

# DEPARTMENT OF BIOSTATISTICS AND DATA SCIENCE

---

## Programs

Interim Chair: Sudesh Srivastav, PhD

### Mission

The Department of Biostatistics and Data Science advances biostatistics, bioinformatics and data sciences by conducting original methodological research, collaborating on interdisciplinary research teams, training students in the application of biostatistics and bioinformatics methods and public health data analytics, and providing high quality services to the academic, research and professional communities.

### About Biostatistics

The Department of Biostatistics and Data Science has expertise in biostatistics, bioinformatics, genomics, biomedical informatics, big data and data analytics, including data capture and data management.

The BIOS faculty take great pride in providing a strong nurturing learning environment and are very accessible to students. Faculty are highly engaged in collaborative and independent research and encourage student participation in research projects both within and outside the department. Faculty serve on interdisciplinary research teams and provide expertise in statistical methodology, sample size estimations, data analysis, techniques for handling missing data, design of experiments, robust estimation, survival analysis, analysis of microarray data, genomics and proteomics.

Faculty research areas include biostatistics methods and applications, bioinformatics related to cancer, osteoporosis, respiratory and cardiovascular disease, health informatics and data analytics, big data, data capture, management analysis for large clinical trial studies.

## Graduate Degrees

- Biostatistics, MPH (<https://catalog.tulane.edu/public-health-tropical-medicine/biostatistics-data-science/biostatistics-mp/>)
- Biostatistics, MS (<https://catalog.tulane.edu/public-health-tropical-medicine/biostatistics-data-science/biostatistics-ms/>)
- Biostatistics, MSPH (<https://catalog.tulane.edu/public-health-tropical-medicine/biostatistics-data-science/biostatistics-msp/>)
- Biostatistics, PhD (<https://catalog.tulane.edu/public-health-tropical-medicine/biostatistics-data-science/biostatistics-phd/>)

## Graduate Certificates

- Biostatistics Certificate (Graduate) (<https://catalog.tulane.edu/public-health-tropical-medicine/biostatistics-data-science/biostatistics-certificate/>)

## Courses

### Biostatistics (BIOS)

#### BIOS 6040 Intermediate Biostatistics (3)

This is an intermediate course in applied biostatistics. The course covers Analysis of Variance and Multiple Regression and Correlation Analysis, and Logistic Regression. The focus will be on numerical computation and interpretation of results of statistical application using statistical packages. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030, SPHL 6850 or 6050).

#### BIOS 6220 Database Management (3)

An introduction to the principles and application of data management, techniques in data collection, data cleaning, data reporting, database design, and implementing databases for managing large data systems. After taking the course, students will be able to create databases with applications to public health intervention and surveillance, use SQL to administrate, manage, and retrieve data for statistical analysis. Prerequisite(s): Basic knowledge of MS Office.

#### BIOS 6290 Data Management and Statistical Computing (3)

This course presents basic knowledge and techniques in data management and practice. Topics include data import and export, processing and cleaning data, variable and data manipulation, descriptive summary report development, and graphic report creation. The course emphasizes hands-on experience, particularly, allowing students to develop a working knowledge and essential programming skills of commonly used statistical packages, such as SAS, R and STATA, for managing and characterizing public health-related data.

**BIOS 6300 Introduction To ArcGIS (1)**

This course covers the elementary concepts and applications for mapping using the ArcGIS software. The course focuses on a wide variety of public health applications and is applicable to virtually all academic and professional settings where mapping is used. Each lecture begins with a PowerPoint presentation to introduce fundamental mapping concepts and is followed with in-class exercises to reinforce hands-on application. Two in-class, paper-based exams are given to monitor and assess students' understanding of the course concepts.

**Prerequisite(s):** (BIOS 6030\* or SPHL 6050\*).

\* May be taken concurrently.

**BIOS 6310 Introduction to Methods in Data Science (3)**

This course introduces practical methods for gaining insights from large datasets that are transforming biomedical research. Data science utilizes tools from statistics and computer science to curate and process large datasets, explore data quantitatively, build predictive models, and communicate analysis results through visualization. Students will learn the Python programming language, databases and the SQL query language, and other technologies used by today's data science practitioners through online course materials and face-to-face lectures. Using real-world examples, students will gain experience in hands-on practice in applying these software tools to explore datasets in genetics and electronic health records. Interactive online and face-to-face discussions will also discuss current research articles that apply data science to make discoveries in biology and medicine.

