Institute of Statistical Research and Training
University of Dhaka

Curriculum and Syllabus

M.S. Program in Applied Statistics and Data Science

Session: 2022–2023

www.isrt.ac.bd/academics/graduate

Contents

1 Institute of Statistical Research and Training

1.1 Introduction

The Institute of Statistical Research and Training (ISRT), University of Dhaka, is the leading institution for training and research in Applied Statistics and Data Science 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 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, business, economics, and the social sciences. The 1-year M.S. program consists of specialized courses in areas ranging from environmental statistics to statistical signal processing, statistical machine learning, and causal inference 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. Highly experienced faculty members, most of whom have 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. It has 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 published bi-annually by ISRT since 1970, is a forum for exchanging research ideas between statisticians in Bangladesh and abroad. Faculty members have research interests in diverse areas such as biostatistics, machine learning, spatial statistics, statistical pattern recognition, public health, Bayesian analysis, and econometrics. They regularly disseminate their research works in peer-reviewed journals and international conferences.

Among its other activities, the Institute frequently organizes short courses and training programs for non-statisticians working in government and non-government organizations who need statistical analysis. In doing so, it has played an active role in promoting and creating awareness about the need for sound statistical practices among 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 (BBS) and other organizations responsible for collecting and disseminating statistical data in Bangladesh. It is frequently called upon to offer its expertise on statistical issues of national interest. Over the years, the Institute has played a significant role in the country's development by producing world-class statisticians for academia and industry and providing statistical expertise on issues of national interest. In addition, the Institute provides statistical consulting services through StatLab primarily for the students and faculty members of the University of Dhaka, aiming to strengthen research on campus by assisting graduate students and faculty members of other disciplines.

1.2 Vision and Mission of the Institute

Vision

The vision of the Institute is to take a leading position globally in providing quality education in Applied Statistics and Data Science, conducting leading-edge research, and creating innovative industrial partnerships.

Mission

The Institute's mission is to produce competent graduates in Applied Statistics and Data Science equipped with the skills necessary for success in a technological society and competitive global environment who will fulfill the statistical demands of the nation and the world.

Objectives

To fulfil the vision and missions, ISRT aims to

(i) Strengthen and update various teaching and training programs at undergraduate, postgraduate, and doctoral levels to produce graduates with strong theoretical and practical knowledge of statistical science in line with the labor market requirements.

(ii) Create an environment conducive to high quality research.

(iii) Contribute to advancing science and technology through interdisciplinary research, jointly with scientists, scholars at the University of Dhaka, and other research institutions at home and abroad.

(iv) Contribute to the statistics profession and the larger scientific community by running quality statistical journals and serving on editorial boards, review panels, and administrative and advisory committees.

(v) Employ high quality faculty members with diverse research interests.

(vi) Promote the exchange of knowledge and ideas by arranging invited talks regularly in addition to workshops and international conferences.

(vii) Disseminate statistical knowledge by offering training programs to students of other departments and professionals of various government and private organizations.

(viii) Serve the University's and national bodies' statistical needs by providing consulting services in research, government, business, and industry.

(ix) Produce graduates having strong moral and ethical values, respect for local norms and culture, and exceptional leadership qualities.

2 M.S. Program in Applied Statistics and Data Science

2.1 Introduction

The Master of Science (M.S.) program in Applied Statistics and Data Science is a one-year program. The minimum requirement for admission to this program is successfully completing the B.S. Honours degree in Applied Statistics from ISRT. Unless otherwise stated, the regulations for the admission and the examinations will be the same as those of the M.S. courses in the Faculty of Science.

The program includes courses on advanced topics in statistics and computing with special emphasis on applying advanced statistical techniques to real-life situations. The program aims to produce graduates with high statistics and computing skills 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.

2.2 Structure of the Program

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

Group A :	M.S. degree based on course work and project/internship.
Group B :	M.S. degree based on course work and thesis.

Total credit hours are 34 and 37 for Group A and B, respectively. All students must take 21-credit hours of theoretical courses, of which 3-credit hours are for the compulsory course and 18-credit hours are for elective courses. For the elective part, students can choose six 3-credit hours courses from the list of elective courses. The selection of elective courses will depend on the availability of the teaching faculties of the Institute. In addition, there will be a oral comprehensive course of 2-credit hours. The remaining credit hours are distributed for the students of two groups as follows:

Group A

A selected number of students will be considered for Group B, who must submit and defend a thesis (AST551). The M.S. thesis course AST551 will carry 6 credit hours, of which 40%

will be for thesis presentation and 60% will be for the thesis report. Group B students must also take four statistical computing courses (CSE530, AST531, AST532, and AST533), each with 2-credit hours.

