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. Must be taken in conjunction with BSTA 002.
Corequisites: BSTA 002

BSTA 003 Computational Thinking 3 Credits
This course introduces computational thinking as a problem-solving methodology in health and biological sciences. You will explore the approach of developing theoretical models for natural events and converting them into computer simulations using tools like R, Python, MATLAB, or SAS. The course emphasizes fundamental programming concepts, making it suitable for beginners, while also highlighting computational thinking in health. Additionally, the course explores ethics in computational science, covering responsible algorithmic decision-making, data management, privacy, bias, and transparency in computing.

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 007 (POPH 007) Frontiers of AI in Health 3 Credits
This course presents a broad contemporary survey of the actual and potential contributions of Artificial Intelligence and Health Data Science in addressing public health challenges. By reading recent articles that describe case studies of AI in health and healthcare and by engaging in discussions both in class and online, students will come to appreciate the many unsolved problems in public health and how one may evaluate the potential benefits and risks of exciting new data-centric solutions made possible by AI.

BSTA 030 Data Exploration in R 3 Credits
This course provides an introduction to problem-solving using the R environment for statistical computing and graphics. Students will gain experience designing, implementing, and testing their R code. Multiple programming paradigms will be explored. The course covers R data types, input and output, and control flow in the context of preparing, cleaning, transforming, and manipulating data. Students will use R to conduct exploratory data analyses, including computing descriptive statistics and data visualization. Students should expect to spend each class writing programs.
Prerequisites: CSE 012

BSTA 040 Data Exploration in Python 3 Credits
This course provides an introduction to the fundamentals of programming in Python. Students will gain experience designing, implementing, and testing their Python code, as well as in using Jupyter Notebooks, and IPython for statistics and data analysis. Multiple programming paradigms will be explored. The course covers Python data types, input, and output, and control flow in the context of preparing, cleaning, transforming, and manipulating data. In addition, students will use Python to conduct exploratory data analyses, including computing descriptive statistics.
Prerequisites: CSE 012

BSTA 101 Population Health Data Science I 3 Credits
This course provides an introduction to the use of statistics in health. Topics include data presentation, descriptive statistics, probability and probability distributions, parameter estimation, hypothesis testing, analysis of contingency tables, analysis of variance, linear and logistic regression models, and sample size and power considerations. Students develop the skills necessary to perform, present, and interpret basic statistical analyses. Must be taken in conjunction with BSTA 102.
Corequisites: BSTA 102

BSTA 102 Population Health Data Science I Algorithms Lab 1 Credit
Students will use a statistical computing platform to apply concepts learned in BSTA 101 and attain autonomy in handling real-world data. Lab must be taken concurrently with lecture (BSTA 101 Population Health Data Science I).
Corequisites: BSTA 101

BSTA 103 Population Health Data Science II 3 Credits
This course is a continuation of BSTA 101. Topics include an overview of generalized linear models, simple and multiple linear regression, regression models for binary data, regression models for count data, quasi-likelihood methods, extensions of generalized linear models. Must be taken in conjunction with BSTA 104. Prerequisites: BSTA 101.
Prerequisites: BSTA 101
Corequisites: BSTA 104

BSTA 104 Population Health Data Science II Algorithms Lab 1 Credit
Students will use a statistical computing platform to apply regression techniques learned in BSTA103 Population Health Data Science II to health datasets. Lab must be taken concurrently with lecture (BSTA103 Population Health Data Science II).
Prerequisites: BSTA 101
Corequisites: BSTA 103

BSTA 120 (CGH 120, EPI 120, POPH 120) Independent Study or Research 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 130 Internship 1-4 Credits
In this introductory course, students will engage in supervised work in Biostatistics. Placements will be arranged to suit individual interests and career goals. Potential internship sites include government agencies, non-profit organizations, and the private sector. A written report is required, and a preceptor evaluation will be required. Department permission is required.
Repeat Status: Course may be repeated.

BSTA 132 Health Data Science I: Inference 4 Credits
This course provides an introduction to methods of statistical inference as applied to health data. Topics covered include hypothesis testing, confidence intervals, analysis of variance, correlation, and non-parametric methods. The course will illustrate these concepts using data from the health context. In addition to traditional methods of learning, computing will be a significant component of the course, ensuring students acquire the skills to both formulate and answer pressing questions in population health.
Prerequisites: MATH 052 and MATH 043 and BSTA 030

BSTA 133 Health Data Science 2: Regression 4 Credits
This course provides an introduction to generalized linear models as applied to health data. Topics covered include models for binary data, models for nominal and ordinal data, models for count data, quasi-likelihood methods, and Bayesian generalized linear models. The course will illustrate these concepts using data from the health context. In addition to traditional methods of learning, computing will be a significant component of the course, ensuring students acquire the skills to both formulate and answer pressing questions in population.
Prerequisites: BSTA 132

