Institute of Statistical Research and Training
University of Dhaka

Curriculum and Syllabus

M.S. Program in Applied Statistics

Session : 2017–2018

www.isrt.ac.bd/academics/graduate

Institute of Statistical Research and Training

The Institute of Statistical Research and Training (ISRT), University of Dhaka, is the leading institution for training and research in Applied Statistics in Bangladesh. It was founded in 1964 by the Late National Professor Dr. Qazi Motahar Husain, an eminent scientist, academician and a leading proponent of the statistical sciences in this country. The Institute offers a 4-year B.S. (Honours) program that has been designed to produce graduates with strong statistical computing skills, sound knowledge of statistical concepts and the versatility to apply these concepts in areas as diverse as medicine, engineering, economics and the social sciences. The 1-year M.S. program consists of specialized courses in areas ranging from environmental statistics to clinical trials, statistical machine learning and meta analysis and has been designed for students with a keen interest in higher studies and research. In addition, the Institute offers Ph.D. and M.Phil. degree programs. A number of highly experienced faculty members with masters and Ph.D. degrees from reputed universities across the world run these programs.

ISRT boasts an academic environment that is highly competitive and conducive to research. Both students and faculty members benefit from the regular seminars and talks given by researchers from home and abroad on topics of current interest. The Institute has a rich library with well over 15,000 books and is equipped with three state-of-the-art computer labs, cloud computing facilities and high-speed internet access for graduate and undergraduate students. The aim is to provide a learning environment that stimulates intellectual curiosity, critical thinking and independent problem-solving skills. The Journal of Statistical Research (JSR), an international journal, has been published bi-annually by ISRT since 1970, and it serves as a forum for the exchange of research ideas between statisticians in Bangladesh and abroad. Faculty members conduct research in diverse areas such as biostatistics, spatial statistics, statistical pattern recognition, Bayesian analysis and econometrics, and regularly publish in the peer-reviewed journals.

Among its other activities, the Institute frequently organizes short courses and training programs for non-statisticians in government and non-government organizations who find themselves using statistics in their work. In doing so, it has played an active role in promoting and creating awareness about the need for sound statistical practices among the people from other disciplines so that they may work more efficiently within their organizations. ISRT also maintains close ties with the Bangladesh Bureau of Statistics and other organizations responsible for the collection and dissemination of statistical data in Bangladesh and is frequently called upon to offer its expertise on statistical issues of national interests. The Institute has been played a significant role in the countrys development by producing world class statisticians for academia and industry. In addition to that, the Institute provides statistical consulting service through *StatLab* primarily for the students and faculty members of the University of Dhaka, with the aim to strengthen research on campus by assisting graduate students and faculty members of other disciplines.

M.S. Program in Applied Statistics

The Master of Science (M.S.) program in Applied Statistics is a one-year program. The minimum requirement for the admission to this program is the successful completion of the B.S. Honours degree in Applied Statistics from ISRT. The regulations for the admission and the examinations will be the same as those of the M.S. courses in the Faculty of Science unless otherwise stated. The program includes courses on advanced topics in statistics and computing with special emphasize on the applications of the advanced statistical techniques to real life situations. The objective of the program is to produce graduates with high statistics and computing skill so that, after successful completion, they are equipped to work efficiently and completely in government and non-government organizations, research organizations, service departments and other related fields.

Structure of the Program

There are two types of course designs available for the M.S. program in Applied Statistics:

Group A	:	M.S. degree based only on course work.
Group B	:	M.S. degree based on course work and thesis.

Total credit hours for both the Group A and B is 30. Students of both groups must take 19-credit hours of theoretical courses of which 4-credit hours are compulsory and 15-credit hours are elective. For the elective part, students can choose five 3-credit hours courses from the list of optional courses. The choice of optional courses will depend on the availability of teaching faculties of the institute. In addition, there will be a two credit hours oral comprehensive course. The remaining credit hours are distributed as follows:

- Group **A**

  Students from Group A are required to take three statistical computing courses (AST 530, AST 531, and AST 532) and prepare either a project report or a report from internship (AST 550). The computing courses AST 530 and AST 531 are of 2-credit hours and the comprehensive computing course AST 532 is of 3-credit hours. The project report (AST 550) will carry 2-credit hours of which 40% weight will be allotted for presentation, 10% weight for supervisor and the remaining 50% weight will be allotted for report.

- Group **B**

  A selected number of students will be considered for Group B who are required to submit a thesis and defend it (AST 551). The course AST 551 will carry 6-credit hours of which 40% weight will be for thesis presentation and 60% weight for thesis. Students of Group B must take the comprehensive statistical computing course (AST 532), which will carry 3-credit hours. It is expected that all thesis students actively participate in seminars organized by the institute during the academic year.

The duration of M. S. program will be of 1 (one) academic year to be distributed as follows:

Classes	26 weeks
Time for preparation of final examination	04 weeks
Course final examination	04 weeks
Submission of thesis/project/internship	14 weeks
Publication of results	04 weeks

The credit is defined as follows:

(i) For theoretical courses, 15 class hour of 50 minutes each = 1 credit.

(ii) For computing courses, 15 class hour of 50 minutes each for lab work + 15 hours for practices = 1 credit.

Assessment System

Evaluation: The performance of a student in a given course will be evaluated by in-course examinations/assignments/performance evaluation in the class/final examinations. Thirty percent marks of the theoretical courses and forty percent marks of the computing courses will be allotted for in-course examination.

