Biostatistics

The Division of Biostatistics engages in research, biostatistical consultation and training activities. Interested students may pursue intensive studies through the Master of Science in Biostatistics (MSIBS), the Master of Science in Biostatistics and Data Science (MSBDS), the Master of Science in Genetic Epidemiology (GEMS) (for postdocs only), the Certificate in Genetic Epidemiology, the Certificate in Biostatistics and Data Science, or other individual courses offered by the Division. Research activities include several independent lines of research as well as numerous collaborative projects with various departments of the medical school. Biostatistical consultation represents an important activity of the Division, which provides expertise in both theoretical and applied areas. The Division participates actively in postdoctoral training through a T32 postdoctoral training grant in genetic epidemiology.

The Division provides consultation through the Washington University Institute of Clinical and Translational Sciences (ICTS), the Washington University Intellectual & Developmental Disabilities Research Center, and the Biostatistics Consulting Service in a wide range of areas, including the statistical design of experiments and clinical trials, protocol development, database management, analysis of data and interpretation of results. Some of the areas of special strength and expertise include cardiovascular biostatistics, computing and statistical packages. The Division is well equipped to provide assistance with grant application preparation through services such as careful discussions of study design, sample size calculations, randomization schemes, computer resources and data analysis.

One of the Division’s specialties is statistical genetics/genetic epidemiology, and it hosts and participates in a postdoctoral T32 training grant in this area. Statistical genetics is the scientific discipline that deals with an analysis of the familial distribution of traits, with a view to understanding any possible genetic basis. However, one cannot study genes except as they are expressed in people living in certain environments, and one cannot study environmental factors except as they affect people who have certain genotypes. Statistical genetics is a unique interdisciplinary field that seeks to understand both the genetic and environmental factors and how they interact to produce various diseases and traits in humans. These studies are carried out in relatively large samples of participants in relevant populations, so population history and dynamics often come into play. Population dynamics alter the frequency and distribution of both genetic and environmental factors and, thus, their net effect on the phenotype of interest. Some population characteristics also can be exploited for the purposes of gene discovery and mapping because the history has affected the genomic structure in a way that specific genotypes associated with disease can be identified.

Human diseases have been the focal point of these studies, and recent efforts are directed toward complex disorders such as coronary heart disease, hypertension, diabetes, obesity, cancer, atopy and allergies, and neurological and psychiatric disorders, to name a few. It is commonly thought that an understanding of the genetic underpinnings of such disorders will revolutionize medicine in the 21st century, enabling better preventive measures, diagnosis, prognosis and novel treatments. Given the significant progress of the Human Genome Project, computing power, and the creation of powerful statistical methods of analysis, we are poised to shepherd this revolution. It is an exciting time in science, and opportunities for careers in statistical genetics/genetic epidemiology abound.

NIH-Sponsored Training Programs

The PRIDE Summer Institute in Cardiovascular Genetics and Epidemiology (CVD-CGE) focuses on cardiovascular and other heart, lung, blood and sleep disorders. This all-expense-paid summer institute continues during the summer of 2019 with funding from the NHLBI. The goal is to mentor junior faculty in underrepresented minorities and/or faculty with disabilities into independent research careers in the biomedical sciences. For more information, visit the PRIDE CVD-CGE website (https://biostatistics.wustl.edu/education/pridecge) or email the program administrator (biostat-pride-cge@email.wustl.edu).

The Division holds a postdoctoral T32 training grant in cardiovascular genetic epidemiology from the NIH. This training grant is available to PhDs and MDs with a background in quantitative sciences, cardiovascular sciences or experience in statistical genetics and genetic epidemiology. Candidates must be U.S. citizens or permanent residents to be eligible. For more information, visit our Research Training (https://biostatistics.wustl.edu/education/post-doctoral-research-training-in-genetic-epidemiology) webpage, contact the program administrator at 314-362-3697, or send an email (post-doc-search@wubios.wustl.edu).

For more information regarding our Genetic Epidemiology training programs, including the Master of Science in Genetic Epidemiology (GEMS) or the certificate, please visit the Genetic Epidemiology (http://bulletin.wustl.edu/medicine/departments/genetic-epidemiology) page of this Bulletin.

Academic Calendar

The academic programs begin in early July each year with preparatory workshops, followed by intensive summer semester courses. For fall and spring courses, the program follows the calendar of the College of Arts & Sciences.

