STATISTICS (STAT)

STAT 500: Applied Statistics

3 Credits

Descriptive statistics, hypothesis testing, power, estimation, confidence intervals, regression, one- and 2-way ANOVA, Chi-square tests, diagnostics.

Prerequisite: one undergraduate course in statistics

STAT 501: Regression Methods

3 Credits

Analysis of research data through simple and multiple regression and correlation; polynomial models; indicator variables; step-wise, piece-wise, and logistic regression.

Prerequisite: STAT 500 or equivalent; matrix algebra

STAT 502: Analysis of Variance and Design of Experiments

3 Credits

Analysis of variance and design concepts; factorial, nested, and unbalanced data; ANCOVA; blocked, Latin square, split-plot, repeated measures designs.

Prerequisite: STAT 462 or STAT 501

STAT 503: Design of Experiments

3 Credits

Design principles; optimality; confounding in split-plot, repeated measures, fractional factorial, response surface, and balanced/partially balanced incomplete block designs.

Prerequisite: STAT 462 or STAT 501 ; STAT 502

STAT 504: Analysis of Discrete Data

3 Credits

Models for frequency arrays; goodness-of-fit tests; two-, three-, and higher- way tables; latent and logistic models.

Prerequisite: STAT 460 or STAT 502 or STAT 516 ; matrix algebra

STAT 505: Applied Multivariate Statistical Analysis

3 Credits

Analysis of multivariate data; T2-tests; particle correlation; discrimination; MANOVA; cluster analysis; regression; growth curves; factor analysis; principal components; canonical correlations.

Prerequisite: MATH 441 , STAT 501 , STAT 502

STAT 506: Sampling Theory and Methods

3 Credits

Theory and application of sampling from finite populations.

Prerequisite: calculus; 3 credits in statistics

STAT 507: Epidemiologic Research Methods

3 Credits

Research and quantitative methods for analysis of epidemiologic observational studies. Non-randomized, intervention studies for human health, and disease treatment. STAT 507 Epidemiologic Research Methods (3) This 3-credit course develops research and quantitative methods related to the design and analysis of epidemiological (mostly observational) studies. Such studies assess the health and disease status of one or more human populations or identify factors associated with health and disease status. To a lesser degree, the course also covers non-randomized, intervention (experimental) studies that may be designed and analyzed with epidemiological methods. This course is a second-level course and complements Biostat Methods, STAT 509, which is focused on clinical (experimental) trials. Together, these two courses provide students with a complete review of research methods for the design and analysis for common studies related to human health, disease, and treatment. Prerequisite are Intro Biostats (STAT 250 or equivalent).

Prerequisite: STAT 250 or equivalent

STAT 508: Applied Data Mining & Statistical Learning

3 Credits

With rapid advances in information technology, the field of Applied Statistics and Data Science has witnessed an explosive growth in the capabilities to generate and collect data. In the business world, very large databases on commercial transactions are generated by retailers. Huge amounts of scientific data are generated in various fields as well using a wide assortment of high throughput technologies. The internet provides another example of billions of web pages consisting of textual and multimedia information that is used by millions of people. Analyzing large complex bodies of data systematically and efficiently remains a challenging problem. This course addresses this problem by covering techniques and new software that automate the analysis and exploration of large complex data sets. Data Mining methods are introduced by using examples to demonstrate the power of the statistical methods for exploring structure in data sets, discovering patterns in data, making predictions, and reducing the dimensionality by Principal Component Analysis (PCA) and other tools for visualization of high dimensional data. Exploratory data analysis, classification methods, clustering methods, and other statistical and algorithmic tools are presented and applied to actual data. In particular, the course investigates classification methods (supervised learning), and clustering methods (unsupervised learning), and other statistical and algorithmic tools as they are applied to actual data. In addition, data mining and learning techniques developed in fields other than statistics, e.g., machine learning and signal processing, will also be reviewed. The Statistics graduate program also offers more in-depth courses on data mining, STAT 557 and STAT 558. This course focuses on how to use software to investigate and analyze large data sets, whereas STAT 557 and STAT 558 focus more on writing data mining algorithms and the computational aspects of algorithm implementation.

Prerequisite: (STAT 501; STAT 462)

STAT 509: Design and Analysis of Clinical Trials

3 Credits

An introduction to the design and statistical analysis of randomized and observational studies in biomedical research. STAT 509 Design and Analysis of Clinical Trials (3) The objective of the course is to introduce students to the various design and statistical analysis issues in biomedical research. This is intended as a survey course covering a wide variety of topics in clinical trials, bioequivalence trials, toxicological experiments, and epidemiological studies. Many of these topics do not appear in other statistics courses, although a few topics are covered in greater depth in more advanced statistics courses. Computations are performed via the SAS statistical software package. Evaluation methods include four to five homework assignments, an in-class mid-semester examination and an in-class final examination.