All students are expected to actively participate in seminars organized by the Institute during the academic year.

Table 2.1: Distribution of 1 academic year by different components of the program

Type	Duration (in weeks)
Classes	1–26
Preparation of final examination	27–30
Course final examination	31–34
Submission of thesis/project/internship	35–48
Result publication	49–52

One credit for the M.S. program in Applied Statistics and Data Science is defined differently for theoretical and computing courses. For theoretical courses, 1 credit corresponds to 15 class hours, and each class is of 50 minutes. For computing courses, 1 credit corresponds to 15 class hours of 50 minutes each for lab work and 15 hours for practice.

2.3 Assessment System

2.3.1 Evaluation

A student's performance in a given course will be evaluated by in-course examinations (or assignments or continuous evaluation) in the class/final examinations. Thirty percent of the theoretical courses and forty percent of the computing courses will be allotted for in-course examination.

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

Table 2.2: Marks (%) allocation for theoretical and computing courses

Theoretical			Computing		
Attendance	:	05	Attendance/assignment	:	10
In-course exam	:	25	In-course exam	:	30
Final exam	:	70	Final exam	:	60

There will be two in-course examinations for each theoretical and computing course. Students in the in-course may be evaluated by giving short questions as the course teacher decides. 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 the in-course is 1.5 hours for a computing course, and the sum of two marks will be taken as the final mark.

The theoretical course final examinations will be 4 hours for 4-credit courses and 3 hours for 3-credit hour courses. The duration of the final examinations of computing courses will be of

4 hours.

Table 2.3: Marks distribution for attendance

Attendance (%)	Marks (%)
90 and above	5
85 to 89	4
80 to 84	3
75 to 79	2
60 to 74	1
<60	0

2.3.2 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:

Table 2.4: Percentage score, letter grade, and grade points

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
<40	F	0.00
	I	Incomplete
	W	Withdrawn

Only "D" or higher grade will be counted as credits a student earns. 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:

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

2.3.3 Minimum 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 course will not be awarded degree. Students with "F" grades in only ONE course shall be allowed to retake either within three months of publication of the results after paying special fees set by the university or with the subsequent batches. However, a student with a "F" grade in MORE THAN ONE course must take re-admission in the following year.

Policies about the examination system

1. **In-course Examination**
   - (a) No make-up test will be arranged for students who fail to appear in in-course test/tests. Absence in any in-course test will be counted as zero for calculating the average in the 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, her/his parents have expired, or s/he has undergone a surgical procedure, or any other such situation that 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.

2.4 List of Courses for Groups A and B

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

Table 2.5: Course structure for Groups A and B

		Group	
Courses	Credit	A	B
Compulsory Courses			
Theoretical Courses	3	✓	✓
Statistical Computing Courses	8	✓	✓
M.S. Project or Internship	3	✓	–
M.S. Thesis	6	–	✓
Oral	2	✓	✓
Seminar	Non-credit	✓	✓
Elective Courses			
Theoretical Courses	18	✓	✓
Total		34	37

Table 2.6: List of Compulsory Courses for Groups A and B

			Group	
Course ID	Course Title	Credit	A	B
AST501	Bayesian Statistics	3	✓	✓
CSE530	Python for Data Science	2	✓	✓
AST531	Statistical Computing I	2	✓	✓
AST532	Statistical Computing II	2	✓	✓
AST533	Statistical Computing III	2	✓	✓
AST540	Oral	2	✓	✓
AST545	Seminar	Non-credit	✓	✓
AST550	M.S. Project/Internship	3	✓	–
AST551	M.S. Thesis	6	–	✓
Total			16	19

Table 2.7: List of Elective Courses for Groups A and B

Course ID	Course Title	Credit
AST510	Advanced Survival Analysis	3
AST511	Environmental and Spatial Statistics	3
AST518	Introduction to Causal Inference	3
AST519	Analysis of Longitudinal Data	3
AST522	Statistical Signal Processing	3
AST523	Meta Analysis	3
AST524	Clinical Trials	3
AST525	Statistical Machine Learning	3
CSE526	Big Data Analytics	3
AST527	Advanced Statistical Machine Learning	3

