Biostatistics and Data Science

Degree Offered

Master of Arts

Program Description

The Master of Arts program in Biostatistics and Data Science provides a learning experience focused on solid theoretical foundation and practical experience. Robust course offerings, active engagement in statistical consulting, and a capstone project will provide ample opportunities to develop the skills needed to succeed in today's job market. A large volume of consulting projects which range from the simplest statistical summaries to the most complex protocols and data collection schemes will allow MA students to get experience of working with real data analysis project from start to finish. This hands-on experience will provide students with an opportunity to synthesize the acquired knowledge and integrate various courses they have taken; help students to build a skill set needed to develop a timeline for completion of data analysis project from start to finish, including study planning, data acquisition, its thorough analysis, and preparation of the final report summarizing the findings; provide a student with a portfolio which demonstrates competency in data analysis, statistical programming, consulting experience with non-statisticians, oral and written communication skills.

Program Admissions Requirements

*In addition to the general **Graduate School admission requirements**, this program has additional specific requirements.*

- Completed an undergraduate degree including coursework in calculus, probability and/or statistics, linear/matrix algebra, and computer programming experience;
- A strong interest in biostatistics and biomedical applications;
- An overall grade point average of 3.0 or better;
- A 3.0 grade point average or better in mathematics and science courses;
- Scores in the 70th percentile or higher on the Quantitative component of the Graduate Record Examination (GRE) is ideal.

Overall Course Requirements

The curriculum will consist of 8 core, required courses which have been identified as an essential knowledge base for all students in the program. Also required, is an Ethics and Integrity in Science course along with at least 3- credit capstone project course. The capstone project course can be taken throughout multiple semesters but at least

3 credit hours are required for graduation. Students can choose two or more elective courses which best reflect their personal interests.

Required Courses
BIOE 10222 Ethics and Integrity in Science. *1 credit.*
This course provides the basis for understanding the ethical issues related to basic scientific and medical research, including animal and human subject research, fraud and misconduct, and governmental, institutional, and researcher responsibilities.
BIOS 04224 Biostatistical Computing. *3 credits.*
This course will cover the details of manipulating and transforming data required for statistical analysis, such as reshaping the data from a per-case to a per-event within a case and vice-versa. It will also cover the techniques necessary to write functions and macros in both SAS and R for developing new/modified data analysis methods. Students are expected to be facile in the use of computers before they take this course. LaTeX document production system is also introduced.
BIOS 04221 Biomedical Applications and Consulting. *3 credits.*
Theory of consulting, statistical techniques most often used in consulting, practical experience in the real consulting setting and writing statistical reports.
BIOS 04231 Statistical Models and Methods I. *3 credits.*
This course will cover statistical models and analyses for count data and contingency tables, basic nonparametric methods including sign, rank-sum, and signed-ranks tests, simple linear regression model and inference, checking model assumptions, correlation analysis, one-way and two-way analysis of variance. Emphasis is on models, their application to data and interpretation. SAS will be used throughout the course.
BIOS 04232 Statistical Models and Methods II. *3 credits.*
Factorial, nested, split-plot and repeated measures designs, multiple regression and variable selection, multiple comparisons, logistic regression, discriminant analysis, principal components and factor analysis, rates and proportions, introduction to survival analysis.
BIOS 04233 Introduction to Statistical and Machine Learning. *3 credits.*
This course will provide an introduction to statistical learning. Core topics include variable selection, penalized linear regression such as lasso, dimension reduction including principal component analysis, flexible regression techniques including kernel smoothing/smoothing splines/generalized additive models/regression trees, support vector machine, clustering, and random forests. Other topics that can be covered include but are not limited to ridge regression, group lasso, fused lasso, adaptive lasso, SCAD, Bayesian lasso, Bayesian group

lasso, Bayesian CART, BART, Neutral network, feature screening, graphical models, and quantile regression.

BIOS 24160 Concepts in Probability and Statistics. *3 credits.*
The course is designed for graduate students who have a background in statistics but would benefits from a review of the basic concepts in probability and statistics. Given that students entering the new MS program in Biostatistics may be diverse mathematical backgrounds, there is a need for a course which would provide them with necessary preparation for taking more advanced courses in statistics, both theoretical and applied.

