Department of Biostatistics and Data Science

The mission of the Department of Biostatistics & Data Science is to provide an infrastructure of biostatistical and informatics expertise to support and enhance the research, service and educational needs of the University of Kansas Medical Center and its affiliates. The global objectives of the department are as follows:

- To provide a leadership role in biostatistical and informatics research initiatives across the medical center.
- To provide the biostatistics and informatics cores for major initiatives.
- To ensure that researchers have ready access to biostatistical and informatics resources and support.
- To provide the infrastructure and expertise for centralized and project specific database development, management and analysis.
- To consolidate resources pertaining to biostatistics and informatics.

The innovative M.S. & Ph.D. programs in Biostatistics, M.S. in Applied Statistics, Analytics & Data Science, and M.S. in Health Data Science help meet the ever-increasing demand for statisticians, biostatisticians, and health data scientists needed to take leadership roles in academia, government, health care institutions and industry. Faculty members are active researchers collaborating and consulting in research projects and initiatives at the Medical Center, in addition to pursuing their own research agendas and participating in curricular instruction. Expertise in the department includes linear, nonlinear, and longitudinal modeling; clinical trial and experimental design; survival analysis; categorical data analysis; robust statistics; psychometric methods; statistical 'omics; bioinformatics; Bayesian methodology; data science; and machine learning.

Biostatistics and Data Science Courses

BIOS 406. Special Topics in Biostatistics & Data Science. 1-6 Credits.
Research and exploration of special topics in biostatistics and data science. May be repeated for credit if the content differs.

BIOS 704. Principles of Statistics in Public Health. 3 Credits.
Introductory course concerning the concepts of statistical reasoning and the role of statistical principles as the scientific basis for public health research and practice. Prerequisite: Permission of instructor.

BIOS 714. Fundamentals of Biostatistics I. 3 Credits.
First-semester course of a two-semester introductory statistics course that provides an understanding of the proper application of statistical methods to scientific research with emphasis on the application of statistical methodology to public health practice and research. This course focuses on basic principles of statistical inference with emphasis on one or two sample methods for continuous and categorical data. This course fulfills the core biostatistics requirement. Prerequisite: Calculus or Permission of Instructor.

BIOS 715. Introduction to Data Management using RedCap and SAS. 3 Credits.
This course will cover the utilization of Redcap and SAS for data management. Data collection and management using Redcap will be covered. Data cleaning and preparation for analysis will be covered using SAS. In addition, some of the basic descriptive analysis procedures will be covered in SAS. Prerequisite: Corequisite: BIOS 704 or BIOS 714 or equivalent with permission of instructor.

BIOS 717. Fundamentals of Biostatistics II. 3 Credits.
Second level statistics course that provides an understanding of more advanced statistical methods to scientific research with an emphasis on the application of statistical methodology to public health practice, public health research, and clinical research. Special focus will be upon the utilization of regression methodology and computer applications of such methodology. Prerequisite: BIOS 714 or equivalent with permission of instructor.

BIOS 720. Analysis of Variance. 3 Credits.
Methods for designed experiments including one-way analysis of variance (ANOVA), two-way ANOVA, repeated measures ANOVA, and analysis of covariance are emphasized. Post- ANOVA tests, power and testing assumptions required in NOVA are discussed and applied. Outlier detection using robust estimators also are incorporated. Boxplots, histograms and scatterplots are used to display data. Prerequisite: PRE 710/711 or BIOS 714/717 or equivalent. Preferred: BIOS 715. Knowledge of statistical software, basic statistical plotting methods, p-values, two-sample t-test and simple linear regression is assumed.

BIOS 725. Applied Nonparametric Statistics. 3 Credits.
This course will study nonparametric methods in many situations as highlighted by the following topics: Students will learn how nonparametric methods provide exact p-values for tests, exact coverage probabilities for confidence intervals, exact experimentwise error rates for multiple comparison procedures, and exact coverage probabilities for confidence bands. This course will be using EXCEL and SAS to conduct various procedures. Prerequisite: BIOS 714 or equivalent with permission of instructor.

BIOS 730. Applied Linear Regression. 3 Credits.
Simple linear regression, multiple regression, logistic regression, nonlinear regression, neural networks, autocorrelation, interactions, and residual diagnostics. Applications of the methods will focus on health related data. Prerequisite: 1) BIOS 714 or the equivalent and 2) BIOS 717 or BIOS 720 or equivalent with permission of instructor.

BIOS 735. Categorical Data and Survival Analysis. 3 Credits.
An intermediate level statistics course that provides an understanding of the more advanced statistical methods to scientific research with emphasis on the application of statistical methodology to clinical research, public health practice, public health research and epidemiology. Prerequisite: BIOS 714, BIOS 715, and BIOS 717 or permission of the instructor.

BIOS 740. Applied Multivariate Methods. 3 Credits.
This course is an advanced statistical course for students who have had fundamental biostatistics and linear regression. Topics to be covered include Hotelling's T-squared test, MANOVA, principal components, factor analysis, discriminant analysis, canonical analysis, and cluster analysis. More advanced topics such as Multidimensional Scaling or Structural Equation Modeling might be introduced if time allows. Computers will be extensively used through the whole course, and students are suggested to be familiar with some statistical software before taking this course. Although students are allowed to use the software they are comfortable with, SAS will be the primary statistical package used to demonstrate examples in this course. Prerequisite: Corequisite: BIOS 730 or equivalent with permission of instructor.