**Course Limit:** 2

**BIOS 6800 Public Health GIS II (3)**

The course is an introduction to desktop mapping and spatial analysis. The first part of the course covers geographic information systems (GIS) concepts and mapping using the ArcGIS software. The second part of the course covers introductory spatial analytical techniques, including spatial autocorrelation quantification, cluster analysis, and spatial modeling. The student will develop a public health GIS project that requires the synthesis of mapping and spatial analysis.

**BIOS 7040 Statistical Inference I (3)**

The course is the first of a sequence in the theory of statistical interference and probability. The first part of the course covers probability theory; discrete, continuous, and exponential distribution functions; moment generating functions; and differentiation. The latter part of the course covers joint and marginal distributions and concepts of random samples. Students taking this course need to have completed at least one year of college calculus. Students will develop a project that synthesizes the course learning objectives through an applied course project. The course focuses on the theoretical underpinnings of biostatistics and improving understanding of statistical application and problem solving approaches.

**BIOS 7050 Statistical Inference II (3)**

The course is the second part of a sequence for introduction to statistical inference and probability. The first part of the course covers data reduction, point estimation, hypothesis testing, and interval estimation. The latter part of the course covers asymptotic evaluations, analysis of variance, and regression modes. The student will develop a project that synthesizes the course learning objectives through an applied course project. The course focuses on the theoretical underpinnings of biostatistics and improving understanding of statistical application and problem solving approaches.

**Prerequisite(s):** BIOS 7040.

**BIOS 7060 Regression Analysis (3)**

This is an advanced course on selected statistical techniques for analyzing data on multiple variables, both continuous and categorical. This course ultimately provides the student with insight into the application of regression techniques to the medical and health sciences. It focuses on statistical methodology with emphasis on selection of appropriate applications and interpretation of results. Elementary knowledge of the use of statistical computing package is needed.

**Prerequisite(s):** (BIOS 6030, SPHL 6850 or 6050) and BIOS 6040.

**BIOS 7080 Design of Experiments (3)**

This course deals with fundamental topics in design of experiments including principle theory of experimental designs (randomization, replication, and balance). It focuses the main elements of statistical thinking in the context of experimental design such as completely randomized design, randomized complete block design, experiments with two factors, factorial design, Latin Square, nested designs, repeated measurement design, and split-plot designs. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030, SPHL 6850 or 6050) and BIOS 6040.

**BIOS 7150 Categorical Data Analysis (3)**

Fundamental concepts and methods for analysis of categorical outcomes. Topics include analysis of 2-way tables, unconditional and conditional logistic regression, power and sample size computation, and modeling of dependent categorical outcomes via mixed models and GEE methods. Course covers the mathematical basis of the statistical procedures but the emphasis is on application of the methods using statistical software and interpretation of results. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030, SPHL 6850 or 6050) and BIOS 6040.

**BIOS 7220 Nonparametric Statistics (3)**

Nonparametric inferential statistical methods are introduced. Topics include single, paired, independent, and multiple sample hypothesis testing and confidence interval methods; non parametric regression and correlation methods; categorical data and measures of concordance. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030 or SPHL 6050) and BIOS 6040.

**BIOS 7250 Principles of Sampling (3)**

This course introduces core principles of survey sampling, with emphasis on sampling plans, methods of estimating unknown parameters of population and subdomain, and techniques for calculating precisions of the estimators. Topics include: basic concepts in survey sampling, simple random sampling; stratified random sampling; systematic sampling; one-, two-, and multi-stage cluster sampling; probability proportionate to size sampling. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030, SPHL 6850 or 6050).

**BIOS 7300 Survival Data Analysis (3)**

Topics include analysis of survivorship data including estimation and comparison of survival curves, regression methods in the analysis of prognostic and etiologic factors, concepts of competing risks, and the analysis of clinical trial data. Software used for problem solving. Emphasis placed on the application of methods to the analysis of public health data with examples of clinical trials, cancer survivorship, and other data sets for which there is partial follow-up of subjects. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030, SPHL 6850 or 6050) and BIOS 6040.

