### Biostatistics (BSTA)

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Name</th>
<th>Credits</th>
<th>Prerequisites</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSTA 001</td>
<td>Population Health Data Science I 3 Credits</td>
<td>3</td>
<td>Corequisites: BSTA 000</td>
</tr>
<tr>
<td>BSTA 002</td>
<td>Population Health Data Science Algorithms Lab 1</td>
<td>1</td>
<td>Corequisites: BSTA 001</td>
</tr>
<tr>
<td>BSTA 101</td>
<td>Population Health Data Science II 3 Credits</td>
<td>3</td>
<td>Corequisites: BSTA 001</td>
</tr>
<tr>
<td>BSTA 103</td>
<td>Population Health Data Science Algorithms Lab 1</td>
<td>1</td>
<td>Corequisites: BSTA 001</td>
</tr>
<tr>
<td>BSTA 300</td>
<td>Apprentice Teaching 1-4 Credits</td>
<td>1-4</td>
<td>Repeat Status: Course may be repeated.</td>
</tr>
<tr>
<td>BSTA 305</td>
<td>Population Health Data Science III 3 Credits</td>
<td>3</td>
<td>Corequisites: BSTA 101 or BSTA 103</td>
</tr>
<tr>
<td>BSTA 402</td>
<td>Health and Computational Science 3 Credits</td>
<td>3</td>
<td>Prerequisites: BSTA 402</td>
</tr>
<tr>
<td>BSTA 403</td>
<td>Applications in Statistical Learning 3 Credits</td>
<td>3</td>
<td>Prerequisites: BSTA 402</td>
</tr>
<tr>
<td>BSTA 404</td>
<td>Data Architecture, Mining, and Linkage 3 Credits</td>
<td>3</td>
<td>Prerequisites: BSTA 402</td>
</tr>
</tbody>
</table>

**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.

**Corequisites:** BSTA 001

**Prerequisites:** Population Health Data Science I

**BSTA 002 Population Health Data Science I Algorithms Lab 1 Credit**

Students will apply statistical concepts learned in BSTA 001 Population Health Data Science I to describe, visualize, and draw inferences on a variety of health datasets using Python3. Datasets that students will study include: influenza-like illness and confirmed positive cases of COVID-19 tracked by the Centers for Disease Control and Prevention, and a repository of potentially hazardous chemicals found in commercial products. Lab is to be taken concurrently with lecture (BSTA 001 Population Health Data Science I).

**Corequisites:** BSTA 001

**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.

**Prerequisites:** BSTA 001

**BSTA 103 Population Health Data Science II Algorithms Lab 1 Credit**

Students will apply regression and machine learning models learned in BSTA101 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 (BSTA101 Population Health Data Science II).

**Prerequisites:** BSTA 001

**BSTA 300 Apprentice Teaching 1-4 Credits**

Repeat Status: Course may be repeated.

**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 402 Health 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.

**Prerequisites:** BSTA 402

**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.