

# BIOSTATISTICS

## **BIS 505b, Biostatistics in Public Health II** Michael Wininger

This continuation of EPH 505 introduces students to regression-based methods for analyzing public health data. Topics include analysis of variance, linear regression, logistic regression, Poisson regression, survival analysis, and longitudinal regression models. Students develop hands-on R computing skills to perform the analyses discussed. Prerequisite: EPH 505. Not open to auditors.

## **[ BIS 515, Accelerated Biostatistics ]**

This intensive seven-week summer course provides a comprehensive introduction to the use of statistics in the fields of epidemiology, public health, and clinical research. Students gain experience conducting and interpreting a broad range of statistical analyses. Topics include descriptive statistics, rules of probability, probability distributions, parameter estimation, hypothesis testing, sample size estimation, analysis of variance, nonparametric tests, and linear regression. Through computer laboratory sessions, students become familiar with the SAS statistical software package. Enrollment limited to students in the Advanced Professional M.P.H. and Accelerated M.B.A./M.P.H. programs. Not open to auditors. 2 Course cr

## **BIS 525a and BIS 526b, Seminar in Biostatistics and Journal Club** Staff

The BIS departmental seminar fosters engagement with innovative statistical researchers outside Yale and exposes students to new ideas in statistical research that they may not encounter in their traditional course work. Topics discussed in seminar talks vary, but a major theme is statistical-methodological innovation in the service of public health. Although no credit or grade is awarded, satisfactory performance will be noted on the student's transcript. 0 Course cr per term

## **[ BIS 534, Stochastic Models and Inference for the Biomedical and Social Sciences ]**

This course covers a diverse array of stochastic processes that serve as mechanistic models for processes in biology, medicine, public health, social science, operations, and economics. For each model, we study theoretical properties and simulation with primary emphasis on statistical inference for model parameters from realizations under different observation scenarios. Each topic in the course is illustrated by numerous applications to empirical data. Topics include models for infectious disease dynamics, physiology, evolution/phylogenetics, health care operations, social networks, and collective behavior. 1 Course cr

## **BIS 536b, Measurement Error and Missing Data** Xin Zhou

The course presents methods for the analysis of data with measurement error or missing data. This course can be divided into two parts. The first part provides an exposition to the statistical theory and the analytic techniques used for adjusting estimates and inference for covariate measurement error and misclassification. The second part covers data analysis with missing data. Much emphasis is placed on likelihood-based approaches to missing data, for example, the Expectation-Maximization (EM) algorithm and multiple imputation (MI). SAS/R is used for analysis of data. Prerequisites: S&DS 541 and S&DS 542 or equivalent, or permission of the instructor. This course is intended for biostatistics graduate students in the second year and above and requires knowledge of, and comfort with, general mathematical

statistics. Prior exposure to asymptotic theory, survival analysis, and/or Bayesian statistics is desirable but not required. Some basic statistical programming skills will also be helpful.

**BIS 537a, Statistical Methods for Causal Inference** Fan Li

This course formally introduces statistical theory and methods that allow rigorous comparisons of treatment strategies for public health and biomedical studies. Although randomization is the gold standard for unbiased treatment comparisons, observational studies are increasingly common for comparative effectiveness research for real-world evidence (RWE). The course addresses complexities in the design and analysis of observational studies for the purpose of comparing treatments. We focus on the treatment effect averaged over a target population as the parameter of scientific interest, and we discuss conditions when the parameter can be interpreted as causal. Modern statistical tools for inferring causality are developed and demonstrated. In the first half of the course, we formalize the comparison of a point treatment in cross-sectional observational studies; and we develop regression, propensity score subclassification, matching, weighting, and hybrid estimators. In the second half, we turn to the more complex time-varying treatments in longitudinal observational studies and introduce methods to account for time-dependent confounding and censoring bias. We explain why traditional regression adjustment fails and discuss the methods of g-computation, sequential stratification, marginal structural models, and structural nested models. Examples are drawn from various biomedical and health-related studies. Prerequisites: S&DS 542 and BIS 623 or their equivalents.

**BIS 540b, Fundamentals of Clinical Trials** Tassos Kyriakides

This course addresses issues related to the design, conduct, analysis, and interpretation of clinical trials. Topics include protocol development, examination and selection of appropriate experimental design, methods of randomization, sample size determination, appropriate methods of data analysis including time-to-event (possibly censored) data, non-inferiority studies, and interim monitoring and ethical issues. Prerequisites: EPH 505 or equivalent, and second-year status.