BIOS 799. Introduction to Statistical Genomics. 3 Credits.
This survey course will provide a high-level introduction of various statistical and bioinformatics methods involved in the study of biological systems. In particular, this course will provide an overview of the

analytical aspects involved in: the study DNA, RNA, and DNA methylation data measured from both microarray and next-generation sequencing (NGS) technologies. During the last week of the summer semester, students will be required to participate in a group seminar session in which they will present the results from their assigned genomics analysis projects. Prerequisite: Corequisite: BIOS 714 and BIOS 717 or equivalent with permission of instructor; Experience with a higher level programming language is preferred.

BIOS 805. Professionalism, Ethics and Leadership in the Statistical Sciences. 3 Credits.

This web-based course addresses issues in professionalism, leadership and ethics that are specific to students training to become statisticians, biostatisticians, and data scientists. Topics include use of sound statistical methodology, common threats to valid inference, effective communication and collaboration with content-area experts, maintaining transparency and independence, reproducible research, the publishing process (including authorship guidelines, plagiarism, peer review, intellectual property, etc.), conflict of interest, data security, and properties of effective leaders, among others. Prerequisite: Department consent.

BIOS 806. Special Topics in Biostatistics. 1-3 Credits.

This course allows exploration of special topics that are not routinely a part of the curriculum. Prerequisite: Permission of the instructor.

BIOS 810. Clinical Trials. 3 Credits.

The design, implementations, analysis, and assessment of controlled clinical trials. Basic biostatistical concepts and models will be emphasized. Issues of current concern to trialists will be explored. Prerequisite: By permission of instructor.

BIOS 811. Scientific Rigor and Reproducibility. 3 Credits.

This course introduces the principles and practices required to conduct rigorous and reproducible research across the translational spectrum. The National Institutes of Health (NIH) promotes rigor and reproducibility in their guidance to grant applicants as part of the scorable parameters that grant reviewers must address. In addition, NIH requires formal instruction in scientific rigor and transparency for individuals supported by institutional training grants, career development awards, and fellowships. In this course, students learn best practices, including sound study planning and design, consideration of all relevant biomedical variables, sound data management practices, statistical considerations and techniques, and transparency in reporting research results. Prerequisite: BIOS 714 or equivalent with permission of instructor.

BIOS 815. Introduction to Bioinformatics. 3 Credits.

Bioinformatics, an interdisciplinary field at the cross-section of biology, computer science, and statistics, has played a key role in enhancing our understanding of many areas of biology. The broad purpose of this course is to introduce students in the quantitative sciences to the field of bioinformatics and its practice. Topics include foundatational concepts in molecular biology, biological databases, sequence alignment, BLAST, molecular phylogenetics, genomics, transcriptomics, proteomics, microbiomics, with treatment of the accompanying bioinformatic tools/methodologies that have been developed to analyze such data types . Over the semester, students will gain a familiarity with the essential concepts and theories underlying the practice of bioinformatics, different types of 'omic data, the technologies used to generate different 'omic data types, and databases and tools commonly used for bioinformatics analysis. Prerequisite: There are no formal prerequisties for this course. Previous graduate-level coursework in probability and statistics and molecular biology is helpful, but not necessary.

BIOS 820. SAS Programming I. 3 Credits.

This is a graduate level course preparing a student for the SAS base programming certification exam. We will cover the topics required for a

student to pass the SAS base programming certification exam given by SAS. To this end, topics we will study will include, referencing files and setting options, creating list reports, understanding data step processing, creating and managing variables, reading and combining SAS data sets, do loops, arrays, and reading raw data from files. After the completion of the course the student should be able to create SAS programs to read data from external files, manipulate the data into variables to be used in an analysis, generate basic reports showing the results. Prerequisite: Permission of the Instructor.

BIOS 821. SAS Programming II. 3 Credits.

This is a graduate level course preparing a student for the SAS advanced programming certification exam. We will cover the topics required for a student to pass the SAS advanced programming certification exam given by SAS. To this end, topics we will study include array processing, use of data step views, using the data step to write SAS programs, efficient use of the sort procedure, introduction to the macro language in SAS, and accessing data using SAS PROC SQL. After the completion of the course the student should be able to create SAS programs to read data from external files, manipulate the data into variables to be used in an analysis, generate basic reports showing the results. Prerequisite: Corequisite: BIOS 820 or equivalent (SAS Certified BASE programmer for SAS or at least one year of experience as a data analyst/programmer).

BIOS 823. Introduction to Programming and Applied Statistics in R. 3 Credits.

This course will provide students with the opportunity to learn advanced statistical programming. The development of new statistical or computational methods often implies the development of programming codes to support its application. Much of this type of development is currently carried out in the R (or S-Plus) language. Indeed much of the recent development of statistical genetics is based on the R programming language and environment. This course provides an introduction to programming in the R language and it's applications to applied statistical problems. Prerequisite: Corequisite: Some previous exposure to computer programming. Some basic statistics at the Applied Regression or Applied Design level and permission of instructor.

BIOS 825. Nonparametric Methods. 3 Credits.

This course is an introduction to nonparametric statistical methods for data that do not satisfy the normality or other usual distributional assumptions. We will cover most of the popular nonparametric methods used for different scenarios, such as a single sample, two independent or related samples, three or more independent or related samples, goodness-of-fit tests, and measures of association. Power and sample size topics will also be covered. The course will cover the theoretical basis of the methods at an intermediate mathematical level, and will also present applications using real world data and statistical software. Prerequisite: Permission of instructor.

BIOS 830. Experimental Design. 3 Credits.