Students of Groups A and B should select five courses from the list of elective courses

2.5 Sustainable Development Goals (SDGs) and M.S. Program in Applied Statistics and Data Science

The Sustainable Development Goals (SDGs) were adopted by all United Nations member states in 2015 as a universal call for action to end poverty, protect the planet, and ensure that all people enjoy peace and prosperity by 2030. There are 17 key SDGs, which have been designed to bring the world to several life-changing 'zeros', including zero poverty, hunger, AIDS, and discriminations against women and girls. In Bangladesh, Applied Statistics and Data Science graduates can contribute to achieve SDGs potentially through their acquired knowledge during their future employment. This is because the scientific knowledge of statistics is welcome by all spheres of development issues particularly in policy making, implementation, monitoring and evaluation. Therefore, it is essential to mark the SDGs indicators in the syllabus for M.S. in Applied Statistics and Data Science so that pertinent course instructor(s) can emphasize on relevant topic(s) for the sake of better understanding of the issues by the learners.

Generally, SDG relevant statistics are recorded, updated, monitored, and evaluated as official statistics by different organs of the government. Principally, government agencies under different ministries are in charge of implementing relevant interventions for achieving different SDGs targets, and Bangladesh Bureau of Statistics (BBS) leads the monitoring of the progress towards meeting the targets through conducting surveys and/or using official statistics. However, as an educational institution Institute of Statistical Research and Training (ISRT) can equip Applied Statistics and Data Science graduates with important statistical and computing skills so that they can work for the government and non-government agencies and help to achieve and monitor the SDGs in future.

Applied Statistics and Data Science graduates have already achieved skills to compute and analyze data for monitoring and evaluating SDG indicators from different courses of the B.S. Honors in Applied Statistics and Data Science program. In addition to those, the M.S. in Applied Statistics and Data Science program has been designed to equip students with skills on more advanced statistical methods covering topics related to small area estimation and area mapping (AST511), panel or longitudinal data analysis (AST519), causal inference (AST518),

Table 2.8: Connections between SDGs and Courses of M.S. Program in Applied Statistics and Data Science

SDGs	Keywords	Relevant Courses
SDG 1 : End poverty in all its forms everywhere	measuring poverty, zero poverty, poverty line, extreme poverty	511, 518, 550
SDG 2 : End hunger, achieve food security and improved nutrition and promote sustainable agriculture	prevalence of malnutrition among under five children	501, 511, 519, 550
SDG 3 : Ensure healthy lives and promote well-being for all at all ages	reduce neonatal mortality, under five mortality, maternal mortality, death rate	501, 511, 518, 519, 524, 550
SDG 4 : Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all ages	enrollment and dropout rate, participation rate in formal and non-formal education, rate of ICT learning	511, 518, 525, 526
SDG 5 : Achieve gender equality and empower all women and girls	women empowerment, domestic violence, teen marriage	550, 551
SDG 6 : Ensure availability and sustainable management of water and sanitation for all	access to safe drinking water, improved sanitation, hygiene practice	501, 511, 518, 519
SDG 7 : Ensure access to affordable, reliable, sustainable and modern energy for all	access to electricity, population with primary reliance on clean fuels and technology	501, 519, 525, 526
SDG 8 : Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all	GDP, unemployment rate, child labour	511, 550, 551
SDG 10 : Reduce inequality within and among countries	income inequality, poverty line, mapping poverty	519, 550, 551
SDG 13 : Take urgent action to combat climate change and its impacts	climate change, natural disaster	501, 511, 525, 526

statistical signal processing (AST522), statistical machine learning for predictive modeling (AST525), and big data analytics (CSE526). Besides these, Bayesian methods (AST501) are also useful for students to develop skills for estimating statistical models in some complex data conditions. In addition, students prepare MS projects or thesis reports (AST550) on the topics related to SDG indicators. All computing modules taught in MS Applied Statistics and Data Science namely CSE530 amd AST531-533 have been designed to acquire advanced statistical data analysis skills.

The detailed connectivity among SDG indicators and Applied Statistics and Data Science courses have been portrayed in Table 1. Course instructors of M.S. in Applied Statistics and Data Science are recommended to point to the key words (Table 1) picked up from SDG in their teaching module(s) wherever appropriate and emphasize relevant tools for teaching the techniques to estimate, test, and evaluate the SDG indicators thereby complying with national goals in line with the targets set in SDGs.