[ **BIS 542E, Introduction to Health Informatics** ]

The course provides an introduction to clinical and translational informatics. Topics include (1) overview of biomedical informatics, (2) design, function, and evaluation of clinical information systems, (3) clinical decision-making and practice guidelines, (4) clinical decision support systems, (5) informatics support of clinical research, (6) privacy and confidentiality of clinical data, (7) standards, and (8) topics in translational bioinformatics. Open only to students enrolled in the Executive Online M.P.H. Program. Not open to auditors. 1 Course cr

**BIS 543Eb, Topics in Biomedical Informatics and Data Science** Samah Jarad and Kei-Hoi Cheung

The course focuses on providing an introduction to common unifying themes that serve as the foundation for different areas of biomedical informatics, including clinical, neuro-, and genome informatics. The course is designed for students with significant computer experience and course work who plan to build databases and computational tools for use in biomedical research. Emphasis is on understanding basic principles underlying informatics approaches to interoperability among biomedical databases and software tools, standardized biomedical vocabularies and ontologies, biomedical natural language processing, modeling of biological systems, high-performance computation in

biomedicine, and other related topics. Open only to students enrolled in the Executive Online M.P.H. Program. Not open to auditors.

**BIS 544Ea, Computational Methods for Informatics** Robert McDougal

This course introduces the key computational methods and concepts necessary for taking an informatics project from start to finish: using APIs to query online resources, reading and writing common biomedical data formats, choosing appropriate data structures for storing and manipulating data, implementing computationally efficient and parallelizable algorithms for analyzing data, and developing appropriate visualizations for communicating health information. The FAIR data-sharing guidelines are discussed. Current issues in big health data are discussed, including successful applications as well as privacy and bias concerns. This course has a significant programming component, and familiarity with programming is assumed. Open only to students enrolled in the Executive Online M.P.H. Program. Not open to auditors.

**BIS 550b, Topics in Biomedical Informatics and Data Science** Samah Jarad

The course focuses on providing an introduction to common unifying themes that serve as the foundation for different areas of biomedical informatics, including clinical, neuro-, and genome informatics. The course is designed for students with basic computer experience and course work who plan to build databases and computational tools for use in biomedical research. Emphasis is on understanding basic principles underlying informatics approaches to interoperation among biomedical databases and software tools, standardized biomedical vocabularies and ontologies, biomedical natural language processing, modeling of biological systems, high-performance computation in biomedicine, and other related topics.

**BIS 555a, Machine Learning with Biomedical Data** Leying Guan

This course covers many popular topics in machine learning and statistics that are widely used for the exploration of biomedical data. Techniques covered include different linear prediction methods, random forest, boosting, neural networks, and some recent progress on model inference in high dimensions, as well as dimension reduction and clustering. Various examples using biomedical data – e.g., microarray gene expression data, single-cell RNA-Seq data – are provided. The emphasis is on the statistical aspects of different machine-learning methods and their applications to problems in computational biology. Prerequisites: S&DS 542 (or S&DS 612) and BIS 623 (or S&DS 612). This course assumes prior knowledge of statistical inference and regression. It also involves programming, and knowledge of R or Python is required.

**BIS 560a, Introduction to Health Informatics** Andrew Taylor

Health Informatics is a diverse and varied field. This course is designed to provide a general introduction to health informatics. Students will gain foundational knowledge in clinical information systems, health data standards, electronic health records and data security/privacy issues, among other areas. Students will survey a variety of informatics subfields including research, laboratory/precision medicine, imaging, and artificial intelligence. A particular focus for the course will be conceptual underpinnings that generalized well to all informatics disciplines. Permission of the instructor required.

**BIS 562a, Clinical Decision Support** Edward Melnick and Mona Sharifi

Building on BIS 560/CB&B 740, this course provides the purpose, scope, and history of decision support systems within health care. Using a weekly hands-on application of

knowledge acquired in the lecture portion of the course, students identify a clinical need and prototype their own clinical decision support solution. Solutions are then presented in a “shark tank” format to iteratively refine them to yield a final product that is capable of real-world implementation. Prerequisite: BIS 560 or CB&B 740.