BIOS 24150 Bioinformatics in Omics Analysis. *3 credits.*
The course aims to introduce modern statistical and computational methods in high-throughput omics data analysis. The first half of the course focuses on fundamental statistical and computational methods applicable in different types of high-throughput omics data. The second half covers selected important topics in bioinformatics and aims to give students a systematic view of the omics data analysis. The goals of the course include: (1) to motivate students from quantitative fields into omics research (2) to familiarize students from biological fields with a deeper understanding of statistical methods (3) to promote inter-disciplinary collaboration atmosphere in class. Students are required to have a basic statistical training (i.e. elementary statistics courses, basic calculus, and linear algebra) and basic programming proficiency (R programming is required for homework and the final project).

BIOS 24297 Capstone Project. *1-3 credits.*
The course is the culmination of the MS program in Biostatistics. Students will complete a project integrating their statistical analysis, data science, and application domain knowledge. The project results in a written report and presentation which will improve students' ability to communicate effectively about statistics and data science in written and oral form using both technical and nontechnical language. In addition, the project will enable students to expand their professional portfolio of coding samples, written reports and presentations.

Elective Courses
Any two of the following:

BIOS 04214* Design and Analysis of Clinical Trials. *3 credits.*
This course covers issues in clinical trials including the clinical trial protocol, sources of bias in clinical trials, blinding, randomization, sample size calculation; phase I, phase II, phase III and hybrid trials; interim analysis, stochastic curtailment, Bayes designs, and administrative issues in study design.

BIOS 04285* Introduction to Bayesian Analysis. *3 credits.*

This course introduces basic concepts and computational tools for Bayesian statistical methods. Topics covered include one and two sample inference, regression models and comparison of several populations with normal, dichotomous and count data.

BIOS 04275* Applied Survival Analysis. *3 credits.*
Basic parameters in survival studies; Censoring and truncation, Competing risks; Univariate estimation including the Kaplan-Meier and Nelson-Aalen estimator; tests comparing two or more populations, the log rank test; Semi-parametric regression, the Cox model; Aalen's Additive hazards regression model; regression diagnostics.

BIOS 04222 Statistical Consulting. *3 credits.*
This course is designed for students to gain experience in statistical consulting by working with the biostatistics faculty members on various consulting projects.

PUCH 19210 Health and Medical Geography. *3 credits.*
Geography and physical and social environments have important implications for human health and health care. This course will explore the intersections among geography, environments and public health, with an emphasis on geographical analysis approaches for health data, to address two key questions: (1) How can concepts from geography help us to better understand health and well-being? (2) How can geographic tools, such as Geographic Information Systems (GIS) be used to address pressing questions in health and medical research?

PUCH 19229 Survey Research Methods. *3 credits.*
Survey Research Methods is a graduate-level, 3-credit hour course that introduces students to the broad concepts of survey design, conduct, and analysis. Students will gain a detailed and comprehensive understanding of questionnaire design, sampling, data collection, survey non-response, and analysis of survey data. The course will include lectures, reading assignments, class discussions, individual and group presentations, and exams.

PUCH 19150 Introduction to Epidemiology. *3 credits.*
The course provides: 1) an overview of epidemiologic concepts; 2) an introduction to the approaches and techniques that are used to measure and monitor health status in populations; 3) an introduction to study designs to assess disease prevention and intervention; and 4) an introduction to clinical research study designs that elucidate causative factors for disease.

CTSI 20201 Clinical Epidemiology. *2 credits.*
This course is designed to provide epidemiology research methodologies to clinical practical applications. Topics include diagnostic testing, meta-analysis, qualitative research, data collection and survey design. Students will learn to apply research

methodologies to large data sets or populations, while understanding the reliability, and validity of their methods.

CTSI 20150 Introduction to Epidemiology. *2-3 credits.*
The course provides: 1) an overview of epidemiologic concepts; 2) an introduction to the approaches and techniques that are used to measure and monitor health status in populations; 3) an introduction to study designs to assess disease prevention and intervention; and 4) an introduction to clinical research study designs that elucidate causative factors for disease.

PUBH 18201 Principles of Epidemiology (online). *3 credits.*
The course provides: 1) an overview of epidemiologic concepts; 2) an introduction to the approaches and techniques that are used to measure and monitor health status in populations; 3) an introduction to study designs to assess disease prevention and intervention; and 4) an introduction to clinical research study designs that elucidate causative factors for disease.

*NOTE: One of the three elective courses – either BIOST 04214, 04285, or 04275 is required.