DEPARTMENT OF BIOSTATISTICS, BIOINFORMATICS, & BIOMATHEMATICS

SERVICE COURSES:

BIST 501: Introductory Biostatistics
This course is designed for introductory biostatistical theory and application for students pursuing a master's degree in fields outside of the Department of Biostatistics, Bioinformatics, and Biomathematics.
Fall Semester
Instructor: Rebecca S. Slack, MS

BIST 502: Applied Biostatistics
This course is designed for a more advanced introductory biostatistical theory and application for students pursuing a PhD in fields outside of the Department of Biostatistics, Bioinformatics, and Biomathematics.
Spring Semester
Instructor: Rebecca S. Slack, MS

MASTER'S PROGRAM:

Year 1, Semester 1:

BIST 510: Probability and Sampling (3 credits)
The goal of the course is to convey an understanding of probability and distribution theory. The probability theory is necessary to provide a foundation for statistics. Probability theory: set theory and probability theory, conditional probability and independence, random variables, distribution functions, density and mass functions for continuous and discrete random variables. Transformation and expectations: distributions of functions of a random variable, expected values, moments and moment generating functions. Common families of distributions: discrete and continuous distributions, exponential family, and location-scale family. Multiple random variables: joint and marginal distributions, conditional distributions and independence, covariance and correlation, multivariate distributions, hierarchical models and mixture distributions. Sampling theory: normal theory, limit theorems.
Prerequisites: Calculus of several variables, matrix theory.
Instructor: Valeriy Korostyshevskiy, PhD

BIST 511: Statistical Inference (3 credits)
This course will introduce the basics of statistical inference, parameter estimation, and hypothesis testing in preparation for more in depth coverage of specific models in later

courses. Inference procedures: point and interval estimation, sufficient statistics, hypothesis testing, methods of constructing test and estimation procedures. Point estimation: criteria for estimators, maximum likelihood estimators, Bayes estimators, mean square error, unbiased estimators, asymptotic variance of estimators. Hypothesis testing: error probabilities, power function, one-sample inference about the mean with known and unknown variance, comparison of two samples, 2×2 contingency tables, shortcuts and non-parametric methods. Modeling and study design: missing data, extreme observations, transformations, factorial experiments, probability sampling, sample size, two-stage sampling, stratified sampling, nonsampling errors.
Prerequisites: Calculus of several variables, matrix theory.
Instructor: Antai Wang, PhD

BIST 515: Introduction to Computational Software(3 credits)
This course introduces students to three different open-source programming languages, Perl, Java, and R, as well as popular mathematical or statistical program languages, SAS and Matlab. For each language, we start with some basic introduction to syntax and semantics. We then work through those languages by developing some example applications using existing libraries that are available for those languages (e.g., Bioperl, Biojava, caBio, and Bioconductor).
Prerequisite: Basic programming: one-semester programming course or self-taught with needed basics.
Instructor: Hongfang Liu, PhD

Year 1, Semester 2:

BIST 512: Statistical Modeling I (3 credits = lectures + computer lab)
3 modules: regression and ANOVA, generalized linear models, longitudinal models.
Lab teaches SAS procedures to fit these models.

1. Regression and ANOVA: fitting a linear regression, least squares method, regression in observational studies, estimation and prediction of Y for a given X, predicted sample mean of Y, testing large deviation, no intercept model, correlation coefficient, bivariate normal distribution, testing single correlation, comparing correlations, rank correlation, 2×C contingency tables, linear trend, R×C contingency tables, fixed-effect ANOVA, model notation, F-test, planned comparisons, orthogonal comparisons, random effects model, homogeneity of variance, randomized blocks, latin squares.
2. Generalized Linear Models: exponential family, proportion, count and rate data, link functions, estimation, logistic and Poisson regressions, fixed and random effects, repeated measures, model selection.
3. Longitudinal Models: generalized estimating equations, estimation, clustered data, linear mixed models, model fit.

Instructors: Lei Nie, PhD and Ying Zhang, MD, MS.