BSTA 141 Health Data Science III: Supervised Machine Learning in Health 4 Credits
Supervised machine learning is used to create automated systems that sift through labeled/continuous data at high speed to make predictions with minimal human intervention. This course provides students with skills in applying supervised machine learning in contexts of population health. We will cover regression, classification, cross-validation, hyperparameter selection, feature selection, feature engineering, ensemble methods, regularization, and reinforcement learning. Students will learn concepts through hands-on engagement with health data sets, preparing them to contribute effectively to data-driven precision population health.
Prerequisites: MATH 052 and MATH 043 and BSTA 040

BSTA 142 Health Data Science IV: Unsupervised Machine Learning in Health 4 Credits
Unsupervised machine learning is used to discover hidden patterns and structures in high-dimensional unlabeled health data. This course will survey leading techniques for clustering and dimensionality reduction. The course will cover hierarchical and density-based clustering techniques, along with modeling using Gaussian mixtures, factor analysis, and principal component analysis. Applications considered will include patient clustering for personalized treatment, anomaly detection for early disease identification, and dimensionality reduction for efficient analysis of diverse and complex medical datasets.
Prerequisites: BSTA 141 and MATH 052 and MATH 043 and BSTA 040

BSTA 150 Special Topics in Biostatistics 3-4 Credits
In this course, students will engage in an intensive exploration of a topic of special interest that is not covered in other courses. Topics addressed will be at an intermediate level.
Repeat Status: Course may be repeated.

BSTA 160 Biostatistics Study Abroad 1-3 Credits
Biostatistics focused course taken during an abroad experience.
Repeat Status: Course may be repeated.

BSTA 300 Apprentice Teaching 1-4 Credits
Repeat Status: Course may be repeated.

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 309 Outbreak Science & Public Health Forecasting 3 Credits
This course aims to introduce students to models that describe the spread of a pathogen through a population, and how models can support public health decisions. The course will be split into four parts: (i) the factors that motivate public health actions, (ii) epidemic models such as the Reed-Frost and SIR, (iii) statistical time series and forecasts, (ii) a focus on ensemble building. Students will be expected to complete mathematical/statistical exercises and write code that simulates infectious processes.
Prerequisites: BSTA 101 and BSTA 103

BSTA 310 (CSE 310) Assistive Technologies 3 Credits
This class will introduce typical challenges faced by persons with disabilities and the role of assistive technologies (ATs) in solving such challenges. The class will examine opportunities presented by recent advances in mobile and AI technologies. Working in groups, each student will be expected to acquire and apply relevant skills in designing AT solutions. The class can be taken by students with diverse backgrounds including the following: community and population health, social and behavioral sciences, business, engineering and computer science.
Prerequisites: CSE 017 or (BSTA 101 and BSTA 102)

BSTA 320 (CGH 320, EPI 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 330 Internship 1-4 Credits
In this advanced course, students will engage in supervised work in Biostatistics. Placements will be arranged to suit individual interests and career goals. Potential internship sites include government agencies, non-profit organizations, and the private sector. A written report is required, and a preceptor evaluation will be required. Department permission is required.
Repeat Status: Course may be repeated.

BSTA 350 Special Topics in Biostatistics 3-4 Credits
In this course, students will engage in an intensive exploration of a topic of special interest that is not covered in other courses. Topics addressed will be at an advanced level.
Repeat Status: Course may be repeated.

BSTA 360 Biostatistics Study Abroad 1-3 Credits
Upper-level biostatistics focused course taken during an abroad experience.
Repeat Status: Course may be repeated.

BSTA 372 Analyzing Electronic Health Record Data 3 Credits
This course will explain the structure and provide computing skills to analyze Electronic Health Record (EHR) data. Through a series of health-related case studies, students will have the opportunity to experience EHR as a comprehensive platform to support best-in-class evidence-based care and as the core component for big data analytics to help care organizations adapt and transform into learning organizations. The course will present a number of EHR data architectures, data standards, quality assessment, and workflow methods.
Prerequisites: BSTA 142

BSTA 373 Analyzing Clinical Natural Language Data 3 Credits
This course will convey specialized clinical natural language processing (NLP) principles and methods, as well as how to write regular expressions and parse and collate information from text-rich health documents such as electronic health records, clinical notes, and peer-reviewed medical literature. The course will engage real-world data sets for students to develop text-processing strategies. Computing will be a significant component of the course, ensuring students acquire the skills necessary to work with clinical natural language data.
Prerequisites: BSTA 142

BSTA 374 Analyzing Health GIS Data 3 Credits
This course will convey specialized methodologies of data collection and the statistical analysis of spatial data. Through a series of health-related case studies, students will have the opportunity to explore spatial statistical analysis at a variety of spatial resolutions. Computing will be a significant component of the course, ensuring that students acquire the skills necessary to apply these techniques to health-related GIS data.
Prerequisites: BSTA 142