Location

The program is located in the Division of Biostatistics on the fifth floor of the Bernard Becker Medical Library (660 S. Euclid Ave.), Rooms 500-508.
Additional Information
Graduate Programs
Division of Biostatistics
660 S. Euclid Ave., CB 8067
St. Louis, MO 63110-1093
Phone: 314-362-1384
Email: biostat-msibs@email.wustl.edu
Website: https://biostatistics.wustl.edu

Degrees & Requirements
Graduate Studies
The Division of Biostatistics sponsors three master’s degrees and two certificate programs:

• Master of Science in Biostatistics (MSIBS)
• Master of Science in Biostatistics and Data Science (MSBDS)
• Master of Science in Genetic Epidemiology (GEMS) (open to postdoctoral students only)
• Certificate in Biostatistics and Data Science
• Certificate in Genetic Epidemiology

The Division sponsored the GEMS program from 2002-12. In 2012, the GEMS program was streamlined as a postdoctoral degree program, and some of the curriculum was integrated into the MSIBS program. Master’s students who wish to have the GEMS type of training should look into the Statistical Genetic pathway of the MSIBS program. The MSBDS program will launch in 2019.

Master of Science in Biostatistics (MSIBS)
This 18-month, 42-credit-hour program offers excellent training in biostatistics and statistical genetics for students who earned undergraduate or higher degrees in majors in mathematics, statistics, computer science, biomedical engineering or another related field. It prepares graduates for rewarding employment in academia and industry and for further graduate studies. Students will choose between a traditional biostatistics pathway or a statistical genetics pathway. An internship is a required component of the program, and students have the option to do a thesis project or to enroll in approved elective courses. Students also have the opportunity to enhance their research and statistical training through a paid research assistant position.

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSB 503</td>
<td>Statistical Computing with SAS</td>
<td>2</td>
</tr>
<tr>
<td>MSB 560</td>
<td>Biostatistics I</td>
<td>3</td>
</tr>
<tr>
<td>MSB 506</td>
<td>Introduction to R for Data Science</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSB 570</td>
<td>Biostatistics II (1st half of semester)</td>
<td>3</td>
</tr>
<tr>
<td>MSB 515</td>
<td>Fundamentals of Genetic Epidemiology (2nd half of semester)</td>
<td>3</td>
</tr>
<tr>
<td>MSB 550</td>
<td>Introduction to Bioinformatics</td>
<td>3</td>
</tr>
<tr>
<td>Pathway course</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

Spring Year 2

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSB 617</td>
<td>Study Design and Clinical Trials</td>
<td>3</td>
</tr>
<tr>
<td>MSB 512</td>
<td>Ethics in Biostatistics and Data Science</td>
<td>2</td>
</tr>
<tr>
<td>Pathway course</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elective from approved list</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summer Year 2

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSB 630</td>
<td>Internship (3 or 6 credit hours. If 3 credit hours, student will enroll in an approved elective to replace the credit hours.)</td>
<td>3</td>
</tr>
<tr>
<td>MSB 600</td>
<td>Mentored Research (Student will enroll in Mentored Research course or 6 credit hours of electives.)</td>
<td>6</td>
</tr>
</tbody>
</table>

| Elective from approved list |       |

*For a list of approved electives (https://biostatistics.wustl.edu/education/master-of-science-in-biostatistics-msibs/curriculum-and-degree-requirements), please visit our website.

Specific courses for each pathway:

**Biostatistics**

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHS 501</td>
<td>Introduction to Epidemiology</td>
<td>3</td>
</tr>
<tr>
<td>MSB 618</td>
<td>Survival Analysis</td>
<td>3</td>
</tr>
</tbody>
</table>

**Statistical Genetics**

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSB 5483</td>
<td>Human Genetic Analysis</td>
<td>3</td>
</tr>
<tr>
<td>MSB 621</td>
<td>Computational Statistical Genetics</td>
<td>3</td>
</tr>
</tbody>
</table>

**Master of Science in Biostatistics and Data Science (MSBDS)**
This 18-month, 42-credit-hour program offers excellent training in biostatistics and data science for students who earned undergraduate or higher degrees with majors in mathematics, statistics, informatics, computer science, biomedical engineering or another related field. It prepares graduates for rewarding employment in academia and industry and for further graduate studies. Students will choose between an internship experience or a thesis project. Students also have the opportunity to enhance their research and statistical training through a paid research assistant position.
Academic Policies


Prospective Students

Those interested in applying for a training program or who would like more information may contact the program manager (biostatmsibs@email.wustl.edu).