BIST 513: Statistical Modeling II (3 credits = lectures + computer lab)
3 modules: multivariate analysis, survival models, Markov models.
Lab teaches Splus/R procedures to fit these models.

1. Multivariate Analysis: multivariate normal distribution, multivariate analysis of variance, principal components, canonical correlation, factor analysis, discrimination and classification.
2. Survival Models: concepts and methods for analyzing survival time data, life tables, survival distributions, mathematical and graphical methods for evaluating goodness of fit, comparison of treatment groups, regression models, proportional hazards models.
3. Markov Models: probabilistic properties of Markov chains, equilibrium distributions, Hidden Markov models, parameter estimation, applications in biology.

Instructors: Ionut Bebu, PhD and Antai Wang, PhD

BIST 531: Pattern Recognition (3 credits)

The course will introduce the student to the fundamentals of pattern recognition and its application in extracting biological knowledge from high dimensional and low sample-size data. The course will discuss several supervised and unsupervised algorithms and how they can be applied for various purposes including feature extraction, feature selection, dimensionality reduction, clustering, and classification. Particular emphasis will be given to computational methods such as linear discriminant functions, nearest neighbor rule, weighed voting, artificial neural networks, fuzzy logic, support vector machines, genetic algorithms, and swarm intelligence. The course will present some examples of pattern recognition problems in genomics and proteomics (e.g., DNA base calling, analysis of microarray and mass spectral data, etc.) where pattern recognition methods offer a solution.

Instructor: Habtom Ressom, PhD

BIST 541: Principles of Epidemiology (3 credits)

Epidemiology is the scientific discipline of public health. As such, it plays a central role in the identification, characterization, and control of risk factors for human diseases. The course will begin with a brief overview of the history of epidemiology and its historical milestones, followed by consideration of the major types of epidemiological study designs: ecological, cress-sectional, case-control, cohort, and randomized trials. Within each type of study, the course will provide information on the basic methods of analysis associated with the study design, and on special issues such as bias and confounding that must be controlled in epidemiological research. Special topics, such as screening studies, cancer epidemiology, infectious diseases, genetic epidemiology, and biomarker development and validation will also be introduced. This course includes discussion sessions and statistical laboratory exercises.

Instructors: Christopher A. Loffredo, PhD

BIST 918: Practicum (3 credits)

Students will be involved in a research project under the supervision of a faculty member. While the consulting class will expose them to short-term projects, the practicum will provide them with an opportunity to implement a combination of the skills they have acquired and to extend them in a limited context. This practical experience should span 3-4 months. The project will be written up as a Master's paper including the following

sections: background to the problem, experimental design, description of the data, analytical methods, results, and interpretation of the latter. This paper will be defended orally, after no fewer than two faculty members (the advisor and one other) have read it and deemed it ready for presentation.
Instructors: Françoise Seillier-Moiseiwitsch, PhD and Lei Nie, PhD

Summer Internship
All full-time students will be expected to spend part of a summer (at least six full-time weeks) as an intern in an environment where they will need to apply their newly-acquired skills. This can be done either in a research office or laboratory on campus or in a suitable industrial or governmental setting off campus. Part-time students may be exempted from this requirement if they can demonstrate sufficient use of their training in biostatistics or bioinformatics within their current employment.

Year 2, Semester 1:

BIST 535: Case Studies in Bioinformatics (3 credits)
3 one-unit modules, each covering a different topic: e.g., microarray data analysis, phylogenetic analysis, proteomics.

1. Microarray Data Analysis: Microarray technology will be introduced, and the different platforms described. Statistical issues in microarray experiments include: experimental design, sample size calculations, pre-processing and data cleansing, differential expression testing, clustering and prediction. S-PLUS, R and the Bioconductor project, including many statistical and graphical methods that are appropriate in the analysis of microarray data, will be introduced.
2. Phylogenetic Analysis: mutation, evolution, selection, properties of trees, topologies, distance metrics, clustering, distance-based methods, character-based methods, evolutionary models, maximum likelihood estimation, Bayesian approach.
3. Proteomics: technology for 2D gels analysis and mass spectra, biomarker discovery, preprocessing, differential-expression analysis, classification, clustering.

