

PH.140 (BIOSTATISTICS)

Courses

Course location and modality is found on the BSPH website (<https://publichealth.jhu.edu/courses/>).

PH.140.604. Introduction to R for Public Health Researchers. 2 Credits.

Provides “hands-on” training for analyzing data in the R statistical software package, a popular open-source solution for data analysis and visualization. Covers data input/output, data management and manipulation, and constructing useful and informative graphics. Geared towards individuals who have never used R or have a little familiarity.

PH.140.605. Introduction to the SAS Statistical Package. 2 Credits.

PH.140.606. Survival Analysis. 2 Credits.

PH.140.607. Multilevel Models. 2 Credits.

Gives an overview of “multilevel statistical models” and their application in public health and biomedical research. Multilevel models are regression models in which the predictor and outcome variables can occur at multiple levels of aggregation: for example, at the personal, family, neighborhood, community and regional levels. They are used to ask questions about the influence of factors at different levels and about their interactions. Multilevel models also account for clustering of outcomes and measurement error in the predictor variables. Students focus on the main ideas and on examples of multi-level models from public health research. Students learn to formulate their substantive questions in terms of a multilevel model, to fit multilevel models using Stata during laboratory sessions and to interpret the results.

PH.140.608. Analysis of Longitudinal Data. 2 Credits.

Covers statistical models for drawing scientific inferences from longitudinal data. Topics include longitudinal study design; exploring longitudinal data; linear and generalized linear regression models for correlated data, including marginal, random effects, and transition models; and handling missing data.

PH.140.609. Improving Precision and Power in Randomized Trials by Leveraging Baseline Variables. 0.5 Credits.

Explains what covariate adjustment is, how it works, when it may be useful to apply, and how to implement it (in a preplanned way that is robust to model misspecification) for a variety of scenarios. Demonstrates the impact of covariate adjustment using trial data sets in multiple disease areas. Provides step-by-step, clear documentation of how to apply the software in each setting. Applies the software tools on the different datasets in small groups.

PH.140.611. Statistical Reasoning in Public Health I. 3 Credits.

PH.140.612. Statistical Reasoning in Public Health II. 3 Credits.

Provides a broad overview of biostatistical methods and concepts used in the public health sciences, emphasizing interpretation and concepts rather than calculations or mathematical details. Develops ability to read the scientific literature to critically evaluate study designs and methods of data analysis. Introduces basic concepts of statistical inference, including hypothesis testing, p-values, and confidence intervals. Includes topics: comparisons of means and proportions; the normal distribution; regression and correlation; confounding; concepts of study design, including randomization, sample size, and power considerations; logistic regression; and an overview of some methods in survival analysis. Draws examples of the use and abuse of statistical methods from the current biomedical literature.

PH.140.613. Data Analysis Workshop I. 2 Credits.

Emphasizes concepts and illustration of concepts applying a variety of analytic techniques to public health datasets in a computer laboratory using Stata statistical software. Learns basic methods of data organization/management and simple methods for data exploration, data editing, and graphical and tabular displays. Includes additional topics: comparison of means and proportions, simple linear regression and correlation.

PH.140.614. Data Analysis Workshop II. 2 Credits.

Emphasizes concepts and illustration of concepts applying a variety of analytic techniques to public health datasets in a computer laboratory using Stata statistical software. Masters advanced methods of data analysis including analysis of variance, analysis of covariance, nonparametric methods for comparing groups, multiple linear regression, logistic regression, log-linear regression, and survival analysis.

PH.140.615. Statistics for Laboratory Scientists I. 4 Credits.

PH.140.616. Statistics for Laboratory Scientists II. 4 Credits.

PH.140.618. AI Programming in Python for Public Health. 2 Credits.

Explores the transformative potential of Artificial Intelligence (AI) in public health. Aims at public health professionals, researchers, and policymakers, the course delves into AI’s role in disease surveillance, epidemic prediction, healthcare delivery, and health policy. Gains foundational knowledge in AI concepts, machine learning algorithms, and data analytics. Teaches how AI can address public health challenges, enhance disease prevention strategies, and improve health outcomes through case studies, interactive sessions, and hands-on projects. Emphasizes ethical considerations, data privacy, and the equitable application of AI technologies in diverse health settings.