The emphasis of this course is on learning the basics of experimental design and the appropriate application and interpretation of statistical analysis of variance techniques. Prerequisite: Permission of instructor, BIOS 820 recommended.

BIOS 833. Measurement for Statisticians. 3 Credits.

This course aims to introduce the theory and applications of measurement and psychometrics to students in the statistical sciences. The goal is for students to master the concepts of measurement theory, classical/modern test theory, reliability and validity, factor analysis, structural equation modeling, item response theory, and differential item functioning. Prerequisite: Corequisite: BIOS 835, or by permission of instructor.

BIOS 835. Categorical Data Analysis. 3 Credits.

This course provides an understanding of both the mathematical theory and practical applications for the analysis of data for response measures that are ordinal or nominal categorical variables. This includes univariate analysis, contingency tables, and generalized linear models for categorical response measures. Regression techniques covered for categorical response variables, such as logistic regression and Poisson regression methods, will include those categorical and/or continuous explanatory variables, both with and without interaction effects. Prerequisite: By permission of instructor; BIOS 820 and BIOS 840 are recommended.

BIOS 840. Linear Regression. 3 Credits.

This course is an introduction to model building using regression techniques. We will cover many of the popular topics in Linear Regression including: simple linear regression, multiple regression, model selection and validation, diagnostics and remedial measures. Prerequisite: By permission of the instructor.

BIOS 845. Survival Analysis. 3 Credits.

This course provides an understanding of both the mathematical theory and practical applications for the analysis of time to event data with censoring. This includes univariate analysis, group comparisons, and regression techniques for survival analysis. Parametric and semi-parametric regression techniques covered will include those with categorical and/or continuous explanatory variables, both with and without interaction effects. Prerequisite: Corequisite: BIOS 820, 835, 840, and 871, or by permission of instructor.

BIOS 850. Multivariate Statistics. 3 Credits.

This course will introduce the theory and methods of applied multivariate analysis. As the field of multivariate analysis is very wide and well developed, the course will focus on those methods that are more frequently used in biostatistical applications. Some knowledge of basic matrix algebra is necessary and will be reviewed as the course progresses. Theoretical exercises and analysis of data sets will be assigned to the student. Emphasis will be on biostatistical applications. Prerequisite: Corequisite: BIOS 820, BIOS 830, and BIOS 840.

BIOS 855. Statistical Methods in Genomics Research. 3 Credits.

This survey course will provide a high-level introduction to various statistical and bioinformatics methods involved in the study of biological systems. In particular, this course will provide an overview of the analytical aspects involved in: the study DNA, RNA, and DNA methylation data measured from both microarray and next-generation sequencing (NGS) technologies. During the last week of the summer semester, students will be required to participate in a group seminar session in which they will present the results from their assigned genomics projects. Prerequisite: BIOS 820 OR experience programming in a higher level programming language; BIOS 840; OR by permission of the instructor.

BIOS 860. Clinical Trial Design and Analysis. 3 Credits.

This course, is intended for students interested in the statistical aspects of clinical trial research,. This course will provide a comprehensive overview of the design and analysis of clinical trials, including: first-in-human studies (dose-finding, safety, proof of concept, Phase I), Phase II, Phase III, and Phase IV studies. Prerequisite: By permission of instructor. BIOS 820, BIOS 830, BIOS 840.

BIOS 871. Mathematical Statistics. 3 Credits.

This course introduces the fundamentals of probability theory, random variables, distribution and density functions, expectations, transformations of random variables, moment generating functions, convergence concepts, sampling distributions, and order statistics. Prerequisite: By permission of instructor.

BIOS 872. Mathematical Statistics II. 3 Credits.

This course introduces the fundamentals of statistical estimation and hypothesis testing, including point and interval estimation, likelihood and sufficiency principles, properties of estimators, loss functions, Bayesian analysis, and asymptotic convergence. Prerequisite: BIOS 871 or by permission of instructor.

BIOS 880. Data Mining and Analytics. 3 Credits.

Students will be introduced to common steps used in data mining, such as accessing and assaying prepared data; pattern discovery; predictive modeling using decision trees, regression, and neural networks; and model assessment methods. Prerequisite: Corequisite: BIOS 820, 830, 835, 840, and 871, or by permission of instructor. BIOS 821 and 850 recommended.

BIOS 898. Collaborative Research Experience. 3 Credits.

This course provides students with experience in collaborative research under the supervision of an experienced researcher. The student will spend one semester working under an investigator or faculty member, making independent contributions to a research project. Prerequisite: Corequisite: BIOS 820, 830, 835, 840, 871, and 872, or by permission of instructor.

BIOS 899. MSCR Thesis. 3 Credits.

This course involves preparation of a formal thesis based on the research conducted by a student working toward the MS in Clinical Research and directed by a faculty member in the Department of Biostatistics. After the thesis has been completed, the student will be given an oral examination of the research methods and content. Prerequisite: Corequisite: Department of Biostatistics approval.

BIOS 900. Linear Models. 3 Credits.

This course introduces the theory and methods of linear models for data analysis. The course includes the theory of general linear models including regression models, experimental design models, and variance component models. Least squares estimation, the Gauss-Markov theorem, and less than full rank hypotheses will be covered. Prerequisite: Corequisite: BIOS 871 and BIOS 872 or by permission of instructor; BIOS 820 recommended.

BIOS 902. Bayesian Statistics. 3 Credits.