**BIS 567a, Bayesian Statistics** Joshua Warren

Bayesian inference is a method of statistical inference in which prior beliefs for model parameters can be incorporated into an analysis and updated once data are observed. This course is designed to provide an introduction to basic aspects of Bayesian data analysis including conceptual and computational methods. Broad major topics include Bayes’s theorem, prior distributions, posterior distributions, predictive distributions, and Markov chain Monte Carlo sampling methods. We begin by motivating the use of Bayesian methods, discussing prior distribution choices in common single parameter models, and summarizing posterior distributions in these settings. Next, we introduce computational methods needed to study multi-parameter models. R software is most often used. We then apply these methods to more complex modeling settings including linear, generalized linear, and hierarchical models. Discussion of model comparisons and adequacy is also presented.

**BIS 568a, Applied Machine Learning in Healthcare** Andrew Taylor and Wade Schulz

Recent advances in machine learning (ML) offer tremendous promise to improve the care of patients. However, few ML applications are currently deployed within healthcare institutions and even fewer provide real value. This course is designed to empower students to overcome common pitfalls in bringing ML to the bedside and aims to provide a holistic approach to the complexities and nuances of ML in the healthcare space. The class focuses on key steps of model development and implementation centered on real-world applications. Students apply what they learn from the lectures, assignments, and readings to identify salient healthcare problems and tackle their solutions through end-to-end data engineering pipelines. Students are expected to be proficient in programming (R, Python, or Julia preferred) and have some prior experience in machine learning including data preprocessing (e.g., Python-Pandas, R-Tidyverse) and the development and validation of ML models (e.g. logistic regression, random forest, XGBoost). Otherwise, permission of the instructor is required.

**BIS 575a, Introduction to Regulatory Affairs** Robert Makuch

This course provides students with an introduction to regulatory affairs science, as these issues apply to the regulation of food, pharmaceuticals, and medical and diagnostic devices. The course covers a broad range of specialties that focus on issues including legal underpinnings of the regulatory process, compliance, phases of clinical testing and regulatory milestones, clinical trials design and monitoring, quality assurance, post-marketing study design in response to regulatory and other needs, and post-marketing risk management. The complexities of this process require awareness of leadership and change management skills. Topics to be discussed include: (1) the nature and scope of the International Conference on Harmonization, and its guidelines for regulatory affairs in the global environment; (2) drug development, the FDA, and principles of regulatory affairs in this environment; (3) the practice of global regulatory affairs from an industry perspective; (4) description/structure/issues of current special importance to the U.S. FDA; (5) historical background and FDA jurisdiction of food and drug law; (6) the drug development process including specification of the important milestone meetings with the FDA; (7) risk analysis and approaches to its

evaluation; (8) use of Bayesian statistics in medical device evaluation, a new approach; (9) use of data monitoring committees and other statistical methods for regulatory compliance; (10) developments in leadership and change management; and (11) food quality assurance including risk analysis/compliance/enforcement. Through course participation, students also have opportunities to meet informally with faculty and outside speakers to explore additional regulatory issues of current interest.

**BIS 600a or b, Independent Study or Directed Readings** Staff

Independent study or directed readings on a specific research topic agreed upon by faculty and student. By arrangement with faculty. For M.S. and Ph.D. students only.

**BIS 610a or b, Applied Area Readings for Qualifying Exams** Staff

Required of BIS Ph.D. students, in preparation for qualifying exams. Readings arranged with specific faculty in related research area. By arrangement with faculty.

**BIS 620a, Data Science Software Systems** Michael Kane

This course focuses on the principles of software engineering needed to be an effective informatician and data scientist. It provides the fundamentals needed to create extensible systems for processing, visualizing, and analyzing data along with providing principles for reproducibility and communication in R and Python. Prerequisite: BIS 679 or its equivalent, or permission of the instructor.

**BIS 621a, Regression Models for Public Health** Elizabeth Claus

This course focuses on the applications of regression models and is intended for students who have completed an introductory statistics class but who wish to acquire the additional statistical skills needed for the independent conduct and analysis of study designs frequently seen in public health. Topics include model selection, implementation and interpretation for linear regression with continuous outcomes, logistic regression with binary/multinomial/ordinal outcomes, and proportional hazards regression with survival time outcomes. The class explores advanced topics within these domains including the analysis of (1) blocked and nested study designs, (2) linear contrasts and multiple comparisons, (3) longitudinal data or repeated measures, (4) missing data, and (5) pragmatic clinical trials using propensity scores to reduce selection bias, etc. SAS software is used for analysis of data. Prerequisite: EPH 505 or equivalent. Not open to auditors.