The marks allocation for theoretical and computing courses will be as follows:

Theoretical		
Attendance	:	05
In-course exam	:	25
Final exam	:	70

Computing		
Attendance/assignment	:	10
In-course exam	:	30
Final exam	:	60

There will be two in-course examinations for each of the theoretical and computing courses. Students in in-course may be evaluated by giving short questions as decided by the course teacher. Each in-course assessment will be of one-hour duration for a theory course and the average of marks from two exams will be considered as the final mark. However, the duration of in-course is 1.5 hours for a computing course and the sum of two marks will be taken as the final mark.

Each M.S. student (Group A and Group B) will be required to give at least one seminar during the academic year. It is a non-credit course but compulsory. The grade to be assigned will be “Satisfactory” or “Not-Satisfactory”. The internal members of the examination committee will evaluate the performance in the seminars.

The duration of theoretical course final examinations will be as follows:

Credit	Duration of Examination
4	4 hours
3	3 hours

The duration of practical course final examinations will be of 4 hours.

Marks distribution for attendance:

Attendance (%)	Marks (%)
90 and above	05
85 to 89	04
80 to 84	03
75 to 79	02
60 to 74	01
Less than 60	00

Grading and Grade Point: Grades and grade points will be awarded on the basis of marks obtained in the written, oral and practical examinations according to the following scheme:

Marks Obtained (%)	Grade	Grade Point
80-100	A+	4.00
75-79	A	3.75
70-74	A-	3.50
65-69	B+	3.25
60-64	B	3.00
55-59	B-	2.75
50-54	C+	2.50
45-49	C	2.25
40-44	D	2.00
less than 40	F	0.00
	I	Incomplete
	W	Withdrawn

Only 'D' or higher grade will be counted as credits earned by a student. Grade point average (GPA) will be calculated as the weighted average of the grade points obtained by a student in all the courses completed in a year. GPA will be calculated according to the following formula:

$$\text{GPA} = \frac{\sum(\text{grade points obtained in a course} \times \text{total credit for that course})}{\text{total credits taken at a given year}}$$

Requirements for the Award of the M.S. Degree:

1. Minimum number of required credits must be earned in the maximum one year period.
2. Must have GPA of at least 2.5.
3. A student obtaining 'F' grade in any courses will not be awarded degree. Student with 'F' grade in only ONE course shall be allowed to retake either within 3 months of publication of the results after paying special fees set by the university or with the following batches. However, student with 'F' grade in MORE THAN ONE courses is required to take re-admission in the following year.

Some policies about the examination system are given below:

1. **In-course Examination**

   (a) No make-up test will be arranged for a student who fails to appear in in-course test/tests. Absence in any in-course test will be counted as zero for calculating the average in in-course test for that course. However, a student can apply to the Director if recommended by the respective course teacher. The Director will only place the application before the academic committee if the particular student has met with an accident or her/his parents have expired or s/he has gone through a surgical procedure or any other such situation which the Academic Committee feels can be considered. The make-up test must be held during the course period.

   (b) Course teachers must announce results in 4 weeks of holding the examination.

   (c) Marks for in-course assessment must be submitted by the course teacher to the Chairman of the Examination Committee and the Controller of Examinations before the final examination.

   (d) Questions for in-course examinations may preferably be a multiple choice (MCQ) type. Students may also be evaluated by giving short questions as decided by the course teacher.

2. **Final Examination**

   (a) The year final examinations will be conducted centrally by the Controller of Examinations as per existing rules.

   (b) Student having 75% or more attendance on average (collegiate) are eligible to appear in the final examination.

   (c) Student having 60-74% attendance are considered to be non- collegiate and will be eligible to sit for the final examination on payment of fine set by the university.

   (d) Student having attendance less than 60% will not be allowed to sit for the final examination but may seek readmission in the program.

   (e) At the beginning of each academic session, an examination committee is to be constituted for that session by the academic committee of the institute. The Chairman of the Examination Committee will act as a course co-cordinator for that session. The examination committee will have a Chairman, two internal members and an external member.

   (f) For theoretical course final examinations, there will be two examiners: course teacher will be the first examiner and the second examiner will be from within the department or from any other department of Dhaka University relevant to the subject.

   (g) Third Examination: Under double-examiner system and in case of difference of more than 20% of marks, there will be a 3rd examiner. Marks of nearest two examiners (theory and thesis) will be averaged out as final marks.

3. **Time Limits for Completion of Master's Degree**

   A student must complete the courses of her/his studies for a M.S. degree in a maximum period of 2 (two) academic years.

4. **Improvement**

   (a) If a student obtains a grade 'C+' or lower in a course, s/he will be allowed to repeat the term-final examination only once with the following batch for the purpose of grade improvement. But s/he will not be eligible to get a grade better than 'B+' in such a course. A student failing to improve her/his grade in a course can retain the earlier grade.

   (b) A student obtaining 'F' grade in one or more courses (theory and practical) will not be awarded degree. However, a student obtaining 'F' grade in a course may be allowed to retake that course only once with the next batch of students in order to be awarded a degree. A student obtaining 'F' grades in more than one courses will not be allowed to repeat any course.

5. **Readmission**

   (a) A student failing to complete the M.S. course in a year may seek readmission with the next available batch of students, provided s/he applies within one month of publication of the result of the concerned year.

   (b) A readmitted student will be allowed to retain her/his in- course/class assessment/tutorial marks earned in previous year.