Research

Research activities of the division span a wide range of topics relevant to a number of disease areas, and they provide research opportunities at both theoretical and applied levels. Several research projects involve close interaction and collaboration with a number of research groups at the Washington University Medical Center. Independent research programs of the division involve the genetic epidemiology of cardiovascular and metabolic diseases, bioinformatics and statistical issues in imaging sciences, and Alzheimer's disease. A number of theoretical and applied problems are addressed, including nature-nurture resolution and the identification of the genetic basis of risk factor domains such as lipids, obesity, blood pressure and hypertension, and insulin resistance and diabetes; the exploration of gene-gene and gene-environment interactions; and multivariate associations among multiple risk factors.

Our current and recent collaborative research projects include the following:

- A coordinating center for a multicenter study to assess the genetic basis of response to exercise training (HERITAGE)
- A coordinating center for a multicenter study of the effectiveness of a weight loss treatment implemented in primary care
- A coordinating center for a multicenter NETWORK study of the genetics of hypertension (HyperGEN) and the Family Blood Pressure Program (FBPP)
- Coordinating centers for a multicenter study to assess the genetic basis of response to intervention through the incorporation of gene-environment interactions (Gensalt)
- The coordinating center for the PRIDE program, with the goal of mentoring junior faculty from underrepresented minorities and/or faculty with disabilities into independent research careers in the biomedical sciences
- The coordinating center for the Data Analysis and Coordinating Center (DACC), which tracks the education and careers of people who have participated in the NHGRI Diversity Action Plan (DAP) and in NHGRI T32s that concentrate on genomics and genetics
- Important collaborative studies through support roles as biostatistics cores for the Washington University Institute of Clinical and Translational Sciences, the Alzheimer's

Certificate in Biostatistics and Data Science

The 18-credit-hour certificate is designed to prepare students to process and analyze data to effectively extract, present and interpret information from modern biomedical research and practices and to translate this new knowledge to improve health outcomes and public health. The certificate is earned after successful completion (with a minimum of B grade) of six core courses. To earn the certificate, courses must be taken over one to two consecutive years.
Disease Research Center, the Adult Children's Study, Healthy Aging and Senile Dementia (HASD), the Dominantly Inherited Alzheimer Network (DIAN), the Alvin J. Siteman Cancer Center, the Silent Infarct Transfusion Study, the Optimization of Chemotherapy for Control and Elimination of Onchocerciasis, the Washington University SporTias Center, the Washington University Intellectual & Developmental Disabilities Research Center, and Childhood Obesity Treatment

In addition, we play a significant role in studies that focus on lung transplants, asthma, chronic obstructive pulmonary disease, pediatric heart disease and ischemic heart disease. We are also part of several epidemiological research projects developing methods for increasing public awareness and utilization of measures that are known to decrease the likelihood of developing heart disease and for encouraging behaviors that will improve prognosis after a heart attack.

**Faculty**

**Division Director**

Dr. Dabeeru Rao, PhD

Visit our website for more information about our faculty (https://biostatistics.wustl.edu/faculty-staff) and their appointments.

**A**

Amber Salter Albright, PHD, BS1, M PH
Assistant Professor of Biostatistics (primary appointment)
Assistant Professor of Neurology
PHD University of Alabama-Birmingham 2015
BS1 University of Texas Austin 2002
M PH University of North Texas Heal 2005

**C**

Ling Chen, MS, PHD, MPH
Assistant Professor of Biostatistics (primary appointment)
Assistant Professor of Medicine
BS Beijing Medical University 1996
MS Beijing Medical University 1998
PHD University of MO Columbia 2009
MPH University South Carolina 2003

**G**

Charles William Goss, PHD, MS
Instructor in Biostatistics (primary appointment)
Instructor in Medicine
BS University of Michigan 2003
PHD Ohio State University 2014
BA University of Michigan 2018
MS Florida International 2018

Chi Gu, PHD, MS
Associate Professor of Biostatistics (primary appointment)
Associate Professor of Genetics

BS Nanjing Medical University 1982
PHD Washington Univ in St. Louis 1992
MS Nanjing Medical University 1985

**L**

Lei Liu, PHD, BS1, MS1, MS2
Professor of Biostatistics (primary appointment)
Professor of Medicine
PHD University of Michigan 2017
BS1 ZHEJIANG UNIVERSITY 1994
MS1 ZHEJIANG UNIVERSITY 1997
MS2 Virginia Tech 1998