PH.140.620. Advanced Data Analysis Workshop. 2 Credits.

Covers methods for the organization, management, exploration, and statistical inference from data derived from multivariable regression models, including linear, logistic, Poisson and Cox regression models. Students apply these concepts to two or three public health data sets in a computer laboratory setting using STATA statistical software. Topics covered include generalized linear models, product-limit (Kaplan-Meier) estimation, Cox proportional hazards model.

PH.140.621. Statistical Methods in Public Health I. 4 Credits.

PH.140.622. Statistical Methods in Public Health II. 4 Credits.

Presents use of confidence intervals and hypothesis tests to draw scientific statistical inferences from public health data. Introduces generalized linear models, including linear regression and logistic regression models. Develops unadjusted analyses and analyses adjusted for possible confounders. Outlines methods for model building, fitting and checking assumptions. Focuses on the accurate statement of the scientific question, appropriate choice of generalized linear model, and correct interpretation of the estimated regression coefficients and confidence intervals to address the question.

Corequisite(s): Must also register for lab, PH.140.922.

PH.140.623. Statistical Methods in Public Health III. 4 Credits.

Corequisite(s): Must also enroll for PH.140.923

PH.140.624. Statistical Methods in Public Health IV. 4 Credits.

Builds on the concepts, methods, and computing (Stata, R) covered in Statistical Methods 1,2, and 3. Focuses on investigating scientific questions via data analysis and clearly communicating the methodology and results. Uses examples from the contemporary and public health literature and allows students the opportunity to work with their own data over the duration of the class.

Corequisite(s): Must also enrol for a lab, PH.140.924.

PH.140.626. Propensity Score and Related Methods for Estimating Causal Effects. 1 Credit.

Discusses the importance of the careful design of non-experimental studies, and the role of propensity scores and related methods in that design, with the main goal of providing practical guidance on use of sample equating methods. Covers the primary ways of using propensity scores and related methods to adjust for confounders when estimating the effect of a particular "cause" or "intervention," including weighting, subclassification, and matching. Examines issues such as how to specify and estimate a propensity score model, selecting covariates to include in the model, and diagnostics. Draws examples from across public health. Emphasizes non-experimental studies; however, also discusses applications to randomized trials such as examining levels of adherence and generalizability.

PH.140.628. Data Science for Public Health I. 4 Credits.

Presents the basics of data science using the R programming language. Teaches basic unix, version control, graphing and plotting techniques, creating interactive graphics, web app development, reproducible research tools and practices, resampling based statistics and artificial intelligence via deep learning, focusing on practical implementation specifically tied to computational tools and core fundamentals necessary for practical implementation. Culminates with a web app development project chosen by student (who will come out of this course sequence well-equipped to tackle many of the data science problems that they will see in their research).

PH.140.629. Data Science for Public Health II. 4 Credits.

Presents the basics of data science using the R programming language. Teaches basic unix, version control, graphing and plotting techniques, creating interactive graphics, web app development, reproducible research tools and practices, resampling based statistics and artificial intelligence via deep learning, focusing on practical implementation specifically tied to computational tools and core fundamentals necessary for practical implementation. Culminates with a web app development project chosen by student (who will come out of this course sequence well-equipped to tackle many of the data science problems that they will see in their research).

PH.140.630. Introduction to Data Management. 3 Credits.

Introduces students to the principles and skills required to collect and manage research data in a public health setting. Focuses on tools for collecting data that range from spreadsheets to web-based systems, database fundamentals, data collection form design, data entry screen design, proper coding of data, strategies for quality control and data cleaning, protection and sharing of data, and integrating data from external sources. Includes practical and hands-on exercises that require some entry-level computer programming.

PH.140.631. The SAS Statistical Package: A Survey for Statisticians. 3 Credits.

Introduces students to the SAS statistical package using the SAS Studio interface. Using examples of public health data students learn to write programs to summarize data and to perform statistical analyses. Using the interactive matrix language introduces computation within a matrix environment and the development of modular programming techniques.

PH.140.632. Introduction to the SAS Statistical Package. 3 Credits.