Prerequisite: STAT 500

STAT 510: Applied Time Series Analysis

3 Credits

Identification of models for empirical data collected over time. Use of models in forecasting.

Prerequisite: STAT 462 or STAT 501 or STAT 511

STAT 511: Regression Analysis and Modeling

3 Credits

Multiple regression methodology using matrix notation; linear, polynomial, and nonlinear models; indicator variables; AOV models; piece-wise regression, autocorrelation; residual analyses.

Prerequisite: STAT 500 or equivalent; matrix algebra; calculus

STAT 512: Design and Analysis of Experiments

3 Credits

AOV, unbalanced, nested factors; CRD, RCBD, Latin squares, split-plot, and repeatd measures; incomplete block, fractional factorial, response surface designs; confounding.

Prerequisite: STAT 511

STAT 513: Theory of Statistics I

3 Credits

Probability models, random variables, expectation, generating functions, distribution theory, limit theorems, parametric families, exponential families, sampling distributions.

Prerequisite: MATH 230

STAT 514: Theory of Statistics II

3 Credits

Sufficiency, completeness, likelihood, estimation, testing, decision theory, Bayesian inference, sequential procedures, multivariate distributions and inference, nonparametric inference.

Prerequisite: STAT 513

STAT 515: Stochastic Processes and Monte Carlo Methods

3 Credits

Conditional probability and expectation, Markov chains, Poisson processes, Continuous-time Markov chains, Monte Carlo methods, Markov chain Monte Carlo. STAT 515 Stochastic Processes and Monte Carlo Methods (3) This course provides an introduction to stochastic processes and Monte Carlo methods. The course covers topics usually covered in a standard introductory course on stochastic processes, including Markov chains of various kinds. It also covers modern Monte Carlo and Markov chain Monte Carlo methods. Simulation and computing are emphasized throughout the course. The course is divided into two parts: the first part (roughly 8 weeks) provides an introduction to stochastic processes, while the latter (roughly 7 weeks) focuses on Monte Carlo methods, including Markov chain Monte Carlo. The first part of the course begins with a review of elementary conditional probability and expectation before covering basic discrete-time Markov chain theory and Poisson processes. The course then provides students with an overview of continuous-time Markov chains and birth-death processes. The second part of the course covers Monte Carlo methods. Starting with basic random variate generation, the course covers classical Monte Carlo methods such as accept-reject and importance sampling before discussing Markov chain Monte Carlo (MCMC) methods, which includes the Metropolis-Hastings and Gibbs sampling algorithms, and Markov chain theory for discrete-time continuous-space Markov chains.

Prerequisite: MATH 414 , STAT 414 , or STAT 513

STAT 517: Probability Theory

3 Credits

Measure theoretic foundation of probability, distribution functions and laws, types of convergence, central limit problem, conditional probability, special topics.

Prerequisite: MATH 403
Cross-listed with: MATH 517

STAT 518: Probability Theory

3 Credits

Measure theoretic foundation of probability, distribution functions and laws, types of convergence, central limit problem, conditional probability, special topics.

Prerequisite: STAT 517
Cross-listed with: MATH 518

STAT 519: Topics in Stochastic Processes

3 Credits

Selected topics in stochastic processes, including Markov and Wiener processes; stochastic integrals, optimization, and control; optimal filtering.

Prerequisite: STAT 516 , STAT 517
Cross-listed with: MATH 519

STAT 525: Survival Analysis I

3 Credits

Location estimation, 2- and K- sample problems, matched pairs, tests for association and covariance analysis when the data are censored.

Prerequisite: STAT 512 , STAT 514

STAT 540: Statistical Computing

3 Credits

Computational foundations of statistics; algorithms for linear and nonlinear models, discrete algorithms in statistics, graphics, missing data, Monte Carlo techniques.

Prerequisite: STAT 501 or STAT 511 ; STAT 415 ; matrix algebra

STAT 544: Categorical Data Analysis I

3 Credits

Two-way tables; generalized linear models; logistic and conditional logistic models; loglinear models; fitting strategies; model selection; residual analysis.

Prerequisite: STAT 512 , STAT 514

STAT 551: Linear Models I

3 Credits

A coordinate-free treatment of the theory of univariate linear models, including multiple regression and analysis of variance models.

Prerequisite: MATH 415 or STAT 415 or STAT 514 ; STAT 512 ; MATH 436 or MATH 441

STAT 552: Linear Models II

3 Credits

Treatment of other normal models, including generalized linear, repeated measures, random effects, mixed, correlation, and some multivariate models.