This course introduces Bayesian theory and methods for data analysis. The course includes an overview of the Bayesian approach to statistical inference, performance of Bayesian procedures, Bayesian computational issues, model criticism, and model selection. Case studies from a variety of fields are incorporated into the course. Implementation of models using Markov chain Monte Carlo methods is emphasized. Prerequisite: Corequisite: BIOS 871 and 872 or by permissions of instructor; BIOS 820 recommended.

BIOS 905. Theory of Statistical Inference. 3 Credits.

This course covers advanced aspects of statistical inference. It is aimed at preparing Ph.D. BIOS students for the Ph.D. comprehensive exam and will emphasize advanced biostatistical ideas as well as problem solving techniques. Prerequisite: Corequisite: BIOS 871 and BIOS 872 or equivalent and permission of instructor.

BIOS 906. Advanced Special Topics in Biostatistics. 1-9 Credits.

This course allows exploration of special topicss that are not routinely a part of the Biostatistics PhD curriculum. Prerequisite: Passing grade on the PhD Qualifying exam. Permission of the instructor.

BIOS 908. Advanced Clinical Trials. 3 Credits.

This course provides an introduction to recent innovations in clinical trial designs and analysis methods. Topics include concepts of controls, blinding, and randomization; common trial designs by phase of clinical development; sample size calculations; interim analysis; and adaptive clinical trials. Traditional frequentist and likelihood approaches to trial

design and analysis will be covered in the first half of the course; the Bayesian approach (including adaptive clinical trial designs) will be emphasized in the second half of the course. Prerequisite: BIOS 860 and BIOS 902 or by permission of the instructor.

BIOS 910. Generalized Linear Models. 3 Credits.
This course on Generalized Linear Models (GLM) is designed for both the applied and theoretical statistician. In this course we introduce the theoretical foundations and key applications of generalized linear models. Prerequisite: BIOS 835, BIOS 840, and BIOS 900 or by permission of instructor.

BIOS 911. Nonlinear Models. 3 Credits.
This course will involve both theory and applications of nonlinear models, with emphasis in biological, medical, and pharmaceutical research. Applications to dose-response studies, bioassay studies and clinical pharmacokinetics and pharmacodynamics studies will be discussed. Nonlinear mixed effects models will also be examined, as well as criteria for optimal experimental designs based on nonlinear models. This course will cover the theoretical basis of the methods at an intermediate mathematical level, and will also present applications using real world data and statistical software. Prerequisite: BIOS 900 or equivalent and permission of instructor.

BIOS 915. Longitudinal Data Analysis. 3 Credits.
A longitudinal study is a research study that involves repeated observations of the same individuals and events over extended periods of time. It is typically a type of observational study, though may have design components. In medical settings these studies and related models are used to observe the developmental path of a disease or treatment through time. Often this is in the context of follow-up and long-term study of both progress and potential side-effects. As the study involves the same individuals (subject to drop-out) through several time points, statistical methods must employ random effects or "mixed models" incorporating various correlation structures. This is typically done using generalized estimating equations and marginal model approaches. Bayesian methods may also be appropriate here. Students will, after completing this course, be able to design and analyze longitudinal studies. The computer package to be employed is SAS. Prerequisite: BIOS 820, BIOS 830, BIOS 840, BIOS 871, BIOS 872, and BIOS 900 or by permission of instructor.

BIOS 920. Latent Variable Analysis. 3 Credits.
Latent variables refer to random variables whose realization values are not observable or cannot be measured without error, and their inferences rely on statistical models connecting latent and other observed variables. This course aims to introduce a family of such statistical models and their applications in biomedical and public health research. The course is designed as an elective course for students in the Biostatistics graduate program. We will use the statistical packages of M-plus, R, and/or SAS for the course. Prerequisite: BIOS 835 and BIOS 900, or by permission of instructor. Familiarity with vectors and matrices is strongly encouraged.

BIOS 999. Doctoral Dissertation. 1-9 Credits.
Preparation of the doctoral dissertation based upon original research and in partial fulfillment of the requirements for the Ph.D. degree. Credits will be given only after the dissertation has been accepted by the student's dissertation committee. Prerequisite: Successful completion of the Department of Biostatistics Ph.D. Comprehensive Exam and consent of advisor.

Biostatistics and Data Science Courses

DATA 806. Special Topics in Data Science. 1-3 Credits.

This course allows exploration of special topics that are not routinely a part of the Applied Statistics & Analytics and Data Science curriculum. Prerequisite: Permission of instructor.

DATA 817. Introduction to Tableau. 1 Credits.
Under Tableau Desktop-I specialization, the student will discover what data visualization is, and how to use it use to better display and understand the information within a data set. Using Tableau, this course will examine the fundamental concepts of data visualization and explore the Tableau Desktop interface, identifying and applying the various tools Tableau has to offer. By the end of the course, students will be able to prepare and import data into Tableau and explain the relationship between data analytics and data visualization. This course is designed for learners who have never used Tableau before, those in need of a refresher, or those wanting to explore Tableau in more depth. No prior technical or analytical background is required. The course will guide students through the steps necessary to create visualization dashboard and story from the beginning based on data context, setting the stage for students to be ready for Desktop-I certification. Prerequisite: There are no formal prerequisites for this course. Prior experience generating plots, tables, graphs, etc. is helpful, but is not required.

DATA 819. Introduction to Python. 1 Credits.
This is a one credit hour introduction course to programming in Python. The fundamentals of Python programming, including: introduction to Python syntax, types, data structures, control of flow, functions, modules and packages, reading and writing files, and basic statistics will be covered throughout the course.