Overall, the M.S. in Applied Statistics and Data Science program is designed to produce graduates who will be able to provide advanced analysis of SDG data that will help to identify the bottlenecks of achieving SDGs and guide the policy makers for towards forming novel policies for accelerating the progress in achieving SDG targets, subject to the employment of Applied Statistics graduates in the pertinent government organs as well as development organizations in Bangladesh.

3 Detailed Syllabus

AST501: Bayesian Statistics	**Credit 3**

Introduction

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.

Learning Outcomes

Upon completion of the course, students will learn about (i) the concept of the Bayesian statistical methods, (ii) formulation and derivation of prior and posterior distributions, (iii) estimation of the model using different MCMC methods, and (iv) application of the Bayesian methods to analyze data and interpret and compare results with the frequentist approach.

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.

Textbooks

1. Hoff PD (2009). A First Course in Bayesian Statistical Methods. Springer.

Reference Books

1. Gelman A, Carlin JB and Stern HS, Dunson DB, Vehtari A, and Rubin DB (2013). Bayesian Data Analysis, *3rd edition*. Chapman and Hall.
2. Gill J (2007). Bayesian Methods: A Social and Behavioral Sciences Approach, *2nd edition*. Chapman and Hall.

AST510: Advanced Survival Analysis — Credit 3

Introduction

An introduction to methods of analysing correlated time-to-event data is provided in this course. Some commonly used methods for analysing univariate time-to-event data, e.g. Kaplan-Meier estimate of survivor functions, Cox's proportional hazards models, etc., are reviewed using counting processes notations.

Objectives

The objectives of the course are to teach the theoretical basis of different methods related to analysing correlated time-to-event data and competing risks model and to apply statistical softwares to analyse data using such models.

Learning Outcomes

At the end of the course, students are expected (i) to understand the theoretical basis of different methods related to analysing correlated time-to-event data and competing risks model (ii) to use a statistical software (e.g. related R packages) to analyse data using such models (iii) to interpret the results and write scientific publication.

Contents

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.

Textbooks

1. Therneau TM and Grambsch PM (2000). Modeling Survival Data: Extending the Cox Model, Springer.

Reference Books

1. Kalbfleisch JD and Prentice RL (2002). The Statistical Analysis of Failure Time Data, *2nd edition.* Wiley.
2. Hougaard P (2000). Analysis of Multivariate Survival Data. Springer.

AST511: 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.

Learning Outcomes

After completing this course students are expected to have knowledge about (i) spatial and non spatial data, (ii) geostatistical data and analysis, (iii) spatial interpolation, (iv) apply auto regressive model to areal data, (v) point pattern data analysis.

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.

Textbooks

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, *2nd edition.* Chapman and Hall.

Reference Books

1. Cressie N and Wikle CK (2011). Statistics for Spatio-Temporal Data. Wiley.
2. Illian J, Penttine A, Stoyan H and Stoyan D (2008). Statistical Analysis and Modelling of Spatial Point Patterns. Wiley.

AST518: Introduction to Causal Inference **Credit 3**

Introduction

The course provides an introduction to causal inference with a cohesive presentation of concepts of, and methods for, causal inference.

Objectives

The aim of the course is to facilitate students to define causation in biomedical research, describe methods to make causal inferences in epidemiology and health services research, and demonstrate the practical application of these methods.

Learning Outcomes

Upon completion of the course, students are expected (i) to be familiar with the concepts of causality and counterfactuals (ii) to learn different methods for estimating causual effects in the context of biomedical research (iii) to apply the methods to data obtained from observational studies and to interpret the results.

Contents

Causal effects: individual causal effects, average causal effects, causation versus association.

Randomized experiments: randomization, conditional randomization, standardization, and inverse probability weighting.

Observational studies: identifiability conditions, exchangeability, positivity, and consistency.

Effect modification: stratification, matching, and adjustment methods.

Interaction: identifying interaction, counterfactual response type and interaction.

Directed acyclic graphs (DAGs): complete and incomplete DAGs, statistical DAGs, DAGs and models, paths, chains and forks, colliders, d-seperation.

Unconfounded treatment assignment: balancing scores and the propensity score; Estimating propensity scores: selecting covariates and interactions, constructing propensity score strata, assessing balance conditional on estimated propensity score; Assessing overlap in covariate distributions; Matching to improve balance in covariate distributions: selecting subsample of controls to improve balance, theoretical properties of matching procedures; Subclassification on propensity scores: weighting estimators and subclassification; Matching estimators: matching estimators of ATE; A general method for estimating sampling variances for standard estimators for average causal effects.

