Biostatistics (BSTA)

Courses

**BSTA 001 Population Health Data Science I 4 Credits**
This course teaches students 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.

**BSTA 101 Population Health Data Science II 4 Credits**
In this course students 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.

**Prerequisites:** BSTA 001

**BSTA 102 Population Health Data Science III 4 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-drive solution.

**Prerequisites:** BSTA 101