DATA 822. Introduction to SQL. 1 Credits.
This course prepares students to interact with most dialects of Structured Query Language (SQL). At the conclusion of the course, students will be prepared to interact with any major database, including PostgreSQL, MySQL, Oracle, among others. Topics covered relational databases, structure of data, Data Definition Language (DDL), Data Manipulation Language (DML), table joins, data summarization, and writing and interpreting SQL queries.

DATA 824. Data Visualization and Acquisition. 3 Credits.
Being a data scientist requires an integrated skill set that spans the domains of statistics, machine learning, and computer programming. It also demands a solid foundation in the principles of data visualization in order to create effective data presentations that convey the intended message. Put simply, data visualization describes any effort to assist an individual's understanding of the significance of data by placing it in a visual context. In this course, students will be introduced to principles of effective data visualization and tools commonly used for its implementation. Techniques and strategies for visualizing different types of data (e.g., numeric data, non-numeric data, spatial-temporal data, etc.), the use of space and color to visually encode data, interactive visualizations, acquiring and visualizing data from publicly available data repositories, data cleaning and standardizing, are examples of some of the topics this course will address. The focus in the treatment of these topics will be on breadth, rather than depth, and emphasis will be placed on integration and synthesis of concepts and their application to solving problems. Prerequisite: While there are no formal prerequisites for this course, students should have a basic familiarity with the R statistical programming language (STAT 823 highly recommended). Prior experience using statistical software (e.g., R) to generate plots, tables, graphs, etc. is helpful, but is not required.

DATA 881. Statistical Learning I. 3 Credits.
Statistical learning is a fundamental skill for data scientists. Data scientists are specialists in "drinking from the firehose" of big data, and statistical learning techniques are some of their key tools. This course focuses on

applications of statistical learning to big data challenges through data mining and predictive modeling techniques that are in great demand. Students will be introduced to the basics of statistical/machine learning: supervised learning (e.g. linear model, nonlinear models, penalized methods, ensemble methods, etc.), unsupervised learning (e.g. K means clustering, nearest neighbors, hierarchical clustering, etc.), and missing data in machine learning. Throughout the course, we will learn how to be "informed doers", who not only know how to apply methods but understand how those methods work. This understanding can be critical to getting good results from big data, so that the limitations of certain methods are properly understood. Prerequisite: STAT 820 or STAT 823, STAT 835, STAT 840, or by permission of instructor.

DATA 882. Statistical Learning II. 3 Credits.

Knowledge of how and when to apply more sophisticated statistical learning models to big data can make a data scientist an indispensable asset to a research team. In Statistical Learning 2, we will learn how to be "informed doers". We will learn how many of the covered methods work, in addition to the proper situations to apply them. This is particularly important in this course, because these methods are applicable when simpler methods are inappropriate and rarely work well without significant tinkering. Data scientists with mastery of these methods are empowered to investigate questions that are far too complex to answer with the more general "workhorse" methods covered in the first unit of this series, Statistical Learning 1. We will cover many of the most important techniques in use today, including: mixture models, hidden Markov models, spline regression, support vector machines, advanced discriminant analysis methods, neural networks (including deep learning), and methods for handling highly complex computation, such as Hadoop. The course culminates with a short project that will pull together all the skills you have learned to demonstrate how they can be used for statistical decision support, which is a common task for data scientists. Prerequisite: DATA 881, or by permission of instructor.

Biostatistics and Data Science Courses

STAT 655. Foundations of Mathematics for Data Science. 3 Credits.

Topics in single- and multiple-variable differential and integral calculus and linear algebra with applications in statistics and data science. Mathematical concepts including limits, derivatives, integrals, sequences, series, vectors, matrices, and optimization problems will be covered in the context of statistical applications. Prerequisite: College algebra or equivalent.

STAT 805. Professionalism, Ethics and Leadership in the Statistical Sciences. 3 Credits.

This web-based course addresses issues in professionalism, leadership and ethics that re specific to students training to become statisticians, biostatisticians, and data scientists. Topics include use of sound statistical methodology, common treats to valid inference, effective communication and collaboration with content-area experts, maintaining transparency and independence, reproducible research, the publishing process (including authorship guidelines, plagiarism, peer review, intellectual property, etc.), conflict of interest, data security, and properties of effective leaders, among others. Prerequisite: Permission of instructor.

STAT 806. Special Topics in Applied Statistics and Analytics. 1-3 Credits.

This course allows exploration of special topics that are not routinely a part of the Applied Statistics & Analytics curriculum. Prerequisite: Permission of instructor.

STAT 818. Introduction to R. 1 Credits.

This course will provide students with the opportunity to learn applied statistics using R statistical programming language.

STAT 820. SAS Programming I. 3 Credits.

This is a graduate level course preparing a student for the SAS base programming certification exam. We will cover the topics required for a student to pass the SAS base programming certification exam given by SAS. To this end, topics we will study will include, referencing files and setting options, creating list reports, understanding data step processing, creating and managing variables, reading and combining SAS data sets, do loops, arrays, and reading raw data from files. After the completion of the course the student should be able to create SAS programs to read data from external files, manipulate the data into variables to be used in an analysis, generate basic reports showing the results, be able to understand and explain results from univariate analyses using proc univariate. Prerequisite: Permission of Instructor.

STAT 821. SAS Programming II. 3 Credits.

This is a graduate level course preparing a student for the SAS advanced programming certification exam. We will cover the topics required for a student to pass the SAS advanced programming certification exam given by SAS. To this end, topics we will study include array processing, use of data step views, using the data step to write SAS programs, efficient use of the sort procedure, introduction to the macro language in SAS, and accessing data using SAS PROC SQL. After the completion of the course, the student should be able to create SAS programs to read data from external files, manipulate the data into variable to be used in an analysis, generate basic reports showing the results. Prerequisites: STAT 820 or equivalent (SAS Certified BASE programmer for SAS or at least one year of experience as a data analyst/programmer).