**BIS 623a, Advanced Regression Models** Yize Zhao

This course provides a focused examination of the theory and application behind linear regression. Topics include linear regression, estimation, hypothesis testing, regression diagnostics, analysis of variance, adjusting for covariates, transformations, missing data, and generalized linear models. R and SAS software is used for analysis of data. Prerequisites: EPH 505 and BIS 505 or equivalents, algebra, and calculus.

**BIS 628b, Longitudinal and Multilevel Data Analysis** Veronika Shabanova

This course covers methods for analyzing longitudinal data in which repeated measures have been obtained for subjects over time and for analyzing multilevel data, which can be either hierarchically or nonhierarchically structured, e.g., nested, crossed, and/or clustered. The course teaches the common analytic techniques that can be used to analyze both longitudinal data and multilevel data with both continuous and discrete responses. One defining feature of the data is the correlation among responses over time within the same subject in longitudinal data and/or among different observations within a same cluster in multilevel data, which has to be

accommodated in order to make valid inference about the responses. Emphasis is on mixed-effects models and generalized estimating equations (GEE). Rationales on whether population-average research or subject-/cluster-specific inference research may be more appropriate for various study designs and data types are discussed and illustrated. More advanced topics including mixture models, missing data methods, and causal inference are discussed if time allows. Analysis in presence of missing data is incorporated throughout the lectures and the labs. Emphasis is placed on applying the methods, understanding underlying assumptions, and interpreting results for analyzing real data using standard statistical software. Additional material on computational aspects and theoretic aspects of mixed models. R and SAS software is used for analysis of data. Prerequisite: BIS 623 or equivalent.

**BIS 629a, Advanced Methods for Implementation and Prevention Science** Donna Spiegelman

The course presents methods for the design and analysis of studies arising in the implementation and prevention science space. These studies implement a range of cluster-randomized designs, quasi-experimental designs, and observational designs. This course consists of two parts. The first provides an exposition of the theory and analytic techniques used in the design and analysis of experimental studies arising in implementation and prevention science. The second covers the design and analysis of quasi-experimental and observational studies. SAS/R is used for problem sets. Through the provision of student PASS licenses, competency in the use of this leading software for study design is acquired. Prerequisites: S&DS 541 and S&DS 542 or equivalents, or permission of the instructor.

**BIS 630b, Applied Survival Analysis** Yuan Huang

This course demonstrates statistical methods for analyzing and interpreting time-to-failure data. The techniques described include the construction and analysis of failure rates, survival curves, hypothesis tests for comparing survival curves, parametric models, and semiparametric models for the analysis of time-to-failure data including the Cox proportional hazards model. Skills for using statistical software to perform the analyses are developed. In addition, study design is covered, including sample size and power calculations. Prerequisites: BIS 505 or equivalent, BIS 623, and single variable calculus.

**BIS 631b, Advanced Topics in Causal Inference Methods** Laura Forastiere

The evaluation of a public policy program designed to improve the health and well-being of a population requires the use of statistical methods for the estimation of its effects and the knowledge of causal inference tools to attribute the estimated effects to our intervention of interest. When studies are not well designed, several complications may arise. This course covers advanced topics of causal inference in complex settings, known as “irregular designs,” where the common assumptions required for the estimation of causal effects do not hold. Irregular designs include randomized experiments affected by non-compliance, censoring or missing outcomes, and observational studies with unmeasured confounders. We also learn how to deal with other irregular designs, including panel studies with time-invariant unmeasured confounders and regression discontinuity designs where the treatment is assigned based on a cut-off rule on test scores or poverty indexes and hence is affected by the lack of overlap. The second part of the course focuses on ways to go beyond the treatment effect and investigate all the mechanisms that come into play when the