Introduces students with no experience with SAS. Familiarizes them with the skills needed for effective data management and data analysis. Covers performing exploratory analysis on data including the creation of tables and graphs. Proceeds next to creating new datasets and altering old datasets. Covers building regression models (linear, logistic, and Poisson), interpreting results and criticizing such models, and attempting to improve them.

PH.140.633. Biostatistics in Medical Product Regulation. 2 Credits.

PH.140.634. Non-Inferiority and Equivalence Clinical Trials. 2 Credits. Presents the important differences between superiority trials and those intended to show either equivalent effect, or to show that one therapy is no worse than another (but might be better). Explores the problems of setting equivalence margins, preservation of some proportion of active control effect, and emphasizes the use of confidence intervals to interpret the results of studies. Discusses special issues of quality of the trial conduct, assay sensitivity, historical evidence of treatment effects and assumptions of constancy of treatment effects over time. Compares sample size requirements between superiority trials, equivalence trials and non-inferiority trials. Discusses the use of different analysis populations (ITT and per-protocol) and issues of changing conclusions between non-inferiority and superiority. Discusses the regulatory aspects of trial design and interpretation, and reviews existing regulatory guidance.

PH.140.635. Practical Genomics: Computational Tools for Cancer Research. 2 Credits.

Provides an introduction to computational analysis of genomics datasets with applications in cancer research. Includes hands-on training in organizing, analyzing, and visualizing data using R, RStudio, and Bioconductor. Covers data manipulation, single-cell genomics, and differential gene expression analysis. Features live coding demonstrations, guided exercises, and capstone projects. Emphasizes real-world examples relevant to cancer biology and practical skills for integrating computational genomics into research workflows.

PH.140.636. Scalable Computational Bioinformatics. 4 Credits.

Introduces the computational hardware and programming model upon which analysis tools and languages are based. Introduces and uses three main languages (Python, Perl, SQL) and their underlying rationale to develop computer science concepts such as data structures, algorithms, computational complexity, regular expressions, and knowledge representation. Draws examples and exercises from high-throughput sequence analysis, proteomics and modeling of biological systems. Reinforces key concepts through lectures with live computer demonstrations, weekly readings, and programming exercises. Has students working with a High Performance Compute Cluster and the Amazon cloud.

PH.140.638. Analysis of Biological Sequences. 3 Credits.**PH.140.640. Statistical Methods for Sample Surveys. 3 Credits.****PH.140.641. Survival Analysis. 3 Credits.**

Introduces fundamental concepts, theory and methods in survival analysis. Emphasizes statistical tools and model interpretations which are useful in medical follow-up studies and in general time-to-event studies. Includes hazard function, survival function, different types of censoring, Kaplan-Meier estimate, log-rank test and its generalization. For parametric inference, includes likelihood estimation and the exponential, Weibull, log-logistic and other relevant distributions. Discusses in detail statistical methods and theory for the proportional hazard models (Cox model), with extensions to time-dependent covariates. Includes clinical and epidemiological examples (through class presentations). Introduces basic concepts and methods for competing risks data, including the cause-specific hazard models and other models based of cumulative incidence function (CIF). Illustrates various statistical procedures (through homework assignments).

PH.140.642. Design of Clinical Experiments. 3 Credits.

PH.140.643. Practice of Statistical Consulting. 3 Credits.

Emphasizes the understanding of, and practical experience in, the spectrum of non-technical aspects of statistical consulting, the art and science of applying statistics to real-world problems. Discusses the elements of a consultation, from defining the research problem to providing final products to the client, interpersonal communication, reproducible work, ethics and consulting in different environments. Develops students' consulting skills via lectures, role-play opportunities, consulting sessions, and actual research projects. Acquaints students with practical consulting experience through shadowing and leading the Biostatistics Center's clinics on Friday mornings. Provides opportunities to work directly with Johns Hopkins researchers to elicit information about the research question, and to provide a presentation and final report to researchers.

PH.140.644. Statistical Machine Learning: Methods, Theory, and Applications. 4 Credits.