STAT 823. Introduction to Programming and Applied Statistics in R. 3 Credits.

This course will provide students with the opportunity to learn advanced statistical programming. The development of new statistical or computational methods often implies the development of programming codes to support its application. Much of this type of development is currently carried out in the R (or S-Plus) language. Indeed much of the recent development of statistical genetics is based on the R programming language and environment. This course provides an introduction to programming in the R language and it's applications to applied statistical problems. Prerequisites: Some previous exposure to computer programming. Some basic statistics at the Applied Regression or Applied Design level and permission of instructor.

STAT 825. Nonparametric Methods. 3 Credits.

This course is an introduction to nonparametric statistical methods for data that doe not satisfy the normality or other usual distributional assumptions. We will cover most of the popular nonparametric methods used for different scenarios, such as a single sample, two independent or related samples, three or more independent or related samples, goodness-of-fit tests, and measures of association. Power and sample size topics will also be covered. The course will cover the theoretical basis of the methods at an intermediate mathematical level, and will also present applications using real world data and statistical software. Prerequisite: Permission of instructor.

STAT 830. Experimental Design. 3 Credits.

The emphasis of this course is on learning the basics of experimental design and the appropriate application and interpretation of statistical analysis of variance techniques. Prerequisite: Permission of instructor. STAT 820 or STAT 823 is recommended.

STAT 833. Measurement for Statisticians. 3 Credits.

This course aims to introduce the theory and applications of measurement and psychometrics to students in the statistical sciences. The goal is

for students to master the concepts of measurement theory, classical/modern test theory, reliability and validity, factor analysis, structural equation modeling, item response theory, and differential item functioning. Prerequisites: Prerequisite: STAT 820 or STAT 823 Corequisite: STAT 835, or by permission of instructor.

STAT 835. Categorical Data Analysis. 3 Credits.
This course provides an understanding of both the mathematical theory and practical applications for the analysis of data for response measures that are ordinal or nominal categorical variables. This includes univariate analysis, contingency tables, and generalized linear models for categorical response measures. Regression techniques covered for categorical response variables, such as logistic regression and Poisson regression methods, will include those categorical and/or continuous explanatory variables, both with and without interaction effects. Prerequisites: Permission of instructor. STAT 820 or STAT 823 and STAT 840 are recommended.

STAT 840. Linear Regression. 3 Credits.
This course is an introduction to model building using regression techniques. We will cover many of the popular topics in linear regression including: simple linear regression, multiple linear regression, model selection and validation, diagnostics, and remedial measures. Prerequisite: Permission of Instructor.

STAT 845. Survival Analysis. 3 Credits.
This course provides an understanding of both the mathematical theory and practical applications for the analysis of time to event data with censoring. This includes univariate analysis, group comparisons, and regression techniques for survival analysis. Parametric and semi-parametric regression techniques covered will include those with categorical and/or continuous explanatory variables, both with and without interaction effects. Prerequisites: STAT 820 or STAT 823, 835, and 840 or by permission of instructor.

STAT 850. Multivariate Statistics. 3 Credits.
This course will introduce the theory and methods of applied multivariate analysis. Topics include multivariate model formulation, multivariate normal distribution, Hotelling's T-square, multivariate analysis of variance, repeated measures analysis of variance, growth curves, discriminant analysis, classification analysis, principal components analysis, and cluster analysis. Prerequisites: STAT 820 or STAT 823, and STAT 840, or by permission of the instructor.

STAT 855. Statistical Methods in Genomics Research. 3 Credits.
This survey course will provide a high-level introduction to various statistical and bioinformatics methods involved in the study of biological systems. In particular, this course will provide an overview of the analytical aspects involved in: the study DNA, RNA, and DNA methylation data measured from both microarray and next-generation sequencing (NGS) technologies. This course will be held in a block format with 4 hours of lectures a day for two weeks (one week in June and one week in July), with readings and homework assignments assigned throughout the summer semester. During the last week of the summer semester, students will be required to participate in a group seminar session in which they will present the results from their assigned genomics projects. Prerequisite: STAT 820 or STAT 823, and STAT 840, or by permission of the instructor.

STAT 871. Mathematical Statistics. 3 Credits.
This course introduces the fundamentals of probability theory, random variables, distribution and density functions, expectations, transformations of random variables, moment generating functions, convergence concepts, sampling distributions, and order statistics. Prerequisite: Permission of Instructor.

STAT 872. Mathematical Statistics II. 3 Credits.
This course introduces the fundamentals of statistical estimation and hypothesis testing, including point and interval estimation, likelihood and sufficiency principles, properties of estimators, loss functions, Bayesian analysis, and asymptotic convergence. Prerequisite: STAT 871 or by permission of instructor.

STAT 880. Data Mining and Analytics. 3 Credits.
Students will be introduced to common steps used in data mining, such as assessing and assaying prepared data; pattern discovery; predictive modeling using decision trees, regression, and neural networks; and model assessment methods. Prerequisites: STAT 820 or STAT 823, STAT 835, and STAT 840, or by permission of instructor. STAT 850 is recommended.