Instructor: Françoise Seillier-Moiseiwitsch, PhD

BIST 545: Case Studies in Epidemiology (3 credits)
3 one-unit modules, each covering a different topic: e.g., genotyping studies, biomarkers of exposure, family studies.

1. Genotyping Studies. Types of epidemiology studies using genotypes, study design issues, genotyping methods including quality control, database aspects, analysis of genotype data, field trip to genetic laboratory
2. Biomarkers of Exposure. Types of epidemiology studies using biomarkers, study design issues, biomarker detection methods including quality control, database aspects, analysis of biomarker data, field trip to biomarker laboratory
3. Family Studies. Types of epidemiology studies using family-based designs, study design issues, special methods in family studies, database aspects, analysis of family data, field trip to Familial Cancer Registry.

Instructor: Yun-Ling Zheng, PhD, MD, MPH

BIST 530: Biostatistics for Bioinformatics (3 credits)
Bioinformatics is the application of computer science, statistics, and mathematics to the management and analysis of large-scale, complex biological data. This course will enable students to obtain some understanding of the statistical methods needed to analyze such data. During the first weeks of the course, we will provide a basic introduction to database management systems and an overview of important biological databases including GenBank, UniProt, and iProClass. The course will then go on to describe the underlying theories and algorithms for sequence alignment (pairwise, multiple, nucleotides, proteins, statistical evaluation), sequence analysis (correlations, profiles, PAM and BLOSUM matrices), genome comparison (dot matrices), molecular evolution, and gene prediction. For each of these topics, available tools will be introduced during hands-on laboratory sessions.
Prerequisite: Introductory statistics
Instructors: Françoise Seillier-Moiseiwitsch, PhD and Hongfang Liu, PhD

BIST 540: Clinical Trials (3 credits)
The objective of the course is to explain in practical terms the basic principles of clinical trials, with particular emphasis on their scientific rationale, organization and planning, and methodology. Issues discussed include design of randomized and non-randomized trials, size of a clinical trial, monitoring of trial progress, and some basic principles of statistical analysis. The intent is to present the methodology of clinical trials with emphasis on the practical aspects.
Instructor: George Luta, PhD
Prerequisite: Introductory statistics

BIST 550: Consulting (2 credits)
This course offers instruction, discussion and hands-on experience providing statistical consultation in applied scientific situations. These will typically include survival analysis, clinical trial/study design, tumor growth curves, microarray analysis, and proteomics projects. Instruction and experience will focus on consulting strategy. This includes preparing analysis plans and reports, communication and time management skills, and ethics/professional standards. Additionally, students will gain consulting practice to include interactions with investigators with actual projects and problems. Students will attend class and weekly consulting/debriefing sessions, prepare analysis plans and reports, and give oral presentations describing consulting projects over the semester.
Instructor: Rebecca Slack, MS

BIST 520: Special Topic—Meta-Analysis
Concerned with the effective use of existing clinical studies to inform decision making and health care policy, this course introduces the basic methods of systematic review of the medical literature, including meta-analysis. The principles and methods of systematic reviews, as well as statistical approaches to meta-analysis for clinical trials and observational studies are introduced and their application illustrated in the context of

actual clinical examples. The use of meta-analysis to explore data and identify sources of variation among studies is emphasized, as is the use of meta-analysis to identify future research questions.
Instructor: Michael A. Stoto, PhD
Prerequisite: Statistics, Epidemiology and/ or Introductory Biostatistics

BIST 521: Special Topic—Multiple Comparisons
The course will introduce students to the basic methods for multiple comparisons, and their application to real-life data using SAS.
Instructor: George Luta, PhD
Prerequisite: SAS

BIST 522: Special Topic—Introduction to Population Dynamics
The purpose of the course is to study models of population dynamics, develop skills required for their analysis, become familiar with relevant machinery. Upon completing the course, students will be able to identify different factors acting on a population, construct the corresponding mathematical model, analyze it using both analytical and computational methods (e.g., Matlab, Maple or similar).
Instructor: Valeriy Korostyshevskiy, PhD
Prerequisite: Calculus