   (c) A readmitted student may be allowed to take up thesis work as decided by the institute's Academic Committee.

   (d) The transcripts of successful readmitted student will bear the letter 'R' after GPA with a foot note explaining 'R' means Readmission.

6. **Other General Regulations**

   For any matter not covered in the above guidelines, existing rules for Integrated Honours Course of Dhaka University will be applicable.

Structure of the Courses

Distribution of courses, credits, marks and detailed syllabus are as follows:

Courses for Group A

Courses	Credit Hour
Compulsory Courses	
Theoretical Courses	4
Statistical Computing Courses	7
M.S. Project or Internship	2
Oral	2
Seminar	Non-credit
Elective Courses	
Theoretical Courses	15
Total	30

Compulsory Courses for Group A

Course ID	Course Title	Credit Hour
AST 501	Applied Bayesian Statistics	4
AST 530	Statistical Computing I	2
AST 531	Statistical Computing II	2
AST 532	Comprehensive Statistical Computing	3
AST 540	Oral	2
AST 545	Seminar	Non-credit
AST 550	M.S. Project/Internship	2
Total		15

Courses for Group B

Courses	Credit Hour
Compulsory Courses for the Group B	
Theoretical Courses	4
Statistical Computing Courses	3
M.S. Thesis	6
Oral	2
Seminar	Non-credit
Elective Courses	
Theoretical Courses	15
Total	30

Compulsory Courses for Group B

Course ID	Course Title	Credit Hour
AST 501	Applied Bayesian Statistics	4
AST 532	Comprehensive Statistical Computing	3
AST 540	Oral	2
AST 545	Seminar	Non-credit
AST 551	MS Thesis	6
Total		15

Elective Courses for Groups A and B

Course ID	Course Title	Credit Hour
AST 510	Advanced Survival Analysis	3
AST 511	Environmental and Spatial Statistics	3
AST 512	Advanced Time Series Analysis	3
AST 513	Actuarial Techniques	3
AST 514	Advanced Operations Research	3
AST 515	Advanced Econometric Methods	3
AST 516	Advanced Population Studies	3
AST 517	Queueing Theory and Stochastic Processes	3
AST 518	Advanced Multivariate Techniques	3
AST 519	Analysis of Longitudinal Data	3
AST 520	Adaptive Sampling	3
AST 521	Optimum Experimental Designs	3
AST 522	Statistical Signal Processing	3
AST 523	Meta Analysis	3
AST 524	Clinical Trials	3
AST 525	Statistical Machine Learning	3

Detailed Syllabus

AST 501: Applied Bayesian Statistics	Credit 4

Introduction

Applied Bayesian statistics refers to practical inferential methods that use probability models for both observable and unobservable quantities. The flexibility and generality of these methods allow them to address complex real-life problems that are not amenable to other techniques. This course will provide a pragmatic introduction to Bayesian data analysis and its powerful applications.

Objectives

Acquire basic understanding in the principles and techniques of Bayesian data analysis. Apply Bayesian methodology to solve real-life problems. Utilize R for Bayesian computation, visualization, and analysis of data.

Contents

Bayesian thinking: background, benefits and implementations; Bayes theorem, components of Bayes theorem - likelihood, prior and posterior; informative and non-informative priors; proper and improper priors; discrete priors; conjugate priors; semi-conjugate priors; exponential families and conjugate priors; credible interval; Bayesian hypothesis testing; building a predictive model.

Bayesian inference and prediction: single parameter models - binomial model, Poisson model, normal with known variance, normal with known mean; multi-parameter models - concepts of nuisance parameters, normal model with a non-informative, conjugate, and semi-conjugate priors, multinomial model with Dirichlet prior, multivariate normal model; posterior inference for arbitrary functions; methods of prior specification; method of evaluating Bayes estimator.

Summarizing posterior distributions: introduction; approximate methods: numerical integration method, Bayesian central limit theorem; simulation method: direct sampling and rejection sampling, importance sampling; Markov Chain Monte Carlo (MCMC) methods - Gibbs sampler, general properties of the Gibbs sampler, Metropolis algorithm, Metropolis-Hastings (MH) sampling, relationship between Gibbs and MH sampling, MCMC diagnostics - assessing convergence, acceptance rates of the MH algorithm, autocorrelation; evaluating fitted model - sampling from predictive distributions, posterior predictive model checking.

Linear model: introduction, classical and Bayesian inference and prediction in the linear models, hierarchical linear models - Bayesian inference and prediction, empirical Bayes estimation; generalized linear model - Bayesian inference and prediction (logit model, probit model, count data model); model selection - Bayesian model comparison.

Nonparametric and Semiparametric Bayesian models.

Text Books

1. Hoff PD (2009). A First Course in Bayesian Statistical Methods. Springer.
2. Gelman A, Carlin JB and Stern HS, Dunson DB, Vehtari A, and Rubin DB (2013). Bayesian Data Analysis, $3^{rd} edition$. Chapman and Hall.
3. Gill J (2007). Bayesian Methods: A Social and Behavioral Sciences Approach, $2^{nd} edition$. Chapman and Hall.

AST 510: Advanced Survival Analysis | **Credit 3**

Estimating the Survival and Hazard Functions: Introduction and notation, the Nelson-Aalen and Kaplan-Meier estimators, counting process and martingals, properties of Nelson-Aalen estimator.