Longitudinal causal inference: g-formula and marginal structural models.

Mediation analysis: traditional approaches (direct and indirect effects), counterfactual definitions of direct and indirect effects, regression for causal mediation analysis, sensitivity analysis.

Textbooks

1. Hernan MA and Robins JM (2019). Causal inference. Boca Raton: Chapman & Hall/CRC.
2. Imbens GW and Rubin DB (2015). Causal inference for statistics, social, and biomedical sciences: An introduction. Cambridge University Press.

Reference Books

1. Morgan SL and Winship C (2014). Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge.
2. VanderWeele T (2015). Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford.

AST519: 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.

Objectives

The objectives of the course are to teach students to understand 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 for analyzing longitudinal data. It is also expected that students will be familiar with the proper implementation and interpretation of the statistical methods/models for analyzing longitudinal data and software packages analyzing such data.

Learning Outcomes

Upon completion of the course, students will achieve skills (i) to understand the nature of longitudinal/clustered data (ii) to understand the models and methods for analysing longitudinal/clustered data (iii) to analyse such data and interpret the results and (iv) to understand and interpret research findings of the published reports and articles.

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 quasi-likelihood (PQL), marginal quasi-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.

Textbooks

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.

Reference Books

1. Islam MA and Chowdhury RI (2017). Analysis of Repeated Measures Data. Springer.
2. Diggle PJ, Heagerty P, Liang K-Y, and Zeger SL (2002). Analysis of Longitudinal Data, *2nd edition*. Oxford.
3. Fitzmaurice GM, Laird NM, and Ware JH (2011). Applied longitudinal analysis, *2nd eedition*. Wiley.

AST522: Statistical Signal Processing — **Credit 3**

Introduction

Much of the information around us can be described as signals. Statistical signal processing uses stochastic processes, statistical inference and mathematical techniques to describe, transform, and analyze signals in order to extract information from them.

Objectives

This course is designed to provide statistics graduate students with an overview of different types of signals, their representations and the use of statistical methods, such as estimation and hypothesis testing, to extract information from signals. Its objective is to introduce students to real life applications demonstrating the use of statistics in signal analysis.

Learning Outcomes

At the end of this course a student should be able to : (i) understand basic concepts of signals, signal properties and their representations in time and frequency domains (ii) Apply well-known statistical estimation techniques to estimate signal parameters from noisy signal measurements (iii) Apply well-known statistical decision theory methods to detect signals in Gaussian noise and assess the performance of these methods (iv) Understand and appreciate the importance of statistics in solving real life problems occurring in the domain of signal processing and communication.

Contents

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.

Textbooks

1. Kay SM (1993). Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice Hall.
2. Kay SM (1998). Fundamentals of Statistical Signal Processing: Detection Theory. Prentice Hall.
3. Gonzales RC and Woods RE (2017). Digital Image Processing. *4*th edition, Pearson.

Reference Books

1. Gonzalez RC and Woods RE (2008). Digital Image Processing, *3rd edition*. Pearson Education, Inc.
2. Rahman SMM, Howlader T and Hatzinakos D (2019). Orthogonal Image Moments for Human-Centric Visual Pattern Recognition. Springer.
3. Soliman SS and Mandyam DS (1998). Continuous and Discrete Signals and Systems, *2nd edition*. Prentice-Hall.

AST523: Meta Analysis — Credit 3

Introduction

Meta-analysis refers to the quantitative analysis of study outcomes. Meta-analysis consists of a collection of techniques that attempt to analyze and integrate results that accrue from research studies. This course provides an overview of systematic review and meta-analysis from a statistician's point of view.

Objectives

The main objectives are to introduce students with the merits of meta-analysis and how it can form an important and informative part of a systematic review, with the most common statistical methods for conducting a meta-analysis, and with how to analyze and interpret the results.

Learning Outcomes

At the end of the course, students should be familiar with (i) the research synthesis, (ii) systematic review of the existing research (iii) data extraction from systematic review, and (iv) models and methods for analyzing meta-data for new findings and publication.

Contents

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.

Textbooks

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.