Health Data Science Courses

HDSC 805. Professionalism, Ethics and Leadership in the Statistical Sciences. 3 Credits.
This web-based course addresses issues in professionalism, leadership and ethics that are specific to students training to become statisticians, biostatisticians, and data scientists. Topics include use of sound statistical methodology, common threats to valid inference, effective communication and collaboration with content-area experts, maintaining transparency and independence, reproducible research, the publishing process (including authorship guidelines, plagiarism, peer review, intellectual property, etc.), conflict of interest, data security, and properties of effective leaders, among others. Prerequisite: Department consent.

HDSC 812. Clinical Data Management. 3 Credits.
This course presents critical concepts and practical methods to support the planning, collection, storage, and dissemination of data in clinical research. Understanding and implementing solid data management principles is critical for any scientific domain. Regardless of your current (or anticipated) role in the research enterprise, a strong working knowledge and skill set in data management principles and practice will increase your productivity and improve your science. Our goal is to use these modules to help you learn and practice this skill set. Prerequisite: Some basic knowledge of the clinical trials process; BIOS 810 is recommended but not required.

HDSC 815. Introduction to Bioinformatics. 3 Credits.
Bioinformatics, an interdisciplinary field at the cross-section of biology, computer science, and statistics, has played a key role in enhancing our understanding of many areas of biology. The broad purpose of this course is to introduce students in the quantitative sciences to the field of bioinformatics and its practice. Topics include foundatational concepts in molecular biology, biological databases, sequence alignment, BLAST, molecular phylogenetics, genomics, transcriptomics, proteomics, microbiomics, with treatment of the accompanying bioinformatic tools/methodologies that have been developed to analyze such data types . Over the semester, students will gain a familiarity with the essential concepts and theories underlying the practice of bioinformatics, different types of 'omic data, the technologies used to generate different 'omic data types, and databases and tools commonly used for bioinformatics analysis. Prerequisite: There are no formal prerequisites for this course. Previous graduate-level coursework in probability and statistics and molecular biology is helpful, but not necessary.

HDSC 818. Introduction to R. 1 Credits.
This course will provide students with the opportunity to learn applied statistics using R statistical programming language.

HDSC 819. Introduction to Python. 1 Credits.
This is a one credit hour introduction course to programming in Python. The fundamentals of Python programming, including: introduction to

Python syntax, types, data structures, control of flow, functions, modules and packages, reading and writing files, and basic statistics will be covered throughout the course.

HDSC 820. SAS Programming I. 3 Credits.

This is a graduate level course preparing a student for the SAS base programming certification exam. We will cover the topics required for a student to pass the SAS base programming certification exam given by SAS. To this end, topics we will study will include, referencing files and setting options, creating list reports, understanding data step processing, creating and managing variables, reading and combining SAS data sets, do loops, arrays, and reading raw data from files. After the completion of the course the student should be able to create SAS programs to read data from external files, manipulate the data into variables to be used in an analysis, generate basic reports showing the results. Prerequisite: Permission of the Instructor.

HDSC 822. Introduction to SQL. 1 Credits.

This course prepares students to interact with most dialects of Structured Query Language (SQL). At the conclusion of the course, students will be prepared to interact with any major database, including PostgreSQL, MySQL, Oracle, among others. Topics covered relational databases, structure of data, Data Definition Language (DDL), Data Manipulation Language (DML), table joins, data summarization, and writing and interpreting SQL queries.

HDSC 823. Introduction to Programming and Applied Statistics in R. 3 Credits.

This course will provide students with the opportunity to learn advanced statistical programming. The development of new statistical or computational methods often implies the development of programming codes to support its application. Much of this type of development is currently carried out in the R (or S-Plus) language. Indeed much of the recent development of statistical genetics is based on the R programming language and environment. This course provides an introduction to programming in the R language and it's applications to applied statistical problems. Prerequisite: Corequisite: Some previous exposure to computer programming. Some basic statistics at the Applied Regression or Applied Design level and permission of instructor.

HDSC 824. Data Visualization and Acquisition. 3 Credits.

Being a data scientist requires an integrated skill set that spans the domains of statistics, machine learning, and computer programming. It also demands a solid foundation in the principles of data visualization in order to create effective data presentations that convey the intended message. Put simply, data visualization describes any effort to assist an individual's understanding of the significance of data by placing it in a visual context. In this course, students will be introduced to principles of effective data visualization and tools commonly used for its implementation. Techniques and strategies for visualizing different types of data (e.g., numeric data, non-numeric data, spatial-temporal data, etc.), the use of space and color to visually encode data, interactive visualizations, acquiring and visualizing data from publicly available data repositories, data cleaning and standardizing, are examples of some of the topics this course will address. The focus in the treatment of these topics will be on breadth, rather than depth, and emphasis will be placed on integration and synthesis of concepts and their application to solving problems. Prerequisite: While there are no formal prerequisites for this course, students should have a basic familiarity with the R statistical programming language (HDSC 823 highly recommended). Prior experience using statistical software (e.g., R) to generate plots, tables, graphs, etc. is helpful, but is not required.

HDSC 826. Data Literacy. 3 Credits.

Through this course, students will be educated on the concepts of how to explore, understand, and communicate with data meaningfully. Data literacy is defined as the ability to read, write, and communicate data in context – with an understanding of the data sources and constructs, analytical methods, and techniques applied – and then the ability to describe the use-case application and resulting business value or outcome. Students can use tools such as Tableau, R, and Qlik in this course. We will use the Qlik Sense tool, a cloud-based tool, for this class. Prerequisite: BIOS 714 and BIOS 717, or BIOS/HDSC 823, or equivalent; or by permission of the instructor.

HDSC 830. Experimental Design. 3 Credits.