Introduces popular Machine Learning methods and emphasizes their practical usage for data analysis. Acquaints students with methods to evaluate statistical machine learning models defined in terms of algorithms or function approximations using basic coverage of their statistical and computational theoretical underpinnings. Topics covered include: regression and prediction, tree-based methods, overview of supervised learning theory, support vector machines, kernel methods, ensemble methods, clustering, visualization of large datasets and graphical models. Examples of method applications covered include cancer prognosis from microarray data, visualization and analysis of social network data, and graphical models for clinical decision-making.

PH.140.646. Essentials of Probability and Statistical Inference I: Probability. 4 Credits.**PH.140.647. Essentials of Probability and Statistical Inference II: Statistical Inference. 4 Credits.**

Introduces students to the theory of statistical inference. Includes the frequentist, Bayesian and likelihood approaches to statistical inference including estimation, testing hypotheses and interval estimation. Emphasizes rigorous analysis (including proofs), as well as interpretation of results and simulation for illustration.

PH.140.648. Essentials of Probability and Statistical Inference III: Theory of Modern Statistical Methods. 4 Credits.**PH.140.649. Essentials of Probability and Statistical Inference IV. 4 Credits.**

Builds on the concepts discussed in 140.646, 140.647 and 140.648 to lay out foundation for both classical and modern theory/methods for drawing statistical inference. Includes classical unbiased estimation, unbiased estimating equations, likelihood and conditional likelihood inference, information theory and other extended topics. Includes mathematical proofs but will not emphasize highly technical details. Provides extended discussions, interpretation of results, and examples for illustration.

PH.140.650. AI Methods for Geospatial Data. 3 Credits.

Introduces advanced AI methods for analyzing large-scale geospatial data, with particular emphasis on geostatistics. Starts with an overview of geospatial data and exploratory data analysis and visualization techniques. Delves next into Gaussian Processes (GP) and kriging for spatial data modeling, covering both classical optimization and Bayesian MCMC methods for GP models. Covers spatial graphical models and Nearest Neighbor Gaussian Processes (NNGP) for handling massive datasets. Introduces machine learning techniques, including random forests and neural networks (multi-layer perceptrons, convolutional and graph neural networks), and explores hybrid methods that combine traditional statistical modeling with machine learning. Covers various state-of-the-art computational techniques, like stochastic approximations and variational Bayesian optimization, and offers a hands-on demonstration of analysis of big spatial data using R and Python.

PH.140.651. Methods in Biostatistics I. 4 Credits.

Presents fundamental concepts in applied probability, exploratory data analysis, and statistical inference, focusing on probability and analysis of one and two samples. Includes topics discrete and continuous probability models; expectation and variance; central limit theorem; inference, including hypothesis testing and confidence interval for means, proportions, and counts; maximum likelihood estimation; sample size determinations; elementary non-parametric methods; graphical displays; and data transformations. Introduces R and concepts are presented both from a theoretical, practical and computational perspective.

PH.140.652. Methods in Biostatistics II. 4 Credits.

Presents fundamental concepts in applied probability, exploratory data analysis, and statistical inference, focusing on probability and analysis of one and two samples. Includes discrete and continuous probability models; expectation and variance; central limit theorem; inference, including hypothesis testing and confidence for means, proportions, and counts; maximum likelihood estimation; sample size determinations; elementary non-parametric methods; graphical displays; and data transformations.

PH.140.653. Methods in Biostatistics III. 4 Credits.**PH.140.654. Methods in Biostatistics IV. 4 Credits.****PH.140.655. Analysis of Multilevel and Longitudinal Data. 4 Credits.**

Explores statistical models for drawing scientific inferences from multilevel and longitudinal public health data. Includes topics: multilevel causes in public health, longitudinal as a leading example of multilevel data, study design, exploring multilevel and longitudinal data; linear and generalized linear regression models for correlated data including marginal, random effects, and transition models; and handling missing data.

PH.140.656. Multilevel and Longitudinal Models - Data Analysis Workshop. 4 Credits.

Explores conceptual and formal approaches to the design, analysis, and interpretation of studies with a "multilevel" or "hierarchical" (clustered) data structure (e.g., individuals in families in communities). Develops skills to implement and interpret random effects, variance component models that reflect the multi-level structure for both predictor and outcome variables. Includes topics: building hierarchies; interpretation of population-average and level-specific summaries; estimation and inference based on variance components; shrinkage estimation; discussion of special topics including centering, use of contextual variables, ecological bias, sample size and missing data within multilevel models. Supports STATA and R software.

