Biostatistics (BSTA)

Courses

**BSTA 001 Population Health Data Science I 3 Credits**
Students will learn the fundamentals of probability theory, univariate statistics, statistical computing/programming/visualization, and machine learning. A mix of traditional and experiential learning will focus on how to build an analysis pipeline to answer pressing questions in population health. In-class examples and projects will use real data sets. Students will propose a small data-driven project focused in population health, and use their newly-acquired data science skills to collect, analyze, and present their work. May be taken in conjunction with BSTA 002.

**BSTA 002 Population Health Data Science I Algorithms Lab 1 Credit**
Students will apply statistical concepts learned in BSTA 001 and use real data sets. Students will propose a small data-driven project focused in population health, and use their newly-acquired data science skills to collect, analyze, and present their work. May be taken in conjunction with BSTA 002.

**BSTA 005 Statistical Literacy in Health 3 Credits**
This course is designed to introduce students with a fear of all things mathematical to the importance of statistics in health research. Students will learn how to read and understand basic statistical concepts and methods used in health research, such as probability, sampling, hypothesis testing, and correlation. Students will also learn to interpret tables and statistical findings in the health literature.

**BSTA 101 Population Health Data Science II 3 Credits**
Students will expand their statistics and machine learning toolkit by learning how to compare univariate distributions, build traditional regression models for continuous and binary data, explore supervised learning methods such as: Tree-based learning, KNN/Collaborative filtering, and Feed forward Neural networks, and understand how to manipulate, ask, and answer questions from big datasets. Students will be expected to propose a population health project mid-semester, and apply and present techniques they learned in class. May be taken in conjunction with BSTA 103.

**BSTA 103 Population Health Data Science II Algorithms Lab 1 Credit**
Students will apply regression and machine learning models learned in BSTA 101 Population Health Data Science II to health datasets using Python3. Datasets that students will study include: the National Health and Nutrition Examination Survey, influenza-like illness and confirmed positive cases of COVID-19 tracked by the Centers for Disease Control and Prevention, and the National Health Interview Survey. Lab is to be taken concurrently with lecture (BSTA 101 Population Health Data Science II).

**BSTA 106 Population Health Data Science III 3 Credits**
Students will expand their statistics and machine learning toolkit by learning how to compare univariate distributions, build traditional regression models for continuous and binary data, explore supervised learning methods such as: Tree-based learning, KNN/Collaborative filtering, and Feed forward Neural networks, and understand how to manipulate, ask, and answer questions from big datasets. Students will be expected to propose a population health project mid-semester, and apply and present techniques they learned in class. May be taken in conjunction with BSTA 103.

**BSTA 200 Population Health Data Science III Algorithms Lab 1 Credit**
Students will apply regression and machine learning models learned in BSTA 200 Population Health Data Science III to health datasets using Python3. Datasets that students will study include: the National Health and Nutrition Examination Survey, influenza-like illness and confirmed positive cases of COVID-19 tracked by the Centers for Disease Control and Prevention, and the National Health Interview Survey. Lab is to be taken concurrently with lecture (BSTA 200 Population Health Data Science III).

**BSTA 206 Population Health Data Science III Algorithms Lab 1 Credit**
Students will apply regression and machine learning models learned in BSTA 206 Population Health Data Science III to health datasets using Python3. Datasets that students will study include: the National Health and Nutrition Examination Survey, influenza-like illness and confirmed positive cases of COVID-19 tracked by the Centers for Disease Control and Prevention, and the National Health Interview Survey. Lab is to be taken concurrently with lecture (BSTA 206 Population Health Data Science III).

**BSTA 301 Population Health Data Science IV 3 Credits**
Students will expand their statistics and machine learning toolkit by learning how to compare univariate distributions, build traditional regression models for continuous and binary data, explore supervised learning methods such as: Tree-based learning, KNN/Collaborative filtering, and Feed forward Neural networks, and understand how to manipulate, ask, and answer questions from big datasets. Students will be expected to propose a population health project mid-semester, and apply and present techniques they learned in class. May be taken in conjunction with BSTA 303.

**BSTA 303 Population Health Data Science IV Algorithms Lab 1 Credit**
Students will apply regression and machine learning models learned in BSTA 303 Population Health Data Science IV to health datasets using Python3. Datasets that students will study include: the National Health and Nutrition Examination Survey, influenza-like illness and confirmed positive cases of COVID-19 tracked by the Centers for Disease Control and Prevention, and the National Health Interview Survey. Lab is to be taken concurrently with lecture (BSTA 303 Population Health Data Science IV).

**BSTA 305 Population Health Data Science III 3 Credits**
In this course, students are introduced to Bayesian statistics and computational techniques, ensemble learning (boosting, bagging, stacking), how to handle missing data, and how to build reproducible analysis pipelines via Makefiles. The statistical and machine learning techniques taught will be applied to a variety of real population health datasets; students will apply these techniques and submit a research-style manuscript. Students will investigate a current problem in population health and provide a data-driven solution.

**Prerequisites:** BSTA 101 or BSTA 103
**Corequisites:** BSTA 306

**BSTA 306 Population Health Data Science III Algorithms Lab 1 Credit**
Required programming lab course for students enrolled in BSTA 305 Population Health Data Science III.

**Prerequisites:** BSTA 101 or BSTA 103
**Corequisites:** BSTA 305

**BSTA 307 Applied Machine Learning for Health Sciences 3 Credits**
Machine learning uses interdisciplinary techniques to create automated systems that can sift through large amounts of data at high speed to make predictions and decisions with minimal human intervention. Machine learning is increasingly pervasive and impactful in public health and precision medicine. This class will provide students with a broad cross-section of practical machine learning skills while giving them an appreciation of how machine learning is being applied in the context of public health research today.

**Prerequisites:** BSTA 101

**BSTA 308 Advanced R Programming 3 Credits**
R language syntax and structure. R programming techniques. Emphasis on structured design for medium to large programs. R package development fundamentals. Capstone development project.

**Prerequisites:** BSTA 101 and BSTA 103

**BSTA 320 (CGH 320, POPH 320) Independent Study or Research in Biostatistics 1-4 Credits**
This course can be directed readings or research in Biostatistics or an experiential learning experience that puts student's understanding of Biostatistics into practice. Department permission required.

**Repeat Status:** Course may be repeated.

**BSTA 396 1-4 Credits**
**Repeat Status:** Course may be repeated.

**BSTA 402 Health Data and Computational Science 3 Credits**
Students will learn the fundamentals of probability theory, univariate statistics, statistical computing/programming/visualization, supervised and machine learning. We will focus on theoretical concepts and on writing code in Python to apply what we learn in class to real-life population health examples such as: health policy, infectious disease, clinical trials, and data collected from national databases. A mix of traditional and experiential learning applied to real-life population health data will equip students with statistical tools to answer pressing population health problems.

**BSTA 403 Applications in Statistical Learning 3 Credits**
This course will explore common statistical models used to analyze both continuous, discrete, and time to event data: simple and multivariate linear regression, logistic regression, poisson and negative binomial regression, and survival models. An emphasis will be placed on supervised learning. Throughout the semester, students will apply the theoretical background they learn in class to population health data sets, generating their own hypotheses and testing them with rigorous statistical methods.

**Prerequisites:** BSTA 402
BSTA 404 Data Architecture, Mining, and Linkage 3 Credits
This course will focus on collecting, storing, and formatting data for use in population health data analysis. Students will learn fundamental concepts and best practices for working with data, how to use Python to scrape the internet for data related to population health and learn how to link a diverse set of data together to test novel hypotheses students themselves pose during class.