**M**

J. Philip Miller
Professor of Biostatistics (primary appointment)
Professor of Medicine
Tenure Held At-Large in the Medical School
BA Washington Univ in St. Louis 1965

**R**

Dabeeru C Rao, MS, PHD
Professor of Biostatistics (primary appointment)
Director of the Division of Biostatistics
Professor of Biostatistics in Genetics
Professor of Biostatistics in Psychiatry
Professor of Mathematics
Tenure Held At-Large in the Medical School
BS Indian Statistical Institute 1967
MS Indian Statistical Institute 1968
PHD Indian Statistical Institute 1971

Treva Kay Rice, PHD, MA
Professor of Biostatistics (primary appointment)
Professor of Psychiatry
BS University of Texas Arlington 1981
PHD University of Colorado Boulder 1987
MA University of Colorado Boulder 1984

**S**

Kenneth B Schechtman, MA, PHD, MS
Professor of Biostatistics (primary appointment)
Professor of Medicine
Tenure Held At-Large in the Medical School
BS City University of New York 1967
MA Washington Univ in St. Louis 1978
PHD Washington Univ in St. Louis 1978
MS Purdue University 1971

Yun Ju Sung, PHD
Associate Professor of Biostatistics (primary appointment)
Associate Professor of Psychiatry
PHD University of Minnesota 2003
To recognize complex situational dynamics and ethical issues in their work and to develop professional and ethical problem-solving skills. The course specifically examines ethical challenges related to research design, data collection, data management, ownership, security, and sharing, data analysis and interpretation, and data reporting and provides practical guidance on these issues. The course also examines fundamentals of the broader research environment in which biostatisticians work, including principles of ethics in human subjects and animal research, and issues in biomedical research, publication and authorship, and collaboration in science. By the conclusion of the course, participants will understand the ethical and regulatory context of biomedical research; identify ethical issues, including situational dynamics that serve to foster or hinder research integrity, in the design and conduct of research and the management, analysis and reporting of data; and utilize strategies that facilitate ethical problem-solving and professionalism. Contact the program manager for details, to register, or to obtain permission of the course director (by email (biostat-msibs@email.wustl.edu) or phone: 314-362-1384).

Credit 2 units.

M21 MSB 515 Fundamentals of Genetic Epidemiology

Intensive two-week summer course. Lectures cover causes of phenotypic variation, familial resemblance and heritability, Hardy-Weinberg Equilibrium, ascertainment, study designs and basic concepts in genetic segregation, linkage and association. The computer laboratory portion is designed as hands-on practice of fundamental concepts. Students will gain practical experience with various genetics computer programs (e.g., SOLAR, MERLIN, QTDT, and PLINK). Auditors will not have access to the computer lab sessions. Prerequisite: R Primer. Contact the program manager (biostat-msibs@email.wustl.edu) for details, to register, or to obtain permission from the course director.

Credit 3 units.

M21 MSB 5483 Human Genetic Analysis

Basic Genetic Concepts: meiosis, inheritance, Hardy-Weinberg equilibrium, linkage, segregation analysis; Linkage Analysis: definition, crossing over, map functions, phase, LOD scores, penetrance, phenocopies, liability classes, multipoint analysis, nonparametric analysis (sibpairs and pedigrees), quantitative trait analysis, determination of power for Mendelian and complex trait analysis; Linkage Disequilibrium Analyses: allelic association (case control designs and family based studies), QQ and Manhattan plots, whole genome association analysis; population stratification; Quantitative Trait Analysis: measured genotypes and variance components. Hands-on computer lab experience doing parametric linkage analysis with the program LINKAGE, model free linkage analyses with GeneHunter and Merlin, power computations with SLINK, quantitative trait analyses with SOLAR, LD computations with Haploview and WGAViewer, and family-based and case-control association analyses with PLINK and SAS. The methods and exercises are coordinated with the lectures, and students are expected to understand underlying assumptions and limitations and the basic calculations performed by these computer programs. Auditors will not have access to the computer lab sessions. Prerequisite: M21-515 Fundamentals of Genetic Epidemiology. For details, to register, and to receive the required permission of the course director, contact the MSIBS program manager (by email (biostat-msibs@email.wustl.edu) or phone: 314-362-1384).

Same as L41 Biol 5483

Credit 3 units.