intervention is implemented: causal pathways, spillover effects, and heterogeneity. A better understanding of these mechanisms can help us improve the design of our intervention. We first learn statistical methods to disentangle causal pathways through which the intervention has an effect. We then relax the common assumption of independence between units and allow the treatment of one unit to affect the outcome of other units. We present cutting-edge statistical methods to estimate spillover or peer-influence effects in clusters of units or in social networks. The last part of the course deals with identification of heterogeneous treatment effects using standard and machine-learning approaches. Identifying subgroups of individuals for whom the effect is more beneficial can help us design optimal and cost-effective treatment strategies where the treatment is assigned to specific individuals. The course is complemented with interesting examples from the social and health sciences. Prerequisites: S&DS 542 (or S&DS 612) and BIS 623 (or S&DS 612), or waivers for these courses; BIS 679 (or BIS 557), or a waiver for this course; and BIS 537, or exposure in other courses to fundamental concepts of causal inference. Some understanding of Bayesian statistics (taught in BIS 567) is recommended but not required.

**BIS 633a, Population and Public Health Informatics** Brian Coleman

This is not a programming course or a mathematics course. The course provides an in-depth survey of the data standards, data analysis tools, databases, and information management systems and applications used in clinical population research, disease surveillance, emergency response information systems, and the like. It examines informatics techniques used on population-level data to improve health and the application of information and computer science and technology to public health practice, research, policy, and decision support. This scientific area focuses on the capture, management, and use of electronic public health data. While these backgrounds are prominent in the field, the purpose of this course is to provide the history and context of the field.

**BIS 634a, Computational Methods for Informatics** Robert McDougal

This course introduces the key computational methods and concepts necessary for taking an informatics project from start to finish: using APIs to query online resources, reading and writing common biomedical data formats, choosing appropriate data structures for storing and manipulating data, implementing computationally efficient and parallelizable algorithms for analyzing data, and developing appropriate visualizations for communicating health information. The FAIR data-sharing guidelines are discussed. Current issues in big health data are discussed, including successful applications as well as privacy and bias concerns. This course has a significant programming component, and familiarity with programming is assumed. Prerequisite: CPSC 223 or equivalent, or permission of the instructor.

**BIS 638a, Clinical Database Management Systems and Ontologies** Kei-Hoi Cheung and George Hauser

This course introduces database and ontology in the clinical/public health domain. It reviews how data and information are generated in clinical/public health settings. It introduces different approaches to representing, modeling, managing, querying, and integrating clinical/public health data. In terms of database technologies, the course describes two main approaches – SQL database and non-SQL (NoSQL) database – and shows how these technologies can be used to build electronic health records (EHR), data repositories, and data warehouses. In terms of ontologies, it discusses

how ontologies are used in connecting and integrating data with machine-readable knowledge. The course reviews the major theories, methods, and tools for the design and development of databases and ontologies. It also includes clinical/public health use cases demonstrating how databases and ontologies are used to support clinical/public health research. Prerequisite: CPSC 223 or permission of the instructors. The general expectation to obtain instructor permission is that students have basic command of the Python programming language sufficient to pass CPSC 223 or the equivalent.

**BIS 640b / SBS 640b, User-Centered Design of Digital Health Tools** Terika McCall  
This course combines needs assessment methods, user-centered design principles, and an agile approach to designing digital health tools for consumers. The class environment is designed to model that of a health tech start-up. Students are expected to apply what they learn from the lectures and readings to identify a pain point (i.e., a problem or need faced by a prospective user) and solicit input from intended users to design a prototype of the digital health tool. Solutions are presented in class to receive feedback on the design and to iteratively refine a prototype in order to create a minimum viable product. Prerequisite: BIS 560/CB&B 740, SBS 574, or permission of the instructor.

**BIS 643b, Theory of Survival Analysis** Shuangge Ma  
This course presents the statistical theory underlying survival analysis. It covers different models of censoring and the three major approaches to analyzing this type of data: parametric, nonparametric, and semiparametric methods. The application of this theory through some exemplary data sets is also presented. Offered every other year. Prerequisites: S&DS 541 and S&DS 542. Not open to auditors.

[ **BIS 645, Statistical Methods in Human Genetics** ]  
Probability modeling and statistical methodology for the analysis of human genetics data are presented. Topics include population genetics, single locus and polygenic inheritance, linkage analysis, genome-wide association studies, quantitative trait locus analysis, rare variant analysis, and genetic risk predictions. Offered every other year. Prerequisites: EPH 505 and BIS 505, or equivalents; and permission of the instructor.