Prerequisite: STAT 551

STAT 553: Asymptotic Tools

3 Credits

A rigorous but non-measure-theoretic introduction to statistical large-sample theory for Ph.D. students. STAT 553 Asymptotic Tools (3) STAT 553 covers most standard statistical asymptotics theory but does not require any knowledge of measure theory (it does not define convergence with probability one, for example). It covers convergence of random variables in both the univariate and multivariate settings, Slutsky's theorem(s) and the delta method, the Lindeberg-Feller central limit theorem, power and sample size, likelihood-based estimation and testing, and U-statistics. Although there is no measure theory in the course, it is a mathematically rigorous course and major results are proved. Many common applications of the theory in mathematical statistics are discussed, and most assignments require the use of a computer.

Prerequisite: STAT 513 and STAT 514

STAT 555: Statistical Analysis of Genomics Data

3 Credits

Statistical Analysis of High Throughput Biology Experiments.

Cross-listed with: BIOL 555, MCIBS 555

STAT 557: Data Mining I

3 Credits

This course introduces data mining and statistical/machine learning, and their applications in information retrieval, database management, and image analysis. STAT 557 Data Mining I With rapid advances in information technology, we have witnessed an explosive growth in our capabilities to generate and collect data in the last decade. In the business world, very large databases on commercial transactions have been generated by retailers. Huge amount of scientific data have been generated in various fields as well. For instance, the human genome database project has collected gigabytes of data on the human genetic code. The World Wide Web provides another example with billions of web pages consisting of textual and multimedia information that are used by millions of people. How to analyze huge bodies of data so that they can be understood and used efficiently remains a challenging problem. Data mining addresses this problem by providing techniques and software to automate the analysis and exploration of large complex data sets. Research on data mining have been pursued by researchers in a wide variety of fields, including statistics, machine learning, database management and data visualization. This course on data mining will cover methodology, major software tools and applications in this field. By introducing principal ideas in statistical learning, the course will help students to understand conceptual underpinnings of methods in data mining. Considerable amount of effort will also be put on computational aspects of algorithm implementation. To make an algorithm efficient for handling very large scale data sets, issues such as algorithm scalability need to be carefully analyzed. Data mining and learning techniques developed in fields other than statistics, e.g., machine learning and signal processing, will also be introduced. Example topics include linear classification/regression, logistic regression, model regularization, dimension reduction, prototype methods, decision trees, mixture models, and hidden Markov models. Students will be required to work on projects to practice applying existing software and to a certain extent, developing their own algorithms. Classes will be provided in three forms: lecture, project discussion, and special topic survey/research applications. Project discussion will enable students to share and compare ideas with each other and to receive specific guidance from the instructors. Efforts will be made to help students formulate real-world problems into mathematical models so that suitable algorithms can be applied with consideration of computational constraints. By surveying special topics, students will be exposed to massive literature and become more aware of recent research. Students are strongly encouraged to survey or present their own applications of data mining and statistical learning in graduate research and carry out discussions on data collection and problem formulation.

Prerequisite: STAT 318 or STAT 416 and basic programming skills

STAT 558: Data Mining II

3 Credits

Advanced data mining techniques: temporal pattern mining, network mining, boosting, discriminative models, generative models, data warehouse, and choosing mining algorithms. IST (STAT) 558 Data Mining II (3)This course is the second course in a two-course sequence on data mining. It emphasizes advanced concepts and techniques for data mining and their application to large-scale data warehouse. Building on the statistical foundations and underpinnings of data mining introduced in Data Mining I , this course covers advanced topics on data mining; mining association rules from large-scale data warehouse, hierarchical clustering, mining patterns from temporal data, semi-supervised learning, active learning and boosting. In addition, to computational aspects of algorithm implementation, the course will also cover architecture and implementation of data warehouse, data preprocessing (including data cleansing), and the choice of mining algorithms for applications. In addition to discriminative models such as CRF and SVM models, the course will also introduce generative models such as Bayesian Net and LDA. A term project will be developed by each student to apply an advanced data mining algorithm to a multi-dimensional data set. Classes will include lectures, paper discussions, and project presentations. Paper discussions will allow students to discuss state-of-the-art literature related to data mining. Project presentations will enable students to share and compare project ideas with each other and to receive feedback from the instructor.

Prerequisite: STAT 557 or IST 557
Cross-listed with: IST 558

STAT 561: Statistical Inference I

3 Credits

Classical optimal hypothesis test and confidence regions, Bayesian inference, Bayesian computation, large sample relationship between Bayesian and classical procedures.

Prerequisite: STAT 514; Concurrent: STAT 517

STAT 562: Statistical Inference II

3 Credits

Basic limit theorems; asymptotically efficient estimators and tests; local asymptotic analysis; estimating equations and generalized linear models.

Prerequisite: STAT 561

STAT 565: Multivariate Analysis

3 Credits

Theoretical treatment of methods for analyzing multivariate data, including Hotelling's T2, MANOVA, discrimination, principal components, and canonical analysis.