Courses


M21 MSB 503 Statistical Computing with SAS

Intensive hands-on summer training in SAS (Statistical Analysis System) during seven full weekdays. Students will learn how to use SAS for handling, managing, and analyzing data. Instruction is provided in the use of SAS programming language, procedures, macros, and SAS SQL. The course will include exercises using existing programs written by SAS experts. Contact the Program Managers for details, to register or to obtain permission of the Course Master (biostat-msibs@email.wustl.edu).

Credit 2 units.

M21 MSB 506 Introduction to R for Data Science

This is an introduction to the R Statistical Environment for new users. R is “a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modeling, statistical tests, time series analysis, classification, clustering, etc.” The goal is to give students a set of tools to perform statistical analysis in medicine, biology, or epidemiology. At the conclusion of this primer, students will: be able to manipulate and analyze data, write basic models, understand the R environment for using packages, and create standard or customized graphics. This primer assumes some knowledge of basic statistics as taught in a first-semester undergraduate or graduate sequence. Topics should include: probability, cross-tabulation, basic statistical summaries, and linear regression in either scalar or matrix form. Contact the program manager (biostat-msibs@email.wustl.edu) for details, to register, or to obtain permission from the course director.

Credit 2 units.

M21 MSB 512 Ethics in Biostatistics and Data Science

This course prepares biostatisticians to analyze and address ethical and professional issues in the practice of biostatistics across the range of professional roles and responsibilities of a biostatistician. The primary goals are for biostatisticians to recognize complex situational dynamics and ethical issues in their work and to develop professional and ethical problem-solving skills. The course specifically examines ethical challenges related to research design, data collection, data management, ownership, security, and sharing, data analysis and interpretation, and data reporting and provides practical guidance on these issues. The course also examines fundamentals of the broader research environment in which biostatisticians work, including principles of ethics in human subjects and animal research, and issues in biomedical research, publication and authorship, and collaboration in science. By the conclusion of the course, participants will understand the ethical and regulatory context of biomedical research; identify ethical issues, including situational dynamics that serve to foster or hinder research integrity, in the design and conduct of research and the management, analysis and reporting of data; and utilize strategies that facilitate ethical problem-solving and professionalism. Contact the program manager for details, to register, or to obtain permission of the course director (by email (biostat-msibs@email.wustl.edu) or phone: 314-362-1384).

Credit 3 units.
M21 MSB 550 Introduction to Bioinformatics
Provide a broad exposure to the basic concepts, methodology and application of bioinformatics to solve biological problems. Specifically, the students will learn the basics of online genomic/protein databases and database mining tools, and acquire understanding of mathematical algorithms in genome sequence analysis (alignment analysis, gene finding/predicting), gene expression microarray (genechip) analysis, and of the impact of recent developments in the protein microarray technology. Prerequisite: R Primer. Contact the program manager (biostat-msibs@email.wustl.edu) for details, to register, or to obtain permission from the course director.
Credit 3 units.

M21 MSB 560 Biostatistics I
This course is designed for students who want to develop a working knowledge of basic methods in biostatistics. The course is focused on biostatistical and epidemiological concepts and on practical hints and hands-on approaches to data analysis rather than on details of the theoretical methods. We will cover basic concepts in hypothesis testing, will introduce students to several of the most widely used probability distributions, and will discuss classical statistical methods that include t-tests, chi-square tests, regression analysis, and analysis of variance. Both in-class examples and homework assignments will involve extensive use of SAS. Prerequisite: M21-503, Statistical Computing with SAS®, or student must have good practical experience with SAS®. Participants are strongly encouraged to participate in the "Computing/Unix" and "Statistics" workshops offered free of charge prior to this course. For details, to register, and/or to obtain the required permission of the course director, contact the program manager (by email biostat-msibs@email.wustl.edu) or phone: 314-362-1384).
Credit 3 units.

M21 MSB 570 Biostatistics II
This course is designed for students who have taken Biostatistics I or the equivalent and who want to extend their knowledge of biostatistical applications to more modern and more advanced methods. Biostatistical methods to be discussed include logistic and Poisson regression, survival analysis, Cox regression analysis, and several methods for analyzing longitudinal data. Students will be introduced to modern topics that include statistical genetics and bioinformatics. The course will also discuss clinical trial design, the practicalities of sample size and power computation and meta analysis, and will ask students to read journal articles with a view toward encouraging a critical reading of the medical literature. Both in-class examples and homework assignments will involve extensive use of SAS. Prerequisite: M21-560, Biostatistics I or its equivalent as judged by the course directors. For details, to register, and/or to obtain the required permission of the course director, contact the program manager (by email biostat-msibs@email.wustl.edu) or phone: 314-362-1384).
Credit 3 units.