Semiparametric Multiplicative Hazards Regression Model: Introduction, estimation of parameters, inclusion of strata, handling ties, sample size determinations, counting process form of a Cox model, time-dependent covariates, different types of residuals for Cox models, checking proportionality assumption.

Multiple Modes of Failure: Basic characteristics of model specification, likelihood function formulation, nonparametric methods, parametric methods, semiparametric methods for multiplicative hazards model.

Analysis of Correlated Lifetime Data: Introduction, regression models for correlated lifetime data, representation and estimation of bivariate survivor function.

Text Books

1. Therneau TM and Grambsch PM (2000). Modeling Survival Data: Extending the Cox Model, Springer.
2. Kalbfleisch JD and Prentice RL (2002). The Statistical Analysis of Failure Time Data, 2^{nd} *edition.* Wiley.
3. Hougaard P (2000). Analysis of Multivariate Survival Data. Springer.

AST 511: Environmental and Spatial Statistics | Credit 3

Introduction

Spatial statistics encompasses diversified statistical methods for analyzing data obtained from stochastic process indexed by the space. This branch is enrich enough to gain insight from data exploiting the dependence over space. Its myriad applications caught profound attention of people from both academia and practitioners.

Objectives

Technology is indispensable for modern life, and its advances in different aspects of our life made several things possible. Now a days data have been collected along with extensive additional information. Spatial data is one of such examples. In recent years, analysis of spatial data receives great attention over the world. As a result, several theories have been developed for different types of spatial data analysis. This course is designed to introduce the graduate student with few of such theories so that they can develop their skill in spatial data analysis. To comprehend this course, students need a sound knowledge of Mathematical statistics, particularly the concepts of stochastic process. It is expected that the student will be able to analyze different spatial data from diverse fields after successful completion of the course.

Contents

Review of non-spatial statistics and stochastic process, overview of different types of spatial data; random field and spatial process - geostatistical/point reference process, areal/lattice process and point process; spatial data concern.

Geostatistical data: real data examples, measure of spatial dependence- variogram and covariance, stationarity and isotropic, variograms and covariance functions, fitting the variograms functions; Kriging, linear geostatistical model - formulation, simulation, estimation and prediction, generalized linear geostatistical model - formulation, simulations, estimation and prediction. Areal data: neighborhoods, testing for spatial association, autoregressive models (CAR, SAR), estimation/inference; grids and image analysis, disease mapping. Point pattern data: locations of events versus counts of events, types of spatial patterns, CSR and tests - quadrat and nearest neighbor methods, K-functions and L-functions, point process models- estimation and inference, health event clustering.

Special topics in spatial modeling: Hierarchical models, Bayesian methods for spatial statistics, Bayesian disease mapping, Spatio-temporal modeling, more on stationarity. Use of R and GIS software to give emphasis on analysis of real data from the environmental, geological and agricultural sciences.

Text Books

1. Cressie N (1993). Statistics for Spatial Data, *Revised edition.* Wiley.

2. Banerjee S, Carlin BP, and Gelfand AE (2014). Hierarchical Modelling and Analysis for Spatial Data.$2^{nd} edition$ Chapman and Hall.
3. Cressie N and Wikle CK (2011). Statistics for Spatio-Temporal Data. Wiley.
4. Illian J, Penttine A, Stoyan H and Stoyan D (2008). Statistical Analysis and Modelling of Spatial Point Patterns. Wiley.

AST 512: Advanced Time Series Analysis — Credit 3

Introduction: Forecasting time series, estimation of transfer functions, stochastic and deterministic dynamic mathematical models, stationary and nonstationary stochastic models for forecasting and control, basic ideas in model building. Time series and stochastic processes, stationary stochastic processes.

Seasonal models: parsimonious models for seasonal time series, fitting versus forecasting, seasonal models involving adaptive sines and cosines, general multiplicative seasonal model, some aspects of more general seasonal ARIMA models, structural component models and deterministic seasonal components.

Nonlinear and long memory models: Autoregressive conditional heteroscedastic (ARCH) models, generalized ARCH (GARCH) models, model building and parameter estimation, nonlinear time series models, long memory time series processes.

Multivariate time series analysis: Stationary multivariate time series, vector autoregressive-moving average (ARMA) models and representations, relation of vector ARMA to transfer function and ARMAX model forms, forecasting for vector autoregressive-moving average processes.

Text Books

1. Box GEP, Jenkins GM and Reinsel GC (2008). Time Series Analysis: Forecasting and Control, 4^{th} *edition.* Wiley.
2. Brockwell PJ and Davis RA (2013). Introduction to Time Series and Forecasting, 2^{nd} *edition.* Springer.

AST 513: Actuarial Techniques

Credit 3

Theory of interest in continuous time. Forces of interest and discount (constant and varying). Present and accumulated value calculations using non-level interest rates. Continuous annuities, valuation of continuous streams of payment, including the case in which interest conversion period differs from the payment period, continuous varying annuities. Bonds and related securities.

Definition and application of standard mortality probability symbols and force of mortality; relationship between survival distribution and life table functions; Continuous life annuities. Multiple decrement models. Net premiums, fully continuous premiums. Net premium reserves. Valuation theory for pension plans. The expanse factor and dividends. Introduction to risk theory– Purpose of the theory of risk; main problems in risk theory; individual risk models for a short term; applications of risk theory.

Principles of actuarial modeling. Familiarity with actuarial models– survival models, credibility models, risk theory models, ruin theory models, etc and their applications.