Reference Books

1. Harrer M, Cuijpers P, Furukawa T and Ebert D (2019). Doing Meta-Analysis with R. CRC Press.
2. Chen D-G and Karl EP (2019). Applied Meta-Analysis with R and Stata. Chapman & Hall

AST524: Clinical Trials **Credit 3**

Introduction

The clinical trial is "the most definitive tool for evaluation of the applicability of clinical research". It represents "a key research activity with the potential to improve the quality of health care and control costs through careful comparison of alternative treatments". 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

The main objectve o fthe course is to teach students on the 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.

Learning Outcomes

Upon completion of the course, students will achieve skills (i) to understand, design a trial for assess the effectiveness of a drug, and (ii) to implement and analysis data from such trials and interpret the results.

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.

Textbooks

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.

Reference Books

1. Chow S-C and Liu J-P (2013). Design and Analysis of Clinical Trials: Concepts and Methodologies, 3rd Edition. Wiley.
2. Brody T (2016). Clinical Trials: Study Design, Endpoints and Biomarkers, Drug Safety,

and FDA and ICH Guidelines. Elsevier.

AST525: 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.

Learning Outcomes

After completing the course, students will have the knowledge and skills to: i) Describe a number of models for supervised, unsupervised, and reinforcement machine learning, ii) Assess the strength and weakness of each of these models, iii) Know the underlying mathematical relationships within and across statistical learning algorithms, iv) Identify appropriate statistical tools for a data analysis problems in the real world based on reasoned arguments, v) Develop and implement optimisation methods for training of statistical models, vi) Design decision and optimal control problems to improve performance of statistical learning algorithms, vii) Design and implement various statistical machine learning algorithms in real-world applications, viii) Evaluate the performance of various statistical machine learning algorithms, ix) Demonstrate a working knowledge of dimension reduction techniques. Identify and implement advanced computational methods in machine learning.

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, *1st edition*. Springer.
2. Hastie T, Tibshirani R and Friedman J (2009). The Elements of Statistical Learning: Data Mining, Inference and Prediction, *2nd edition*. Springer.

Reference Books

1. Masashi Sugiyama (2016). Introduction to Statistical Machine Learning. Elsevier Inc.

CSE526: Big Data Analytics   Credit 3

Introduction

This course is designed to introduce key computational concepts, tools and techniques for curating, managing, and analyzing data of large volume, various types and different frequencies. This course assumes basic exposure to the concepts of artificial intelligence, machine learning algorithms and computer programming.

Objectives

The principal aim of this course is to introduce Big Data and its characteristics and challenges to students, and to teach them appropriate tools for managing and analyzing such large-scale data. Further objectives include helping students understand applications of big data in different fields and also ethical issues related to Big Data.

Learning Outcomes

The principal aim of this course is to (i) introduce Big Data and its characteristics and challenges to students, (ii) teach them appropriate tools for managing and analyzing such large-scale data, (iii) help students understand applications of big data in different fields and also ethical issues related to Big Data.

Contents

Introduction to Big Data: definition, characteristics and applications of Big Data in various fields.

Data pre-processing: data collection and extraction – scraping data, data cleaning- handling missing values, noisy data and outliers; redundancy and correlation analysis, tuple duplication, conflict detection and resolution. Structured and unstructured data and databases - relational and NoSQL databases. Data reduction– overview, Wavelet transformation‘, Attribute Subset Selection, Data Cube Aggregation; Data Transformation and Discretization.

Introduction to Big Data Analytics: techniques to address data analysis issues related to data volume (Scalable and Distributed analysis), data velocity (High-Speed Data Streams), Data Variety (Complex, Heterogeneous, or Unstructured data), and Data Veracity (Data Uncertainty).

Database management essentials for Big Data organization and manipulation: Introduction to data organization (lists, queues, priority queues, trees, graphs, hash). Basic graph models and algorithms for searching, shortest path algorithms, flow networks, matching. Processing and streaming Big Data, introduction to data architecture software including MapReduce, Hadoop distributed file system, Spark, Terradata, and how these tools work.

Data Analysis and Visualization Techniques: Descriptive statistics, probabilistic modeling of Big Data (e.g., graphical models, latent variable models, hidden Markov models.) Bayesian Inference (e.g., variational inference, expectation propagation, sampling.) Bayesian Machine Learning (e.g., Bayesian linear regression). Fundamentals of data visualization, Infographics, layered grammar of graphics. Introduction to Modern, mosaic plots, parallel coordinate plots, introduction to GGobi data visualization system, linked plots, brushing, dynamic graphics, model visualization.