The emphasis of this course is on learning the basics of experimental design and the appropriate application and interpretation of statistical analysis of variance techniques. Prerequisite: Permission of instructor, HDSC 820 recommended.

HDSC 835. Categorical Data Analysis. 3 Credits.

This course provides an understanding of both the mathematical theory and practical applications for the analysis of data for response measures that are ordinal or nominal categorical variables. This includes univariate analysis, contingency tables, and generalized linear models for categorical response measures. Regression techniques covered for categorical response variables, such as logistic regression and Poisson regression methods, will include those categorical and/or continuous explanatory variables, both with and without interaction effects. Prerequisite: By permission of instructor; HDSC 820 or HDSC 823 and HDSC 840 are recommended.

HDSC 840. Linear Regression. 3 Credits.

This course is an introduction to model building using regression techniques. We will cover many of the popular topics in Linear Regression including: simple linear regression, multiple regression, model selection and validation, diagnostics and remedial measures. Prerequisite: By permission of the instructor.

HDSC 845. Survival Analysis. 3 Credits.

This course provides an understanding of both the mathematical theory and practical applications for the analysis of time to event data with censoring. This includes univariate analysis, group comparisons, and regression techniques for survival analysis. Parametric and semi-parametric regression techniques covered will include those with categorical and/or continuous explanatory variables, both with and without interaction effects. Prerequisite: Corequisite: HDSC 823, 835, 840 or by permission of instructor.

HDSC 855. Statistical Methods in Genomics Research. 3 Credits.

This survey course will provide a high-level introduction to various statistical and bioinformatics methods involved in the study of biological systems. In particular, this course will provide an overview of the analytical aspects involved in: the study DNA, RNA, and DNA methylation data measured from both microarray and next-generation sequencing (NGS) technologies. During the last week of the summer semester, students will be required to participate in a group seminar session in which they will present the results from their assigned genomics projects. Prerequisite: HDSC 820 OR experience programming in a higher level programming language; HDSC 840; OR by permission of the instructor.

HDSC 861. Observational Health Data Analysis. 3 Credits.

This course provides an understanding of the design and analysis of observational studies in health settings. This includes an introduction to common observational designs (e.g., cohort, case-control, and cross-sectional designs), sources of bias in observational analyses, and considerations for data quality and security in health data analysis. Analytic strategies for observational data are covered, including identifying and addressing bias, handling missing data, and applying multivariate

regression techniques for linear and categorical data. Prerequisite: By permission of instructor; HDSC 835 Categorical Data Analysis, HDSC 840 Linear Regression, and either HDSC 823 Introduction to Programming and Applied Statistics in R or HDSC 820 SAS Programming I.

HDSC 880. Data Mining and Analytics. 3 Credits.
Students will be introduced to common steps used in data mining, such as accessing and assaying prepared data; pattern discovery; predictive modeling using decision trees, regression, and neural networks; and model assessment methods. Prerequisite: Corequisite: HDSC 823, 835, and 840, or by permission of instructor.

HDSC 881. Statistical Learning I. 3 Credits.
Statistical learning is a fundamental skill for data scientists. Data scientists are specialists in "drinking from the firehose" of big data, and statistical learning techniques are some of their key tools. This course focuses on applications of statistical learning to big data challenges through data mining and predictive modeling techniques that are in great demand. Students will be introduced to the basics of statistical/machine learning: supervised learning (e.g. linear model, nonlinear models, penalized methods, ensemble methods, etc.), unsupervised learning (e.g. K means clustering, nearest neighbors, hierarchical clustering, etc.), and missing data in machine learning. Throughout the course, we will learn how to be "informed doers", who not only know how to apply methods but understand how those methods work. This understanding can be critical to getting good results from big data, so that the limitations of certain methods are properly understood. Prerequisite: HDSC 823, HDSC 835, HDSC 840, or by permission of instructor.

HDSC 882. Statistical Learning II. 3 Credits.
Knowledge of how and when to apply more sophisticated statistical learning models to big data can make a data scientist an indispensable asset to a research team. In Statistical Learning 2, we will learn how to be "informed doers". We will learn how many of the covered methods work, in addition to the proper situations to apply them. This is particularly important in this course, because these methods are applicable when simpler methods are inappropriate and rarely work well without significant tinkering. Data scientists with mastery of these methods are empowered to investigate questions that are far too complex to answer with the more general "workhorse" methods covered in the first unit of this series, Statistical Learning 1. We will cover many of the most important techniques in use today, including: mixture models, hidden Markov models, spline regression, support vector machines, advanced discriminant analysis methods, neural networks (including deep learning), and methods for handling highly complex computation, such as Hadoop. The course culminates with a short project that will pull together all the skills you have learned to demonstrate how they can be used for statistical decision support, which is a common task for data scientists. Prerequisite: HDSC 881, or by permission of instructor.

HDSC 883. Processing and Analysis of Medical Information Systems. 3 Credits.
Medical information systems (MIS) are essential tools of modern medicine. Healthcare data scientists must construct MIS and process the information contained in MIS. This course provides an overview of current MIS, such as electronic health record systems, clinical decision support systems, and medical imaging systems. The course will focus on theories and methodologies that support MIS construction and information processing as well as analysis, including artificial intelligence, machine learning and deep learning, knowledge representation and uncertainty reasoning, natural language processing, statistics, and medical imaging. At the end of this course, students will understand how these methodologies work and how to use these methodologies to construct MIS, process information and analyze data. Prerequisite: HDSC 881, 819 or by permission of instructor.