PH.140.658. Statistics for Psychosocial Research: Structural Models. 4 Credits.**PH.140.664. Causal Inference in Medicine and Public Health I. 4 Credits.**

Presents an overview of methods for estimating causal effects: how to answer the question of “What is the effect of A on B?” Includes discussion of randomized designs, but with more emphasis on alternative designs for when randomization is infeasible: matching methods, propensity scores, regression discontinuity, and instrumental variables. Methods are motivated by examples from the health sciences, particularly mental health and community or school-level interventions.

PH.140.665. Causal Inference in Medicine and Public Health II. 3 Credits.

Presents principles, methods, and applications in drawing cause-effect inferences with a focus on the health sciences. Building on the basis of 140.664, emphasizes statistical theory and design and addresses complications and extensions, aiming at cultivating students’ research skills in this area. Includes: detailed role of design for causal inference; role of models and likelihood perspective for ignorable treatment assignment; estimation of noncollapsible causal effects; statistical theory of propensity scores; use of propensity scores for estimating effect modification and for comparing multiple treatments while addressing regression to the mean; theory and methods of evaluating longitudinal treatments, including the role of sequentially ignorable designs and propensity scores; likelihood theory for instrumental variables and principal stratification designs and methods to deal with treatment noncompliance, direct and indirect effects, and censoring by death.

PH.140.673. INTRODUCTION TO STATISTICAL THEORY I. 4 Credits.**PH.140.686. Advanced Methods for Statistical Genetics and Genomics. 3 Credits.**

Covers statistical methods and theory underlying advanced analysis of genetic and genomic data to address mechanistic hypotheses and to build models for prediction. Topics include methods for complex association testing, inference on genetic architecture using mixed model techniques, methods for understanding causal mechanisms using Mendelian randomization, and integrative genomic analysis and strategies for clinical translation using risk prediction models. Requires making presentations and critiquing published studies that have used advanced statistical methods to make new scientific observations.

PH.140.688. Statistics For Genomics. 3 Credits.

Covers the basics of R software and the key capabilities of the Bioconductor project (a widely used open source and open development software project for the analysis and comprehension of data arising from high-throughput experimentation in genomics and molecular biology and rooted in the open source statistical computing environment R), including importation and preprocessing of high-throughput data from microarrays and other platforms. Also introduces statistical concepts and tools necessary to interpret and critically evaluate the bioinformatics and computational biology literature. Includes an overview of preprocessing and normalization, statistical inference, multiple comparison corrections, Bayesian Inference in the context of multiple comparisons, clustering, and classification/machine learning.

PH.140.689. Adaptive Enrichment Designs for Confirmatory Randomized Trials: Methods and Software. 1 Credit.

Provides an overview of the strengths and limitations of randomized trial designs that adaptively change enrollment criteria during a trial (adaptive enrichment designs) and have the potential to provide improved information about which subpopulations benefit from new treatments. Explains recent advances in statistical methods for these designs, and presents adaptive design software planning tools. Discusses FDA guidance documents on adaptive designs. Examines methods for improving precision of estimators of the average treatment effect, by leveraging information in baseline variables; these methods can be used in adaptive designs as well as standard (non-adaptive) trial designs.

PH.140.690. Artificial Intelligence for Improved Patient Outcomes. 1 Credit.

Equips students with the essential skills to build and evaluate AI and predictive modeling tools in medicine. Emphasizes practical implementation and rigorous evaluation to address unique challenges in healthcare. Addresses the limits of AI’s potential to benefit patients and presents actionable insights to overcome these challenges.

PH.140.698. Spatial Analysis III: Spatial Statistics. 4 Credits.

Introduces statistical techniques used to model, analyze, and interpret public health-related spatial data. Casts analysis of spatially dependent data into a general framework based on regression methodology. Covers the geostatistical techniques of kriging and semivariogram analysis, point process methods for spatial case-control data and area-level analysis. Focuses on statistical modeling (although some time will be spent covering topics related to cluster detection of health outcome events). Provides instruction in the public domain statistical computing environment R/RStudio.