1 Course cr

**BIS 646b, Nonparametric Statistical Methods and Their Applications** Heping Zhang  
Nonparametric statistical procedures including recursive partitioning techniques, splines, bootstrap, and other sample reuse methods are introduced. Some of the supporting theory for these methods is proven rigorously, but some is described heuristically. Advantages and disadvantages of these methods are illustrated by medical and epidemiological studies. Students may be required to compare these methods with parametric methods when analyzing data sets. Familiarity with basic statistical theory and computer languages is assumed. Prerequisites: S&DS 541 and S&DS 542. Not open to auditors.

**BIS 649a and BIS 650b, Master's Thesis Research** Shuangge Ma  
The master's thesis is not required of M.S. or M.P.H. students. Students work with faculty advisers in designing their project and writing the thesis. Detailed guidelines for the thesis are outlined in Appendix II of the YSPH Bulletin.

**BIS 662b, Computational Statistics** Forrest Crawford  
This is a course in the theory and practice of statistical computing. The goal is to develop analytical and computational skills that will enable students to solve



computational challenges in their own research. The course covers basic mathematical and statistical techniques that statisticians use when analyzing data and models for which there is no ready-made software. Every component of the course covers theoretical concepts, implementation details, and applications to real data or common statistical models that students will encounter in practice. This course is not an introduction to programming, nor is it a survey of software packages for doing statistics; the course covers fundamentals of using the R language, but students are expected to be already familiar with basic concepts in programming.

**BIS 678a, Statistical Practice I** Lisa Calvocoressi, Peter Peduzzi, Denise Esserman, and James Dziura

This first term of a yearlong capstone course prepares students to transition from the classroom to the real-world practice of biostatistics. The course, which assumes a strong foundation in statistical analysis, study design, and methods, augments that knowledge with topics frequently encountered in practice: e.g., calculating sample size, handling missing data. Students have the opportunity to develop critical reading and problem-solving skills and are encouraged to bring a “big picture” perspective to their analytic work by considering study aims, hypotheses, and design as the framework for planning and conducting appropriate statistical analyses. Within that framework, students are challenged to integrate knowledge from multiple courses to write cogent statistical analysis plans and carry out complex analyses. Moreover, as biostatisticians must be able to clearly communicate their findings to fellow statisticians and non-statisticians, this course provides multiple opportunities for students to present their work orally and in writing. As in statistical practice, there are opportunities for problem-solving and decision-making at the individual and group level. Required of second-year Biostatistics M.P.H., M.S., and doctoral students. Prerequisite: BIS 623; open to second-year Biostatistics M.P.H., M.S., and doctoral students, or by permission of the instructors.

**BIS 679a, Advanced Statistical Programming in SAS and R** Elizabeth Claus

This class offers students the chance to build on basic SAS and R programming skills. Half of the term is spent working with SAS learning how to create arrays, format data, merge and subset data from multiple sources, transpose data, and write and work with macros. The second half of the term is spent working with R learning how to work with data, program functions, write simulation code using loops, and bootstrap. Prerequisites: EPH 505 and basic knowledge of both SAS and R. Not open to auditors.

**BIS 681b, Statistical Practice II** Lisa Calvocoressi, Peter Peduzzi, Denise Esserman, and James Dziura

This second term of a yearlong capstone course prepares students to transition from the classroom to the real-world practice of biostatistics. The course, which assumes a strong foundation in statistical analysis, study design, and methods, augments that knowledge with topics frequently encountered in practice: e.g., calculating sample size, handling missing data. Students have the opportunity to develop critical reading and problem-solving skills and are encouraged to bring a “big picture” perspective to their analytic work by considering study aims, hypotheses, and design as the framework for planning and conducting appropriate statistical analyses. Within that framework, students are challenged to integrate knowledge from multiple courses to write cogent statistical analysis plans and carry out complex analyses. Moreover, as biostatisticians must be able to clearly communicate their findings to fellow statisticians and non-statisticians, this course provides multiple opportunities for students to present their work orally and

in writing. As in statistical practice, there are opportunities for problem-solving and decision-making at the individual and group level. Required of second-year Biostatistics M.P.H., M.S., and doctoral students. Prerequisite: BIS 678; open to second-year Biostatistics M.P.H., M.S., and doctoral students, or by permission of the instructors.