Prerequisite: STAT 505 , STAT 551

STAT 580: Statistical Consulting Practicum I

2 Credits

General principles of statistical consulting and statistical consulting experience. Preparation of reports, presentations, and communication aspects of consulting are discussed.

Prerequisite: STAT 502 ; STAT 503 , STAT 504 , STAT 506

STAT 581: Statistical Consulting Practicum II

1 Credits

Statistical consulting experience including client meetings, development of recommendation reports, and discussion of consulting solutions. STAT 581 Statistical Consulting Practicum II (1 per semester/maximum of 2) This course serves as a continuation of STAT 580, which provides actual practical experience as a statistical consultant. In STAT 581, each student will hold a consulting session biweekly (by appointment) with a researcher to discuss the statistical design, analysis and computation aspects required for the client's project. Written reports are required for each project and reviewed for appropriateness and accuracy by a supervising faculty member. In addition, a weekly seminar is utilized to discuss selected projects and non-standard applications of statistical methodology. This course will be offered in the spring and summer, with an anticipated enrollment of 15-20 students per semester.

Prerequisite: STAT 580

STAT 584: Machine Learning: Tools and Algorithms

3 Credits

Computational methods for modern machine learning models, including applications to big data and non-differentiable objective functions.

Cross-listed with: CSE 584

STAT 590: Colloquium

1-3 Credits/Maximum of 3

Continuing seminars which consist of a series of individual lectures by faculty, students, or outside speakers.

STAT 592: Teaching Statistics

1 Credits

This course is designed to help students become better teachers and communicators of statistics. INTAF 592 Teaching Statistics (1) This course is designed to help students become better teachers and communicators of statistics, and specifically to prepare students to supervise undergraduate statistics students in labs or small group settings, or even to lead their own undergraduate courses. Students learn about and discuss pedagogy in statistics, gain experience with practice teaching, and improve via individual feedback.

STAT 596: Individual Studies

1-9 Credits/Maximum of 9

Creative projects, including nonthesis research, which are supervised on an individual basis and which fall outside the scope of formal courses.

STAT 597: Special Topics

1-9 Credits/Maximum of 9

Formal courses given on a topical or special interest subject which may be offered infrequently; several different topics may be taught in one year or term.

STAT 600: Thesis Research

1-15 Credits/Maximum of 999

No description.

STAT 601: Ph.D. Dissertation Full-Time

0 Credits/Maximum of 999

No description.

STAT 610: Thesis Research Off Campus

1-15 Credits/Maximum of 999

No description.

STAT 800: Applied Research Methods

3 Credits

Investigates methods for assessing data collected from experimental and/or observational studies in various research setting. STAT 800 Applied Research Methods (3) This course provides students with a broad exploration of the tools and methods in Applied Statistics. In particular, it investigates basic probability distributions and methods for assessing data collected from experimental and/or observational studies in social science and other research settings. Students learn methods of point and interval estimation, including sample size determinations required to achieve a prescribed margin of error. Additionally, students examine hypothesis testing and the determination of sample sizes to achieve a prescribed power of a given test. The distinction between observational studies and randomized experiments is clarified and the limitations of the conclusions are emphasized. Research articles that are relevant to students' fields of study are used to determine how these statistical methods are being applied. Students then identify and critique appropriate research methods. Students work with various data sets to establish fundamental practices that properly analyze data and interpret results via either Minitab or SPSS statistical software as they formulate and communicate conclusions based on a given research context.

STAT 805: Multivariate Statistics and Applications

3 Credits

This course is designed to build upon a student's undergraduate quantitative backgrounds by giving an overview of multivariate statistical techniques. Many applied fields often require the use of large, multivariate data sets and students need to be aware of the wide range of statistical tools available to them. Major objectives of this course are to gain a working knowledge of probability theory, univariate and multivariate statistics, the use of copulas, Monte Carlo techniques, and multiple linear regression. Throughout the course, students will have the opportunity to apply these concepts to real world data sets using modern statistical software packages.

STAT 810: Time Series Analysis and Applications

3 Credits

This course is designed to build upon a student's background by giving an overview of the techniques of time series analysis often used in applied settings. Many areas of research and application often utilize long time series of data in an effort to model changes and volatility in data measured consistently over time. Major objectives in this course include an overview of linear time series; AR, MA, and ARIMA models; ARCH and GARCH models; nonlinear time series models; multivariate time series models; and models of high-frequency data. Throughout the course, students will have the opportunity to apply these concepts to real world data sets using modern statistical software packages.

Prerequisite: (MFE 801, STAT 805; STAT 505)

STAT 897: Special Topics

1-9 Credits/Maximum of 9

Formal courses given on a topical or special interest subject which may be offered infrequently; several different topics may be taught in one year or term.

STAT 897D: **SPECIAL TOPICS**

3 Credits