BSTA 375 Analyzing Health Sensor Data 3 Credits
This course will convey specialized methodologies of data collection and the statistical analysis of health-related time-series data collected from sensors. Of particular interest are data generated by environmental sensors, wearable devices, and medical instrumentation. Through a series of health-related case studies, students will have the opportunity to explore signal processing, filtering, modeling, and forecasting techniques. Computing will be a significant component of the course, ensuring that students acquire the skills necessary to apply these techniques to health-related sensor data.
Prerequisites: BSTA 142

BSTA 376 Deep Learning for Healthcare 3 Credits
This course will convey the specialized methods of deep learning in the context of health data. Through health-related case studies, students will learn to engage deep learning models and healthcare applications such as clinical predictive models, computational phenotyping, patient risk stratification, treatment recommendation, and medical imaging analysis. The course will engage with real-world data sets via computing using Jupyter and PyTorch, ensuring that students acquire the skills necessary to apply deep learning techniques to health data.
Prerequisites: BSTA 142

BSTA 381 Analysis of Dependent Data 3 Credits
This course will convey specialized methodologies needed to analyze and model dependent data. By considering dependent data from a series of health-related case studies, students will have the opportunity to explore different types of statistical association, random effects models, generalized estimating equations, copula models, and nonparametric methods for dependent data. Computing will be a significant component of the course, ensuring that students acquire the skills necessary to carry out a wide range of analyses of health-related dependent data.
Prerequisites: BSTA 133

BSTA 383 Survival Analysis 3 Credits
This course will present methodologies needed to model time-to-event data. By considering censored (i.e., incomplete) health data from a series of case studies, students will explore nonparametric estimation (e.g., life table methods, Kaplan–Meier estimator), nonparametric methods for comparing the survival experience of populations, and semiparametric and parametric methods of regression for censored outcome data. Computing will be a significant component of the course, ensuring students acquire the skills necessary to conduct time-to-event analyses of health-related data.
Prerequisites: BSTA 133

BSTA 384 Network Analysis 3 Credits
This course will convey specialized methodologies needed to analyze and model network data. By considering relational data from a series of health-related case studies, students will have the opportunity to explore mathematical description of networks, social network measures, exponential random graph models of networks, network sampling, and visualization. Computing will be a significant component of the course, ensuring that students acquire the skills necessary to carry out a wide range of network-based analyses of health-related data.
Prerequisites: BSTA 133

BSTA 386 Bayesian Analysis 3 Credits
This course will provide a basic introduction to Bayesian concepts and methods with an emphasis on the data analysis in the context of health. We will discuss model choice, including the assessment of prior distributions. We will discuss how to conduct inference in a Bayesian setting, through posterior means, credible intervals and hypothesis testing. The Analyses will be performed using the freely available software Jags as implemented in the R packages rjags and R2jags.
Prerequisites: BSTA 133

BSTA 387 Analyzing Data in SAS 3 Credits
This course will introduce the student to the SAS programming language in a lab-based format. The objective is for the student to develop programming and statistical computing skills to address data management and analysis issues using SAS. The course will also provide a survey of some of the most common data analysis tools in use today and provide decision-making strategies in selecting the appropriate methods for extracting information from data.
Prerequisites: BSTA 133

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

BSTA 399 Portfolio Project 1 Credit
This course will must be taken concurrently with an elective in either the Data or Methods clusters of the program. Students must inform the instructor for the associated elective about their registration in the Portfolio Project course. Portfolio Project students may be assigned additional material/assignments, and will be required to complete a significant report in the associated elective course.

BSTA 402 Biostatistics in Health 3 Credits
This course provides an introduction to the use of statistics in health. Topics include descriptive statistics, probability distributions, parameter estimation, hypothesis testing, analysis of contingency tables, analysis of variance, regression models, and sample size and power considerations. Students develop the skills necessary to perform, present, and interpret statistical analyses; and attain autonomy in handling real-world data using a statistical computing environment.

BSTA 403 Health 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.

BSTA 409 Outbreak Science & Public Health Forecasting 3 Credits
This course aims to introduce students to models that describe the spread of a pathogen through a population, and how models can support public health decisions. The course will be split into four parts: (i) the factors that motivate public health actions, (ii) epidemic models such as the Reed-Frost and SIR, (iii) statistical time series and forecasts, (ii) a focus on ensemble building. Students will be expected to complete mathematical/statistical exercises and write code that simulates infectious processes.

BSTA 410 (CSE 410) Assistive Technologies 3 Credits
This class will introduce typical challenges faced by persons with disabilities and the role of assistive technologies (ATs) in solving such challenges. The class will examine opportunities presented by recent advances in mobile and AI technologies. Working in groups, each student will be expected to acquire and apply relevant skills in designing AT solutions. The class can be taken by students with diverse backgrounds including the following: community and population health, social and behavioral sciences, business, engineering and computer science.

BSTA 420 (CGH 420, POPH 420, PUBH 420) 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 450 Special Topics in Biostatistics 3 Credits
In this course, students will engage in an intensive exploration of a topic of special interest that is not covered in other courses. Topics addressed will be at an advanced level.
Repeat Status: Course may be repeated.