Text Book

1. Kellison SG (1991). Theory of Interest. Irwin.

AST 514: Advanced Operations Research

Credit 3

Special Types of Liner Programming Problems: Transportation problem, Transshipment problem, Assignment problem, Multidimensional problems.

Network Analysis: Terminology of networks, shortest path problem, minimum spanning tree problem, maximum flow problem, minimum cost flow pattern, network simplex method, project planning and control with PERT-CPM.

Dynamic Programming: Characteristics of dynamic programming problems, deterministic dynamic programming, probabilistic dynamic programming.

Non-linear Programming: Sample application, Graphical illustration of non-linear programming problems, types of non-liners programming problems, one-variable unconstrained optimization, multivariate unconstrained optimization, Karush-Kuhn Tucker (KKT) conditions for constrained optimization, quadratic programming, separable programming, convex programming, non-convex programming.

Inventory Models: The ABC inventory system, a generalized inventory model, deterministic models, probabilistic models, just-in-time manufacturing system.

Text Book

1. Hillier FS, Lieberman GJ, Nag B and Basu P (2001). Introduction to Operations Research, 9^{th} *edition.* McGraw-Hill.

AST 515: Advanced Econometric Methods **Credit 3**

Three-stage least squares estimation: The three-stage least squares estimator (3SLS), comparison between GMM 3SLS and traditional 3SLS.

Linear unobserved effects panel data models: Strict exogeneity assumptions on the explanatory variables, some examples of unobserved effects panel data models, estimating unobserved effects models by pooled OLS, random effects (RE) methods, estimation and inference under the basic random effects, a general FGLS analysis, fixed effects (FE) methods, consistency of the fixed effects estimator, asymptotic inference with fixed effects. The Hausman test comparing the RE and FE estimators.

Nonlinear models: Discrete response models, the linear probability model for binary response, probit and logit, maximum likelihood estimation of binary response models, specification issues in binary response models, neglected heterogeneity, continuous endogenous explanatory variables, a binary endogenous explanatory variable, heteroskedasticity and nonnormality in the latent variable model, estimation under weaker assumptions, binary response models for panel data, pooled probit and logit.

Multinomial response models: Multinomial Logit, probabilistic choice models.

Ordered response models: Ordered logit and ordered probit, applying ordered probit to interval-coded data, corner solution outcomes and censored regression models– derivations of expected values, inconsistency of OLS, estimation and inference with censored tobit, pooled tobit, applying censored regression to panel data.

Text Books

1. Wooldridge JM (2010). Introductory Econometrics: A Modern Approach, 5^{th} *edition.* Cengage Learning.
2. Greene WH (2011). Econometric Analysis, 7^{th} *edition.* Prentice Hall.

AST 516: Advanced Population Studies — Credit 3

Fertility Determinants and models. Davis-Blake intermediate variables. Proximate determinants. Bongaarts model. Birth interval dynamics. Fecundity. Application of these models in the Bangladesh setting.

Estimation of population parameters from incomplete data. Estimation of mortality from census. Survivorship ratio. Estimate of infant and child mortality by indirect techniques such as Brass, Sullivan, Trussell and Feeney. Estimation of adult mortality from information on widowhood and orphanhood. Estimation of fertility by indirect techniques such as Brass, Hill, Coale-Trussell, relational Gompertz and reduced Gompertz model.

Population and Development: Inter-relation between population and development as envisaged value. Various population theories such as demographic transition theory. Emerging theories of population. Micro-economic theory of population. Recent contribution of Eastisliu, Becker, Caldwell etc.

Morbidity: Morbidity differentials and trends in Bangladesh. Health expectancy and burden of disease.

Manpower planning: Manpower in Bangladesh. Factors effecting manpower. Working life tables. Statistical analysis of manpower planning. Labor migration and its impact on the economy.

Population program of Bangladesh: Population policy. Population control and family planning. Evaluation of family planning programs. Use effectiveness and cost effectiveness. Multiple decrement life tables for measuring use-effectiveness. FP target setting models and impact assessment models.

Text Books

1. Chiang CL (1984). The Life Table and Its Applications. Krueger.
2. Bongaarts J and Potter RG (1983). Fertility, Biology and Behavior: An Analysis of the Proximate Determinants of Fertility. Academic Press, New York.

AST 517: QUEUEING THEORY AND STOCHASTIC PROCESSES

Credit 3

Queueing theory: Classical M/M/1 queue, global and local balance, performance measures, Poisson arrivals see time averages (PASTA) property, M/M/1/S queueing systems, blocking probability, performance measures, multi-server systems M/M/m, performance measures, waiting time distribution of M/M/m, performance measures of M/M/m/m with finite customer population, Erlang loss systems, a more general queueing models: M/G/1, M/G/m, G/M/1 queueing systems and analysis.

Queueing networks: Open queueing networks, analysis of tandem queues, applications of tandem queues in data networks, Jackson queueing networks, performance measures for open networks, closed queueing networks: Jackson closed queueing networks, steady-state probability distribution, application of closed queueing networks.

Reliability theory: Structure functions, minimal path and minimal cut sets, reliability of systems of independent components, bounds on the reliability function, system life as a function of component lives, expected system lifetime, systems with repair.

Brownian motion and stationary processes: Brownian motion, hitting times, maximum variable, variations on Brownian motion, Brownian motion with drift, geometric Brownian motion, pricing stock options, white noise, Gaussian processes, stationary and weakly stationary processes, harmonic analysis of weakly stationary processes.