PH.140.699. Spatial Analysis IV: Spatial Design and Application. 3 Credits.

Expands students’ abilities to design, conduct and report the results of a complete public health related spatial analysis. Focuses on further developing and integrating components of the spatial science paradigm, Spatial Data, GIS and Spatial Statistics. Introduces relevant topics in GIS, spatial data technologies and spatial statistics not previously covered in Spatial Analysis I-III.

PH.140.711. Advanced Data Science I. 3 Credits.

In this course, we will focus on hands-on data analyses with a main objective of solving real-world problems. We will teach the necessary skills to gather, manage and analyze data using the R programming language. We will cover an introduction to data wrangling, exploratory data analysis, statistical inference and modeling, machine learning, and high-dimensional data analysis. We will also learn the necessary skills to develop data products including reproducible reports that can be used to effectively communicate results from data analyses. Students will train to become data scientists capable of both applied data analysis and critical evaluation of the next generation next generation of statistical methods.

PH.140.712. Advanced Data Science II. 3 Credits.

Builds on Advanced Data Science I by introducing the idea of data products and encouraging students to build products based on their data analyses.

PH.140.721. Probability Theory I. 3 Credits.

Presents the first part of the classical results of probability theory: measure spaces, LP spaces, probability measures, distributions, random variables, integration, and convergence theorems.

PH.140.722. Probability Theory II. 3 Credits.

Presents the first part of the classical results of probability theory: independence, types of convergence, laws of large numbers, Borel-Cantelli lemmas, Kolmogorov's zero-one law, random series and rates of convergence. Also discusses characteristic functions and weak convergence.

PH.140.723. Probability Theory III. 3 Credits.

Presents the second part of the classical results of probability theory: central limit theorems, Poisson convergence, coupling, Stein-Chen method, densities, derivatives and conditional expectations.

PH.140.724. Probability Theory IV. 3 Credits.

Covers basic stochastic processes including martingales and Markov chains, followed by consideration of Markov Chain Monte Carlo (MCMC) methods.

PH.140.731. Statistical Theory I. 4 Credits.

Introduces probability and inference, including random variables; probability distributions; transformations and sums of random variables; expectations, variances, and moments; properties of random samples; and hypothesis testing.

PH.140.732. Statistical Theory II. 4 Credits.

Introduces modern statistical theory; sets principles of inference based on decision theory and likelihood (evidence) theory; derives the likelihood function based on design and model assumptions; derives the complete class theorem between Bayes and admissible estimators; derives minimal sufficient statistics as a necessary and sufficient reduction of data for accurate inference in parametric models; derives the minimal sufficient statistics in exponential families; introduces maximum likelihood and unbiased estimators; defines information and derives the Cramer-Rao variance bounds in parametric models; introduces empirical Bayes (shrinkage) estimators and compares to maximum likelihood in small-sample problems.

PH.140.733. Statistical Theory III. 4 Credits.

Derives the large sample distribution of the maximum likelihood estimator under standard regularity conditions; develops the delta method and the large sample distribution of functions of consistent estimators, including moment estimators; introduces the theory of estimation in semiparametric regression models based on increasing approximation of parametric models; develops likelihood intervals and confidence intervals with exact or approximate properties; develops hypothesis tests through decision theory.

PH.140.734. Statistical Theory IV. 4 Credits.

Focuses on the asymptotic behavior of estimators, tests, and confidence interval procedures. Specific topics include: M-estimators; consistency and asymptotic normality of estimators; influence functions; large-sample tests and confidence regions; nonparametric bootstrap

PH.140.741. Advanced Survival Analysis. 3 Credits.

Introduces statistical models and methods useful for analyzing univariate and multivariate failure time data. Extends Survival Analysis I to topics on length-bias and prevalent samplings, martingale theory, multivariate survival data, time-dependent ROC analysis, and recurrent event processes. Emphasizes nonparametric and semiparametric approaches for modeling, estimation and inferential results. Clinical and epidemiological examples included in class presentation illustrate statistical procedures.