Big Data Ethics and Privacy: Ethical considerations in data collection and analysis, privacy and security concerns in Big Data, legal and regulatory frameworks for Big Data.

Textbooks

1. Balusamy B, Nandhini AR, Kadry S, and Gandomi AH (2021). Big Data: Concepts, Technology, and Architecture. Wiley.

Reference Books

1. Li K-C, Jiang H, Yang LT, and Cuzzocrea A (2015). Big Data: Algorithms, Analytics, and Applications. Chapman & Hall/CRC.
2. Erl T, Khattak W, and Buhler P (2016). Big Data Fundamentals: Concepts, Drivers & Techniques. The Prentice Hall.

AST527: Advanced Statistical Machine Learning — Credit 3

Introduction

The course is typically designed for students who have a basic understanding of mathematics, programming, and machine learning concepts. The course provides a broad but thorough introduction to the methods and practice of advanced statistical machine learning and its core methods, models, and algorithms.

Objectives

The aim of the course is to provide students of Applied Statistics and Data Science with detailed knowledge of how advanced Machine Learning methods work and how statistical models can be brought to bear in computer systems not only to analyze large, high-dimensional, unstructured, and big data sets but also to let computers perform tasks efficiently, that traditional methods of statistical and computer science are unable to address. Students will understand the underlying theory and perform assignments that involve a variety of real-world datasets from a variety of domains. They will learn recent statistical techniques based on a synthesis of resampling techniques, and neural networks that have achieved remarkable progress and led to a great deal of commercial and academic interest.

Learning Outcomes

After successful completion of the course, students are expected to (i) Describe a number of advanced machine learning techniques including deep learning, neural network and reinforcement machine learning, (ii) Assess the strength and weaknesses of each of these models, (iii) Know the underlying mathematical relationships within and across statistical learning algorithms, iv) Identify appropriate statistical tools for a data analysis problems in the real world based on reasoned arguments, v) Develop and implement optimisation algorithms for advanced models, (vi) Design decision and optimal control problems to improve performance of statistical learning algorithms, (vii) Design and implement various advanced statistical machine learning algorithms in real-world applications, (viii) Have an understanding of how to choose a model to describe a particular type of data, ix) Evaluate the performance of various advanced statistical machine learning algorithms.

Contents

Overview of the techniques of Artificial Intelligence (AI), advanced statistical machine learning, and data mining. Overview of supervised and unsupervised learning techniques, and sources of big data (such as social media, sensor data, and geospatial data).

Unsupervised Learning: clustering techniques (k-means, hierarchical clustering, DBSCAN) and dimensionality reduction methods (principal component analysis, t-SNE), multidimensional scaling.

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

Ensemble methods: Decision trees, pruning a decision tree, classification trees and comparison with linear models, bootstrap aggregation (Bagging) and random forests, boosting and variable importance, AdaBoost, XGBoost, and lightGBM.

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

Deep Learning: deep neural networks, feedforward neural networks, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep reinforcement learning, TensorFlow or PyTorch.

Applications of machine learning in natural language processing: recurrent neural networks, backpropagation through time, long short-term memory, attention networks, memory networks.

Reinforcement Learning: Overview of the basics of reinforcement learning algorithms.

Probabilistic Graphical Models.

Textbooks

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

Reference Books

1. Erl T, Khattak W, and Buhler P (2016). Big Data Fundamentals: Concepts, Drivers & Techniques, Prentice Hall.
2. Goodfellow I, Bengio Y and Courville A (2016). Deep Learning. MIT Press.

CSE530: Python for Data Science — **Credit 2**

Introduction

Python has become one of the most popular programming languages for data science due to its simplicity, versatility, and robust data analysis and machine learning capabilities. As organizations continue to digitize and generate increasingly large amounts of data, open-source tools like Python can enable data manipulation and exploration, machine learning, and statistical analysis in a scalable manner. This course provides a gentle introduction to programming in Python and its applications in data science and analytics.

Objectives

This course aims to give the students a working knowledge of Python programming. Students will learn programming fundamentals, basic data structures, writing functions, importing and exporting data of different formats, data manipulation, and data visualization using Python. This course will also prepare students for more advanced courses in data science and machine learning.