**BIS 685a and BIS 686b, Capstone in Health Informatics** Pamela Hoffman, Kei-Hoi Cheung, David Chartash, and Hamada Altalib

Building on BIS 560/CB&B 740 and BIS 550/CB&B 750, this course provides the opportunity for master's-level integration of basic informatics theory and practice through the use of modules focusing on the workflow of major health informatics projects. Students have two major projects throughout the course, including a team project where additional reflection on coordination of responsibilities and teamwork is essential. Each student is also able to work on a term-long individual module or choose to individually continue to advance the previous team project. The final projects are meant to show how the student integrates informatics theory, skills, and stakeholder's needs into a final product or project that may be developed into a deliverable for general public use. Prerequisites: BIS 560/CB&B 740 and BIS 550/CB&B 750, or equivalents.

**BIS 687b, Data Science Capstone** Michael Kane

This course prepares students to transition from the classroom to the real-world practice at the intersection of biostatistics and data science. Students develop a holistic solution to an analytical problem by proposing study aims, hypotheses, and system design and then develop a robust, reproducible solution addressing said hypotheses. Moreover, as biostatisticians must be able to clearly communicate their findings to fellow statisticians and the domain experts with whom they collaborate, this course provides multiple opportunities for students to present their work orally and in writing. As in statistical practice, there are opportunities for problem-solving and decision-making at the individual and group level. Prerequisite: BIS 678.

**BIS 691b, Theory of Generalized Linear Models** Zuoheng Wang

This course considers a class of statistical models that generalize the linear model through the link functions of response mean. Major varieties of GLMs including models for Gaussian, Gamma, binomial, un/ordered polynomial, and Poisson responses are discussed. Goodness of fit of the models and overdispersion are considered. Extensions to correlated responses are examined through the approaches of quasi-likelihood and generalized estimating equation. The course covers both theoretical and applied aspects of data analytic issues arising from practice. Prerequisites: S&DS 542, BIS 623, and some knowledge of matrix calculation.

**BIS 692b, Statistical Methods in Computational Biology** Hongyu Zhao

Introduction to problems, algorithms, and data analysis approaches in computational biology and bioinformatics. We discuss statistical issues arising in analyzing population genetics data, gene expression microarray data, next-generation sequencing data, microbiome data, and network data. Statistical methods include maximum likelihood, EM, Bayesian inference, Markov chain Monte Carlo, and methods of classification and clustering; models include hidden Markov models, Bayesian networks, and graphical models. Offered every other year. Prerequisite: S&DS 538, S&DS 542, or S&DS 661. Prior knowledge of biology is not required, but some interest in the subject and a willingness to carry out calculations using R is assumed.

**BIS 695a or b, Summer Internship in Biostatistics** Shuangge Ma

The summer internship in biostatistics for M.S. students provides a hands-on, real-world experience in support of the student's educational and career goals. It is strongly encouraged that students seek out an internship with a public health or biomedical focus. The internship requires a full-time (thirty to forty hours per week) for ten to twelve weeks in the summer following the first year of the program. Students need to develop a work plan in conjunction with their internship supervisor and the student's faculty adviser must obtain approval of the plan. The student and internship supervisor must also complete a post-internship evaluation. First-year M.S. Health Informatics students choosing to participate in a summer internship should also enroll in this course. Prerequisite: completion of one year of the M.S. program or permission of the instructor. o Course cr

**BIS 699b, Summer Internship in Biostatistical Research** Staff

The purpose of this course is to provide students with the opportunity of gaining practical experience in the analysis and development of biostatistical methods as part of a health sciences research team including medicine, public health, pharmaceutical industry, or health care delivery. This experience provides a basis for developing a dissertation proposal that has practical significance for addressing important scientific questions. Students work with a biostatistics or health informatics faculty mentor to select a suitable placement for the summer intern, and a one-page description of the plans is submitted to the instructor at least three weeks prior to starting the program, for approval within two weeks. The internship must be full-time: thirty-five to forty hours per week for ten to twelve weeks during the summer. Upon completion of the internship, a written report of the work must be submitted to the instructor no later than October 1. Prerequisite: completion of one year of the Ph.D. program or permission of the instructor. o Course cr