**BIOS 7380 Bayesian Inference (3)**

This course examines theoretical foundations and applications of Bayesian paradigm, including Bayes' theorem, prior distribution, likelihood function, deriving posterior distributions, and point and interval estimations. A variety of topics are covered, which encompass Bayesian inference for single- and multi-parameter models, linear regression, hierarchical models, and commonly used Gibbs sampler and Metropolis-Hastings algorithm. Assessment of convergence, the evaluation of models, and the presentation of the results are also illustrated. Real world examples drawn from medical research are used to show practicality of Bayesian approach, particularly how to update beliefs and make inferences from observed data. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030 or SPHL 6050) and BIOS 6040.

**BIOS 7400 Clinical Trials (3)**

Covers design, implementation, analysis and reporting of clinical trials. Topics encompass trial design, hypothesis formulation and testing, methods of randomization, ethics, sequential trials, sample size determination, blinding, subject recruitment, data collection and management, quality control, monitoring outcomes and adverse events, interim analysis, statistical methods in analyzing trial data, and addressing scientific issues in reporting and interpreting trial results. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030\*, SPHL 6850 or 6050\*) and BIOS 6040.

\* May be taken concurrently.

**BIOS 7650 Statistical Learning in Data Science (3)**

This course provides detailed overviews over the evaluation and application of statistical learning theories and techniques for inference and prediction in data science, particularly for biological and public health data. Topics include linear and nonlinear models, resampling techniques, tree-based methods, unsupervised learning such as clustering, support vector machine, graphical models, etc. Working on real and/or simulated data through assignments, students will apply the knowledge learned and practice their skills in solving various biological and public health problems, such as sequence alignment, gene prediction, subtype identification and classification, and disease risk and prognosis prediction. Discussion on model assessment and selection are also included. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030 or SPHL 6050) and BIOS 6040.

**BIOS 7990 Masters Independent Studies (1-3)**

Masters students and advisor select a topic for independent study and develop learning objectives and the expected written final product.

**BIOS 8200 Causal Inference For Biomedical Informatics (3)**

This course covers basic concepts and selected state-of-the-art statistical methods and theory of causal inference for biomedical informatics. It will empower students to draw causal conclusions and make predictions by mining data from observational and experimental studies. Topics include: targeted machine learning, structural equation modeling, Mendelian randomization, and heteroskedastic genomic prediction. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030 or SPHL 6050) and BIOS 6040, 7060 and MATH 6080.

**BIOS 8350 Clustered and Longitudinal Data Analysis (3)**

This is an advanced course in analysis of clustered and longitudinal data, with or without missing values. Students will compute power and sample size for clustered and longitudinal data using generalized linear mixed effect models and estimating equations. Class discussion, lecture, and assignments emphasize application of methods to the analysis of public health data with examples of clinical trials and epidemiological observational studies. Use of standard statistical software and methods required. Elementary knowledge of the use of statistical computing packages is needed.

**Prerequisite(s):** (BIOS 6030, SPHL 6850 or 6050) and (BIOS 6040 or 7060).

**BIOS 8500 Monte Carlo and Bootstrapping Methods (3)**

This hands-on course introduces the methods used for Monte Carlo simulations and nonparametric bootstrapping. Students learn how to design, program, and interpret a simulation study, uses of bootstrapping for estimation and inference, jackknifing, and other resampling methods. Monte Carlo Markov Chain methods and Bayesian inference in Monte Carlo methods will be introduced. This is an advanced, computer-intensive course, so knowledge of programming language (SAS or R preferred) as well as ability to work independently are required.

**Prerequisite(s):** (BIOS 7060, 7080, 7150, 7220 or 7300).

**BIOS 8820 Multivariate Methods (3)**

This is a doctorate level course that covers techniques used to conduct analysis with more than one outcome variable. The focus will be on association methods and predictive models between multiple independent and multiple dependent variables. Additionally the students will learn techniques for variable reduction, path models, and factor analysis. Students will conduct numerical computation and interpretation of results of statistical application using statistical packages. Doctoral status required. Students should have completed at least two 7000 level biostatistics courses and have working knowledge of programmable statistical software, (SAS, R, STATA).

**BIOS 8990 Doctoral Independent Study (1-3)**

Doctoral students and advisors select a topic for independent study and develop learning objectives and the expected final written product.

**BIOS 9980 Master's Thesis Research (0)**

MS Students engaging in thesis research. Course may be repeated up to unlimited credit hours.

**Course Limit:** 99