Text Books

1. Chee-Hock N and Boon-Hee S (2008). Queueing Modelling Fundamentals with Applications in Communication Networks. John Wiley & Sons, Chichester.
2. Ross SM (2010). Introduction to Probability Models, 10^{th} *edition.* Academic press, New York.

AST 518: Advanced Multivariate Techniques

Credit 3

Introduction

Multivariate data analysis contains prominent significance in statistics as data generated from various systems these days are usually of multivariate high dimensional. This course introduces several dimension reduction and classification methods to analyze multivariate data.

Objectives

To understand the core concept of multivariate data both mathematically and graphically. To learn the various dimension reduction and classification approaches with details. To analyze real life multivariate data using the learned methods.

Contents

Principal Components: Population principal components, summarizing sample variations by principal components, graphing the principal components, large sample inference.

Factor Analysis: The orthogonal factor models, methods of estimation (maximum likelihood estimates and principal factor analysis), selection of loadings and factor (factor rotation, varimax rotation, quartimax rotation, oblimin rotations), factor scores, structural equations models.

Canonical Correlation Analysis: Canonical variates and canonical correlations, sample canonical variates and sample canonical correlations, large sample inference.

Discrimination and Classification: Separation and classification two populations, classification of two multivariate normal populations, evaluating classification functions, Fisher's discriminant function, classification with several populations, Fisher's method for discriminating several populations.

Clustering: Similarity measures, hierarchical clustering methods, nonhierarchical clustering methods, multidimensional scaling.

Text Books

1. Johnson RA, Wichern W (2002). Applied Multivariate Statistical Analysis. Pearson.
2. Srivastava KS (2002). Methods of Multivariate Statistics. Wiley.

AST 519: Analysis of Longitudinal Data	Credit 3

Introduction

Longitudinal data arise when multiple measurements of a response are collected over time for each individual in the study and hence are likely to be correlated, which presents substantial challenge in analyzing such data. This course covers topics related to statistical methods and models for drawing scientific inferences from longitudinal data. Topics include longitudinal study design; exploring longitudinal data; repeated measures analysis of variance, linear and generalized linear regression models for correlated data, including marginal, random effects, transition models,; and handling conditional and joint models; missing data.

Objectives

To realize the unique features of and the methodological implication of analyzing the data from longitudinal studies, as compared to the data from traditional studies. To understand statistical methods/models, particularly linear/generalized linear mixed models, GEE approaches and joint models, for analyzing longitudinal data. To be familiar with the proper implementation and interpretation of the statistical methods/models for analyzing longitudinal data.

Contents

Longitudinal data: Concepts, examples, objectives of analysis, problems related to one sample and multiple samples, sources of correlation in longitudinal data, exploring longitudinal data.

Linear model for longitudinal data: Introduction, notation and distributional assumptions, simple descriptive methods of analysis, modelling the mean, modelling the covariance, estimation and statistical inference.

ANOVA for longitudinal data: Fundamental model, one sample model, sphericity condition; multiple samples models.

Linear mixed effects models: Introduction, random effects covariance structure, prediction of random effects, residual analysis and diagnostics.

Extension of GLM for longitudinal data: Review of univariate generalized linear models, quasi-likelihood, marginal models, random effects models, transition models, comparison between these approaches; the GEE methods: methodology, hypothesis tests using wald statistics, assessing model adequacy; GEE1 and GEE2.

Introduction to the concept of conditional models, joint models, their applications to bivariate binary and count data. Estimation, inference and test of independence.

Generalized Linear Mixed Models (GLMM): Introduction, estimation procedures–Laplace transformation, penalized quazi-likelihood (PQL), marginal quazi likelihood (MQL).

Numerical integration: Gaussian quadrature, adaptive gaussian quadrature, Monte Carlo integration; markov chain Monte Carlo sampling; comparison between these methods.

Statistical analysis with missing data: Missing data, missing data pattern, missing data mechanism, imputation procedures, mean imputation, hot deck imputation. estimation of sampling variance in the presence of non-response, likelihood based estimation and tests for both complete and incomplete cases, regression models with missing covariate values, applications for longitudinal data.

Text Books

1. Verbeke G and Molenberghs G (2000). Linear Mixed Model for Longitudinal Data. Springer.
2. Molenberghs G and Verbeke G (2005). Models for Discrete Longitudinal Data. New York: Springer-Verlag.
3. Islam MA and Chowdhury RI (2017). Analysis of Repeated Measures Data. Springer.
4. Diggle PJ, Heagerty P, Liang K-Y, and Zeger SL (2002). Analysis of Longitudinal Data, 2^{nd} *edition.* Oxford.

AST 520: Adaptive Sampling

Credit 3

Design and model unbiased estimators; fixed and stochastic population sampling theory.

Adaptive sampling deigns; Detectability in adaptive sampling; constant and unequal detectabilities for adaptive design.

Adaptive cluster sampling; initial random sample with and without replacement; initial unequal probability sampling; expected sample size and cost; comparative efficiencies of adaptive and conventional sampling.

Systematic and strip adaptive cluster sampling; stratified adaptive cluster sampling; adaptive allocation in stratified sampling; sample sizes based on observations in each strata and from previous strata; comparison of systematic and stratified adaptive sampling with conventional sampling procedures; adaptive cluster sampling based on order statistics.