Learning Outcomes

At the end of this course, students are expected to (i) understand the fundamentals of Python programming, (ii) produce insights from data through exploratory data analysis, and (iii) create effective data visualizations using Python.

Contents

Fundamentals of Python: Installing Python and Jupyter Notebook; the basic syntax of a Python program, Python data types; expressions and variables; lists, tuples, sets, and dictionaries; writing conditions, loops, and functions.

Data analysis with NumPy and pandas: installing NumPy and pandas, NumPy arrays; indexing, slicing, and iterating NumPy arrays; arithmetic and matrix operations with NumPy;

pandas objects– DataFrame, Series, and Index; data indexing and selection; handling missing data; combining and joining datasets, aggregation and grouping, exploratory data analysis.

Data visualization with matplotlib and seaborn: Bar plots, histograms, density plots. box plots and scatterplots.

Hypothesis testing with scipy: z-tests, t-tests, χ^2-tests, analysis of variance (ANOVA).

Statistical Modeling with statsmodels: Linear regression, logistic regression, generalized linear models.

Textbooks

1. McKinney W. (2022). Python for data analysis: Data wrangling with Pandas, NumPy, and Jupyter, *3rd edition.* O'Reilly.
2. VanderPlas J. (2016). Python data science handbook: Essential tools for working with data. O'Reilly.

Reference Books

1. Grus, J. (2019). Data science from scratch: First principles with python. O'Reilly.

AST531: Statistical Computing I — Credit 2

Introduction

This course focuses on the application of Bayesian statistics, geostatistical and spatio-temporal models in real-life situations.

Objectives

The successful completion of the course will help a student to apply methods related to Bayesian statistics and environmental statistical models in various domains.

Learning Outcomes

After completing the course, students will be familiar with (i) handling any computation related to Bayesian models (ii) several Bayesian models including prior and posterior distributions, Bayesian linear regression, MCMC techniques—Gibbs, Metropolis-Hastings, and their implementations, (iii) Bayesian test of hypothesis (iv) geostatistical data and their methods and models (v) Spatio-temporal models and Bayesian approach of spatio-temporal models.

Contents

Computing problems related to AST501: Bayesian Statistics and AST511: Environmental and Spatial Statistics.

AST532: Statistical Computing II — **Credit 2**

Introduction

The course focuses on applications related to longitudinal data and causal inference.

Objectives

The course intends to train students with modern tools of analysing longitudinal data and making causal conclusion from real life data and small area estimation techniques.

Learning Outcomes

At the end of the course students will achieve skills on (i) how to use statistical software and packages to analyse longitudinal/cluster/spatial data (ii) to fit models and interpret the results (iii) to compare and identify the appropriate models required for the data (iv) to analyse data for estimating different causal estimands using observational data, and (v) to write report with statistical results for scientific publications.

Contents

Computing problems related to AST518: Introduction to Causal Inference and AST519: Analysis of Longitudinal Data.

AST533: Statistical Computing III — **Credit 2**

Introduction

The course focuses on applications related to machine learning, big data analytics, and statistical signal processing.

Objectives

The successful completion of the course will help a student to apply methods related to advanced statistical machine learning, Big data analytics, and statistical signal processing in various domains.

Learning Outcomes

After completing the course, students will be familiar with (i) handling any computation related to big data analytics and advanced statistical Machine Learning techniques using R or Python software (ii) with different machine learning techniques including classification trees, CART, regularized regression models for high-dimensional data, spline, gam, ensembled methods-bagging, boosting, random forest, and SVM, (iii) implementation of skills of machine learning tools to any raw datasets, and (iv) sampling signals, signal quantization, noise removal using Fourier transform, estimation of signal parameters.

Contents

Computing problems related to AST522: Statistical Signal Processing, AST525: Statistical Machine Learning and AST526: Big Data Analytics.

AST540: 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.

AST545: SEMINAR **Non-credit**

Assessment of the non-credit seminar course will be either "Satisfactory" or "Non-satisfactory". To get a "Satisfactory" grade, each student needs to attend at least 70% of the seminars organised by the institute during the academic year.

AST550: M.S. PROJECT/INTERNSHIP **Credit 3**

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 two 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. The supervisor cannot evaluate the project or internship report that s/he has supervised. 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 a seminar presentation.

AST551: M.S. THESIS **Credit 6**

After three months of the start, each student of Group B will submit and present a short thesis proposal (length 3-5 A4 pages excluding references, 1.5 line spacing). The proposal 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 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.