PH.140.742. Risk Prediction and Precision Medicine. 3 Credits.

Gives an overview of study designs and methodologies for building and evaluation of risk prediction models. Reviews fundamental concepts of epidemiologic cohort and case-control studies and related modeling approaches through Cox proportional hazard models and logistic regressions. Introduces concepts of absolute risk modeling within the Cox proportional hazard framework and incorporation of different sources data from research studies and population registries. Includes methodologies for building polygenic risk scores from recent genome-wide association studies (GWAS) and enhancing transportability of the scores to diverse populations. Focuses on model validation and evaluation for clinical utility and illustrates the concepts through a number of real-world examples.

PH.140.751. Advanced Methods in Biostatistics I. 3 Credits.**PH.140.752. Advanced Methods in Biostatistics II. 4 Credits.**

Surveys basic statistical inference, estimates, tests and confidence intervals, and exploratory data analysis. Reviews probability distributions and likelihoods, independence and exchangeability, and modes of inference and inferential goals including minimizing MSE. Reviews linear algebra, develops the least squares approach to linear models through projections, and discusses connections with maximum likelihood. Covers linear, least squares regression, transforms, diagnostics, residual analysis, leverage and influence, model selection for estimation and predictive goals, departures from assumptions, efficiency and robustness, large sample theory, linear estimability, the Gauss Markov theorem, distribution theory under normality assumptions, and testing a linear hypothesis.

PH.140.753. Advanced Methods in Biostatistics III. 4 Credits.

Introduces generalized linear model (GLM). Foundational topics include: contingency tables, logistic regression for binary and binomial data, models for polytomous data, Poisson log-linear model for count data, and GLM for exponential family. Introduces methods for model fitting, diagnosis, interpretation and inference and expands on those topics with techniques for handling overdispersion, quasi-likelihood and conditional likelihood. Introduces the role of quantitative methods and sciences in public health, including how to use them to describe and assess population health, and the critical importance of evidence in advancing public health knowledge.

PH.140.754. Advanced Methods in Biostatistics IV. 4 Credits.

Extends topics in 140.753 to encompass generalized linear mixed effects models. Introduces expectation-maximization and Markov Chain Monte Carlo. Introduces functional data analysis. Foundational topics include: linear mixed model, generalized linear mixed model, EM, MCMC, models for longitudinal data, and functional data analysis. Emphasizes both rigorous methodological development and practical data analytic strategies. Discusses the role of quantitative methods and sciences in public health, including how to use them to describe and assess population health, and the critical importance of evidence in advancing public health knowledge.

PH.140.762. Bayesian Methods I. 3 Credits.

Illustrates current approaches to Bayesian modeling and computation in statistics. Describes simple familiar models, such as those based on normal and binomial distributions, to illustrate concepts such as conjugate and noninformative prior distributions. Discusses aspects of modern Bayesian computational methods, including Markov Chain Monte Carlo methods (Gibbs' sampler) and their implementation and monitoring. Bayesian Methods I is the first term of a two term sequence. The second term offering, Bayesian Methods II (140.763), develops models of increasing complexity, including linear regression, generalized linear mixed effects, and hierarchical models.

PH.140.772. Advanced Statistical Theory. 4 Credits.

Examines statistics as a discipline along the path towards making decisions. First examines the justification of statistics from axioms on informed preferences and its close connection to Bayesian theory, and then examines the role of standardizing intermediate steps, through various additional restrictions on estimation, and studies the properties of the resulting methods.

PH.140.773. Foundations of Statistical Inference I. 4 Credits.

Investigates the foundations of statistics as applied to assessing the evidence provided by an observed set of data. Topics include: law of likelihood, the likelihood principle, evidence and the likelihood paradigm for statistical inference; failure of the Neyman-Pearson and Fisherian theories to evaluate evidence; marginal, conditional, profile and other likelihoods; and applications to common problems of inference.

PH.140.774. Foundations of Statistics II. 4 Credits.**PH.140.776. Statistical Computing. 3 Credits.****PH.140.777. Statistical Programming Paradigms and Workflows. 3 Credits.**

Covers the basics of statistical programming and other workflow skills required for the research and application of statistical methods. Includes programming with unix and the command line, git/github, working with Python, SQL, APIs, HTMLs, and interactive dashboard. Includes topics in statistical data analysis provide working examples.