Multivariate aspects of adaptive sampling; multivariate conditions for adding neighbourhoods; design-unbiased estimation for multivariate approach.

Text Books

1. Thompson SK and Seber GAF (1996). Adaptive Sampling. Wiley.
2. Thompson SK (1992). Sampling. Wiley.

AST 521: Optimum Experimental Designs **Credit 3**

Optimum design theory: Continuous and exact designs, the general equivalence theorem, algorithms for continuous designs and general equivalence theorem, function optimization and continuous design.

Criteria of optimality: A-, D-, and E-optimality; D_A-optimality, D_S-optimality, c-optimality, linear optimality; compund design criteria.

D-optimum designs: Properties of D-optimum designs, sequential construction of optimum designs, polynomial regression in variable, second-order model with several variables.

Algorithms for constructing of exact D-optimum designs: The exact design problem, basic formulae for exchange algorithm, sequential algorithms, non-sequential algorithms, the KL and BLKL exchange algorithms.

Experiments with both qualitative and quantitative factors, blocking response surface designs, mixture experiments, non-linear models, Bayesian optimum designs, model checking and designs for discriminating between models, compund design criteria, generalized linear models.

Text Book

1. Atkinson AC, Donev AN, and Tobias RD (2008). Optimum Experimental Designs with SAS. Oxford.

AST 522: Statistical Signal Processing — Credit 3

Introduction to signals: Signals and their classification; real world analog signals: audio, video, biomedical (EEG, ECG, MRI, PET, CT, US), SAR, microarray, etc; digital representation of analog signals; role of transformation in signal processing. Orthogonal representation of signals. Review of exponential Fourier series and its properties.

Signal estimation theory: Estimation of signal parameters using ML, EM algorithm, minimum variance unbiased estimators (Rao-Blackwell theorem, CRLB, BLUE), Bayesian estimators (MAP, MMSE, MAE), linear Bayesian estimators.

Signal detection theory: Detection of DC signals in Gaussian noise: detection criteria (Bayes risk, Probability of error, Neyman-Pearson), LRT; detection of known signals in Gaussian noise: matched filter and its performance, minimum distance receiver; detection of random signals in Gaussian noise: the estimator correlator.

Applications: Scalar quantization, image compression, pattern recognition, histogram equalization, segmentation, application of signal estimation and detection theory to signal communication, signal recovery from various types of linear and nonlinear degradations, copyright protection, enhancement, etc.

Text Books

1. Soliman SS and Mandyam DS (1998). Continuous and Discrete Signals and Systems, 2^{nd} *edition.* Prentice-Hall.
2. Kay SM (1993). Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice-Hall.
3. Kay SM (1998). Fundamentals of Statistical Signal Processing: Detection Theory. Prentice Hall.
4. Gonzalez RC and Woods RE (2008). Digital Image Processing, 3^{rd} *edition.* Pearson Education, Inc.

AST 523: Meta analysis — Credit 3

Introduction to systematic review and meta analysis: Motivation, strengths and weakness of meta-analysis, problem formulation (why study meta analysis), systematic review process.

Types of results to summarize; overview of effect size; effect size calculation for both continuous and discrete data.

Combining effect size from multiple studies; fixed effect and random effects models and their estimation; heterogeneity between studies and its estimation techniques; test of homogeneity in meta analysis; prediction intervals; subgroup analysis, Meta regression: random effect meta regression, baseline risk regression.

Publication bias in meta analysis; Power analysis for meta analysis; effect size rather than P-values; Meta analysis based on direction and P-values, reporting the results of meta analysis.

Introduction to Bayesian approach to meta analysis; Meta analysis for multivariate/longitudinal data; network meta analysis.

Text Books

1. Borenstein M, Hedges LV, Higgins JPT and Rothstein HR (2009). Introduction to Meta-Analysis, John Wiley & Sons, UK.
2. Hartung J and Knapp G and Sinha BK (2011). Statistical Meta-Analysis with Applications. John Wiley & Sons, UK.

AST 524: Clinical Trials

Credit 3

Introduction

The course is designed to give an overall idea of clinical trial studies. It will provide an introduction to the statistical and ethical aspects of clinical trials research.

Objectives

Topics include design, implementation, and analysis of trials, including first-in- human studies, phase II and phase III studies. The course will enable applying existing methodologies in designing clinical trials and will also foster research in this area.

Contents

Statistical approaches for clinical trials: Introduction, comparison between Bayesian and frequentist approaches and adaptivity in clinical trials. Phases of clinical trials, pharmacokinetics (PK) and pharmacodynamics (PD) of a drug, dose-concentration-effect relationship and compartmental models in pharmacokinetic studies.

Phase I studies: Determining the starting dose from preclinical studies. Rule-based designs: 3+3 design, Storer's up-and-down designs, pharmacologically-guided dose escalation and design using isotonic regression. Model-based designs: continual reassessment method and its variations, escalation with overdose control and PK guided designs.

Phase II studies: Gehan and Simon's two-stage designs. Seamless phase I/II clinical trials: TriCRM, EffTox and penalised D-optimum designs for optimum dose selection.

Phase III studies: Randomised controlled clinical trial, group sequential design and multi-arm multi-stage trials in connection with confirmatory studies.

Text Books

1. Berry SM, Carlin BP, Lee JJ, and Muller P (2010). Bayesian Adaptive Methods for Clinical Trials. CRC press.
2. Rosenbaum SE (2012). Basic Pharmacokinetics and Pharmacodynamics: An Integrated Textbook and Computer Simulations. John Wiley & Sons.

