisons of distributions, Q-Q and P-P plots.(10)
Matrix computations - basic operations, finding determinant, inverse, eigen roots and eigen vectors of a matrix, matrix decomposition, solving system of equations. (8)

References:
1. A Handbook of Statistical Analysis using R – Everitt, B. S. and Hothorn, T.
2. A Handbook of Statistical Analysis using SAS – Der, Geoff, Everitt, B. S.
3. Modern Applied Statistics with S-PLUS ¬ Venables, W.N. and Ripley, B. D.
4. Numerical Analysis for Statisticians– Lange, K.

Semester II

1. Computer Intensive Statistical Methods

Statistical inference - likelihood based, Bayesian.(10)
Categorical data analysis: contingency tables, measures of association, test of independence.(4)
Principal component analysis. (3)
Simulation: acceptance/rejection sampling; importance sampling. (6)
Introduction to discrete time Markov chains, finite state space and countable state space. Markov chain Monte Carlo (MCMC) methods and simulation of Markov chains, applications in statistics of the MCMC methods. (20)
Histogram and kernel smoothing; density estimation; nonparametric regression. (9)
Bootstrap and resampling. (6)
Illustration of the methodology with real data.

References:
1. Computational Statistics – Gentle, J. E.
2. Computational Statistics – Givens, G.H. and Hoeting, J. A.
3. Statistical Computing – Kennedy, W.J. and Gentle, J.E.
4. Handbook of Computational Statistics – Gentle, J.E., Hardle, W. and Mori, Y.
5. Statistical Computing: Existing Methods & Recent Developments – Basu,A. and Kundu,D.
6. Resampling Methods: A Practical Guide to Data Analysis – Good, P.I.
7. Simulation and Monte Carlo Method - Rubinstein, R.Y.
8. Smoothing Methods in Statistics – Simonoff, J.S.

2. Regression & Time Series

Regression

Classical Linear Regression Model (2).OLS method of estimation; tests of hypotheses (6)
Use of dummy variables in regression (1); residuals and fitted values (3).Variable selection. (3)
Validation of assumptions using graphical techniques. (7)
Logistic regression; odds ratio, concordance-discordance measures. (7)
Illustration of the methodology with real data.

References:
1. Introduction to Linear Regression Analysis – Montgomery, D. C., Peck, E. and Vinning, G.
2. Regression Analysis by Examples - Chatterjee S. and Hadi, G.
3. Applied Linear Regression – Weisberg, S.
4. Applied Regression Analysis – Draper, N.R. and Smith, H.
5. Applied Logistic Regresson – Hosmer, D.W. and Lemeshow, S.

Time series

Exploratory analysis and graphical display; trend, seasonal and cyclical components. Smoothing: exponential and MA. (6)
Stationary Time Series: AR, MA and ARMA models; Box-Jenkins correlogram analysis, ACF and PACF, choice of AR and MA orders. (10)
Non-Stationary Time Series: introduction to ARIMA model; deterministic and stochastic trends; introduction to ARCH models. (6)
Forecasting: basic tools, using exponential smoothing and Box-Jenkins method. Residual analysis. (6)
Illustration of the methodology with real data.

References:
1. Introduction to Time Series and Forecasting – Brockwell, P. and Davis R. A.

2. Analysis of Time Series – Chatfield, C.
3. Time Series Analysis and Its Applications with R – Shumway, R.H. and Stoffer, D.S.
4. Intro. to Time Series Analysis & Forecasting – Montgomery, D.C.,Jennings,C.L., Kulachi, M.
5. Forecasting: Methods and Applications – Makridakis, S.G., Wheelwright, S.C. and Hyndman, R.J.

3. Statistical Machine Learning & Statistical Finance

Statistical Machine Learning

Unsupervised learning: clustering procedures (hierarchical and non-hierarchical); association rules. (6)
Supervised learning: Linear discriminant analysis; Bayesian classifier, nearest neighbor classifier. Tree based classification methods; predictive modeling using decision trees. Entropy based classifier. (12)
Support vector machine. Boosting and adaptive boosting algorithm. (6)
Assessment and model selection: bias-variance trade off, training error rate, criteria of selection (AIC, BIC), cross validation. (4)
Applications in information retrieval and text analysis. Illustration of the methodology with real data.

References:
1. The Elements of Statistical Learning: data Mining, Inference and Prediction – Hastie, T. Tibshirani, J.H. and Friedman, J.H.
2. Data Mining: Concepts and Techniques – Han, J. and Kamber, M.
3. Machine Learning – Mitchell, T.M.
4. Statistical and Machine-Learning Data Mining – Ratner, B.
5. Classification and Regression Trees – Breiman, L. et al

Statistical Finance

Derivatives: forward and future contracts. Markets, prices, arbitrage. Complete market, market risk and credit risks in the use of derivatives. (4)
Options markets, properties of stock option prices; American and European options. (4)
Binomial model: One-step and two-step models; Risk neutral valuation. (4)
Volatility; value at risk. (4)
Behaviour of stock prices: Conditional expectation and properties. (6)
Options on stock indices; currencies and futures; Some exotic equity and foreign exchange derivatives; Interest rate derivatives.(8)
Illustration of the methodology with real data.

References:
1. Options, Futures and other derivatives – Hull, John
2. Financial Calculus - Baxter, M. and Rennie, A.
3. Risk-Neutral Valuation - Bingham, N. and Keisel, R.

4. Clinical Trials & Actuarial Methods

Clinical Trials

Introduction to clinical trials; bias and random error in clinical studies; conduct of clinical trials, selection of subjects, ethical issues, outcome measures, protocols. (6)
Different Phases; comparative and controlled trials; random allocation. (4)
Design of clinical trials: parallel group designs; crossover designs; symmetric designs; adaptive designs; group sequential designs. (8)
Design of phase I, II and III trials. (4)
Bioequivalence trials. (3)
Power and sample size determination. (3)

Illustration of the methodology with real data.

References:
1. Clinical Trials: A Practical Approach – Pocock, S.
2. Fundamentals of Clinical Trials – Friedman, L.M, Furburg, C. and Demets, D.L.
3. Clinical Trials: A Methodological Perspective – Piantadosi, S
4. The Design and Analysis of Sequential Clinical Trials – Whitehead, J

Actuarial Methods
General Insurance: Loss models; parametric estimation. (3) Re-insurance and deductibles. (2)
Collective and individual risk models for aggregate loss. (4) No Claims Discount systems (3) Ruin theory (statement of the problem) (2)
Life Insurance: Introduction to survival analysis (1).
Complete and curtate future lives; force of mortality and hazard rate (2).
Life tables (3). Present values of insurances and annuities (6). Premium (2).
Illustration of the methodology with real data.

References:
1. Statistical and Probabilistic Methods in Actuarial Science – Boland, P.J.
2. Loss Models: From Data to Decisions – Klugman, S.A, Panier, H.H., Wilmot, G. E
3. Life Contingencies – Spurgeon, E.T.

5. Project