M21 MSB 600 Mentored Research
Student undertakes supervised research in a mentor’s lab. The goal is to acquire important research skills as well as good writing and presentation skills. The student finds a mentor who is willing to work with them, and they together identify a research topic. A written thesis based on the research, prepared in the format of an actual scientific publication, must be submitted and presented to a select audience. The course directors will organize a few meetings throughout to facilitate the whole process. The course directors will determine the grade (pass/fail) in consultation with the mentors. Permission of the course directors is required.
Credit variable, maximum 6 units.

M21 MSB 617 Study Design and Clinical Trials
The course will focus on statistical and epidemiological concepts of study design and clinical trials. Topics include: different phases of clinical trials, various types of medical studies (observational studies, retrospective studies, adaptive designs, and comparative effectiveness research), and power analysis. Study management and ethical issues are also addressed. Students will be expected to do homework and practice power analysis during lab sessions. Prerequisites: M21-560 Biostatistics I and M21-570 Biostatistics II. Permission of the course director required. For details, to register, and to receive the required permission of the course director, contact the program manager (biostat-msibs@email.wustl.edu).
Credit 3 units.

M21 MSB 618 Survival Analysis
This course will cover the basic applied and theoretical aspects of models to analyze time-to-event data. Basic concepts will be introduced including the hazard function, survival function, right censoring, and the Cox-proportional hazards (PH) model with fixed and time dependent covariates. Additional topics will include regression diagnostics for survival models, the stratified PH model, additive hazards regression models and multivariate survival models. Permission of the course director required. Prerequisites: M21-560 Biostatistics I and M21-570 Biostatistics II. For details, to register, and to receive permission from the course director, contact the program manager (by email biostat-msibs@email.wustl.edu) or phone: 314-362-1384).
Credit 3 units.

M21 MSB 621 Computational Statistical Genetics
This course is designed to give the students computational experience with the latest statistical genetics methods and concepts, so that they will be able to computationally implement the method(s)/model(s) developed as part of their thesis. Concentrating on the applications of genomics and SAS computing, it deals with creating efficient new bioinformatic tools to interface with some of the latest, most important genetic epidemiological analysis software, as well as how to derive, design and implement new statistical genetics models. The course also includes didactic instruction on haplotype estimation and modeling of relationship to phenotype, LD mapping, DNA pooling analysis methods, analysis approaches in pharmacogenomics (with an emphasis on possible genomic role in drug response heterogeneity), and epistasis (GxG) and GxE interactions; data mining methods, including clustering, recursive partitioning, boosting, and random forests; and fundamentals of meta-analysis, importance sampling, permutation tests and empirical p-values, as well as the design of monte-carlo simulation experiments. Prerequisite: permission of the instructor. Contact the program manager for the required permission of the course director (by email biostat-msibs@email.wustl.edu) or phone: 314-362-1384).
Credit 3 units.
M21 MSB 630 Internship
The primary goal of the Internship program is for all students to acquire critical professional experience so that they will be well prepared to enter the job market upon graduation. This provides an opportunity for students to develop contacts, build marketable skills, and perceive likes and dislikes in the chosen field. Students will have an opportunity to work with experienced mentors (PIs) on a range of projects that may include data management, data analysis, study design, and protocol development, among other things. Students may have opportunities to contribute to and participate in the preparation of publishable quality manuscripts. As part of the Internship requirements, each student will submit a one-page Abstract of the work performed as part of the internship and will give a 5-minute presentation of the Internship experience. Internship presentations will be scheduled in late summer. The grade (pass/fail) for each student will be determined in consultation with the mentor. The internship is offered during the student's second summer. In extremely unusual circumstances and when the students' prior training justifies it, students can petition the internship committee to complete the internship during the spring or second fall semester. Approval of the committee is required. Credit variable, maximum 6 units.

M21 MSB 660 Biomedical Data Mining
This course introduces methods and applications in biomedical data mining. Various computational and statistical methods will be introduced, such as data wrangling and visualizations, model selection and regularization, and tree-based methods. In addition to the common applications of the covered methods in biomedical sciences, this course will prepare students for future challenges and opportunities in data science. Prerequisites: M21 506, M21 560, M21 570, and M21 550. Matrix algebra is also highly recommended. Credit 3 units.