PH.140.778. Statistical Computing, Algorithm, and Software Development. 3 Credits.

Teaches students common algorithms and essential skill sets for statistical computing and software development through hands-on experiences. Takes a large-scale logistic regression as an example and has students work toward implementing a high-performance `hiperLogit` R package for fitting this model. Presents progressively advanced algorithms and computing techniques. Trains students in various best practices for developing statistical software, including how to start with a basic version of the package and progressively integrate more advanced features. Prepares students for further training in statistical computing techniques and algorithms as covered in Advanced Statistical Computing (140.779).

PH.140.779. Advanced Statistical Computing. 3 Credits.

Covers the theory and application of common algorithms used in statistical computing. Includes topics: root finding, optimization, numerical integration, Monte Carlo, Markov chain Monte Carlo, stochastic optimization, and bootstrapping. Discusses specific algorithms: Newton-Raphson, EM, Metropolis-Hastings algorithm, Gibbs sampling, simulated annealing, Gaussian quadrature, Romberg integration, etc. Discusses applications of these algorithms to real research problems.

PH.140.800. MPH Capstone Biostatistics. 2 Credits.

The MPH Capstone is an opportunity for students to work on public health practice projects that are of particular interest to them. The goal is for students to apply the skills and competencies they have acquired to a public health problem that simulates a professional practice experience.

PH.140.801. Baltimore Community Data Science. 1.5 Credits.

Works in teams with a community-based organization (CBO) to: develop data science products (such as a dashboard, data analysis, series of visualizations, etc.), develop training material for CBO users to use and implement the products, develop sustainability/maintenance plans for CBOs to continue with the data science product in the future and how they can continue to use more data science methods more generally.

PH.140.802. Baltimore Community Data Science II. 1.5 Credits.

Works in teams with a community-based organization (CBO) to: develop data science products (such as a dashboard, data analysis, series of visualizations, etc.), develop training material for CBO users to use and implement the products, develop sustainability/maintenance plans for CBOs to continue with the data science product in the future and how they can continue to use more data science methods more generally.

PH.140.820. Thesis Research Biostatistics. 1 - 22 Credits.**PH.140.828. Teaching in Biostatistics. 1 - 22 Credits.**

Teaching Assistant (TA) for PhD students in Biostatistics

PH.140.830. Postdoctoral Research Biostatistics. 1 - 22 Credits.**PH.140.840. Special Studies and Research Biostatistics. 1 - 22 Credits.****PH.140.850. Advanced Special Topics in Biostatistics. 1 - 22 Credits.**

Exposes Biostatistics PhD students to advanced special topics that are not covered in the core courses. Comprises two- and four-week modules, with revolving instructors and topics. Possible topics include: theory underlying analysis for correlated data; latent variable modeling; advanced survival analysis; image analysis; time series; and likelihood inference.

PH.140.860. Current Topics in Biostatistics Research. 1 Credit.

Features presentations by Biostatistics faculty, postdocs and senior students on their research, with a focus on the public health and scientific questions driving the work, why the research makes a difference for the subject area and how to translate the research into practice. Offers an opportunity for discussion and clarification of key Biostatistical concepts being taught in the core courses and how they apply to problems in public health and science. Provides an opportunity for students and faculty to come together and discuss novel research questions and the role that Biostatisticians have in helping to support, enrich and promote solutions to these novel research questions.

PH.140.895. MPH Practicum: Biostatistics. 1 - 4 Credits.**PH.140.922. Lab for Biostats 140.622.**

Corequisite(s): Lab for PH.140.622

PH.140.923. Lab for Biostats PH.140.623.

Corequisite(s): Must also enroll for PH.140.623

PH.140.924. Lab for Biostatistics 140.624.

Corequisite(s): Must also enrol for PH.140.624

PH.140.936. Lab - Scalable Computational Bioinformatics.**PH.140.941. Biostats Lab for 140.641.****PH.140.955. Lab for Biostat 140.655.****PH.140.958. Biostats Lab for 140.658.****PH.140.998. Lab for PH.140.698.**