AST 525: Statistical Machine Learning | Credit 3

Introduction

The course provides a broad but thorough introduction to the methods and practice of statistical machine learning and its core models and algorithms.

Objectives

The aim of the course is to provide students of statistics with detailed knowledge of how Machine Learning methods work and how statistical models can be brought to bear in computer systems not only to analyze large data sets, but also to let computers perform tasks, that traditional methods of computer science are unable to address.

Contents

Statistical learning: Statistical learning and regression, curse of dimensionality and parametric models, assessing model accuracy and bias-variance trade-off, classification problems and K-nearest neighbors.

Linear regression: Model selection and qualitative predictors, interactions and nonlinearity.

Classification: Introduction to classification, logistic regression and maximum likelihood, multivariate logistic regression and confounding, case-control sampling and multiclass logistic regressionl, linear discriminant analysis and Bayes theorem, univariate linear discriminant analysis, multivariate linear discriminant analysis and ROC curves, quadratic discriminant analysis and naive bayes.

Resampling methods: Estimating prediction error and validation set approach, k-fold cross-validation, cross-validation- the right and wrong ways, the bootstrap, more on the bootstrap.

Linear model selection and regularization: Linear model selection and best subset selection, forward stepwise selection, backward stepwise selection, estimating test error using mallow's Cp, AIC, BIC, adjusted R-squared, estimating test error using cross-validation, shrinkage methods and ridge regression, the Lasso, the elastic net, tuning parameter selection for ridge regression and lasso, dimension reduction, principal components regression and partial least squares.

Moving beyond linearity: Polynomial regression and step functions, piecewise polynomials and splines, smoothing splines, local regression and generalized additive models.

Tree-based methods: Decision trees, pruning a decision tree, classification trees and comparison with linear models, bootstrap aggregation (Bagging) and random forests, boosting and variable importance.

Support vector machines: Maximal margin classifier, support vector classifier, kernels and support vector machines, example and comparison with logistic regression.

Text Books

1. James G, Witten D, Hastie T and Tibshirani R (2013). An Introduction to Statistical Learning: with Applications in R, $1^{st}edition$. Springer.
2. Hastie T, Tibshirani R and Friedman J (2009). The Elements of Statistical Learning: Data Mining, Inference and Prediction, $2^{nd}edition$. Springer.

AST 530: Statistical Computing I

Credit 2

Introduction

This course focuses the application of Bayesian statistics in real life situations. It will also enable students to apply machine learning algorithms to manage large datasets originated from various sources.

Objectives

The successful completion of the course will help a student to apply these techniques in various situations.

Contents

Computing problems related to AST 501: Applied Bayesian Statistics, and AST 525: Statistical Machine Learning.

AST 531: Statistical Computing II

Credit 2

Introduction

The course will show application of various advanced multivariate techniques to real life situations. Here students will also learn how to analyze longitudinal data.

Objectives

The course intends to train students with modern tools of analyzing multivariate and longitudinal data.

Contents

Computing problems related to AST 518: Applied Multivariate Techniques, and AST 519: Analysis of Longitudinal Data.

AST 532: Comprehensive Statistical Computing

Credit 3

Introduction

Comprehensive statistical computing deals with useful advanced statistical computations.

Objectives

This course contains statistical computations related to all courses offered in the academic year.

Contents

This is a comprehensive statistical computing course, which is compulsory for all the M.S. students. It will cover computing problems related to all the courses studied in the academic year. For this course, 30% weight will be alloted to the in-course examinations, 10% weight will be allotted to the lab-based assignments, and remaining 60% weight for the final examination. All the examinations will be held in computer lab.

AST 540: Oral

Credit 2

Each student (Group A and Group B) must be examined orally by a committee of selected members at the end of the academic year.

AST 545: Seminar

Non-Credit

The internal members of the examination committee will evaluate the performance in the seminars.

AST 550: M.S. Project/Internship

Credit 2

Each student should be either in the project report group or in the internship group that the student will decide after discussing with the respective assigned supervisor. Students must submit their project report or internship report within 2 months of completing the final examination. The internal members of the examination committee will evaluate the performance of the students in the seminars and the project report or internship report will be evaluated by two examiners nominated by the examination committee. A supervisor cannot evaluate the project or internship report that s/he has supervised. This course will carry 2 credit hours. For this course, 50% weight of the course will be allotted to report, 10% weight for supervisor and the remaining 40% weight will be for seminar presentation.

AST 551: M.S. Thesis

Credit 6

After three months of the start, each student of Group B will submit a three to five page thesis proposal and present her/his proposal, which will be evaluated by the internal members of the examination committee. Written evaluation on the proposal will be provided to the students explaining the possible improvement and in case of "Not satisfactory" proposals, the reasons for "Not satisfactory" performance will be stated in the written evaluation. In case of "Not satisfactory" performance the examination committee may give the student a second opportunity for proposal presentation. Students with "Not-Satisfactory" performance at the end will be transferred to the Group A. The final submission of the thesis will be required within 4 months of the completion of final exam. Thesis will be examined by two external (outside the institute) examiners. Should it be required, the examination committee may consider one internal and one external examiner. Submitted thesis has to be defended at a presentation evaluated by the members of the examination committee. Forty percent (40%) weight will be allotted for thesis defense and remaining 60% weight for thesis itself.
