

# Statistics

Courses offered by the Department of Statistics are listed under the subject code STATS on the Stanford Bulletin's ExploreCourses (<http://explorecourses.stanford.edu/CourseSearch/search?view=catalog&catalog=&page=0&q=STATS&filter-catalognumber-STATS=on>) web site.

The department's goals are to acquaint students with the role played in science and technology by probabilistic and statistical ideas and methods, to provide instruction in the theory and application of techniques that have been found to be commonly useful, and to train research workers in probability and statistics. There are courses for general students as well as those who plan careers in statistics in business, government, industry, and teaching.

The requirements for a degree in Statistics are flexible, depending on the needs and interests of the students. Some students may be interested in the theory of statistics and/or probability, whereas other students may wish to apply statistical and probabilistic methods to a substantive area. The department has long recognized the relation of statistical theory to applications. It has fostered this by encouraging a liaison with other departments in the form of joint and courtesy faculty appointments: Economics (Anderson, Romano), Education (Olkin, Rogosa), Electrical Engineering (Montanari), Geological and Environmental Sciences (Rajaratnam, Switzer), Health Research and Policy (Efron, Hastie, Johnstone, Lavori, Olshen, Tibshirani, Wong), Mathematics (Candés, Dembo, Diaconis), Political Science (Jackman), and the SLAC National Accelerator Laboratory (Friedman). The research activities of the department reflect an interest in applied and theoretical statistics and probability. There are workshops in biology/medicine and in environmental factors in health.

In addition to courses for Statistics students, the department offers a number of service courses designed for students in other departments. These tend to emphasize the application of statistical techniques rather than their theoretical development.

The department has always drawn visitors from other countries and universities. As a consequence, there is usually a wide range of seminars offered by both the visitors and the department's own faculty.

## Undergraduate Programs in Statistics

### Majoring in Statistics

Students wishing to build a concentration in probability and statistics are encouraged to consider declaring a major in Mathematical and Computational Science (<http://web.stanford.edu/group/mathcompsci>). This interdepartmental program is administered in the Department of Statistics and provides core training in computing, mathematics, operations research, and statistics, with opportunities for further elective work and specialization. See the "Mathematical and Computational Science" section of this bulletin.

### Graduate Programs in Statistics

University requirements for the M.S. and Ph.D. degrees are discussed in the "Graduate Degrees (<http://exploreddegrees.stanford.edu/graduatedegrees>)" section of this bulletin.

## Learning Outcomes (Graduate)

The purpose of the master's program is to further develop knowledge and skills in Statistics and to prepare students for a professional career or doctoral studies. This is achieved through completion of courses, in the primary field as well as related areas, and experience with independent work and specialization.

The Ph.D. is conferred upon candidates who have demonstrated substantial scholarship and the ability to conduct independent research and analysis in Statistics. Through completion of advanced course work and rigorous skills training, the doctoral program prepares students to make original contributions to the knowledge of Statistics and to interpret and present the results of such research.

## Minor in Statistics

The undergraduate minor in Statistics is designed to complement major degree programs primarily in the social and natural sciences. Students with an undergraduate Statistics minor should find broadened possibilities for employment. The Statistics minor provides valued preparation for professional degree studies in postgraduate academic programs.

The minor consists of a minimum of six courses with a total of at least 20 units. There are two required courses (8 units) and four qualifying or elective courses (12 or more units). All courses for the minor must be taken for a letter grade. An overall 2.75 grade point average (GPA) is required for courses fulfilling the minor.

## Required Courses

|           |                                       | Units |
|-----------|---------------------------------------|-------|
| STATS 116 | Theory of Probability                 | 3-5   |
| STATS 200 | Introduction to Statistical Inference | 3     |

## Qualifying Courses

At most, one of these two courses may be counted toward the six course requirement for the minor:

|           |  | Units |
|-----------|--|-------|
| MATH 52   | Integral Calculus of Several Variables | 5     |
| STATS 191 | Introduction to Applied Statistics     | 3-4   |

## Elective Courses

At least one of the elective courses should be a STATS 200-level course. The remaining two elective courses may also be 200-level courses. Alternatively, one or two elective courses may be approved courses in other departments. Special topics courses and seminars for undergraduates are offered from time to time by the department, and these may be counted toward the course requirement. Students may not count any Statistics courses below the 100 level toward the minor.

Examples of elective course sequences are:

|   |  | Units |
|---|--|-------|
| <b>Data Analysis and Applied Statistics</b> |  |       |
| STATS 202                                   | Data Mining and Analysis                                   | 3     |
| STATS 203                                   | Introduction to Regression Models and Analysis of Variance | 3     |
| <b>Statistical Methodology</b>              |  |       |
| STATS 205                                   | Introduction to Nonparametric Statistics                   | 3     |
| STATS 206                                   | Applied Multivariate Analysis                              | 3     |
| STATS 207                                   | Introduction to Time Series Analysis                       | 3     |
| <b>Economic Optimization</b>                |  |       |

|  |  |     |
|--|--|-----|
| STATS 206                                  | Applied Multivariate Analysis                                      | 3   |
| ECON 160                                   | Game Theory and Economic Applications                              | 5   |
| <b>Psychology Modeling and Experiments</b> |  |     |
| STATS 206                                  | Applied Multivariate Analysis                                      | 3   |
| <b>Signal Processing</b>                   |  |     |
| STATS 207                                  | Introduction to Time Series Analysis                               | 3   |
| EE 264                                     | Digital Signal Processing  | 3   |
| EE 279                                     | Introduction to Digital Communication                              | 3   |
| <b>Genetic and Ecologic Modeling</b>       |  |     |
| STATS 217                                  | Introduction to Stochastic Processes                               | 3   |
| BIO 283                                    | Theoretical Population Genetics                                    | 3   |
| <b>Probability and Applications</b>        |  |     |
| STATS 217                                  | Introduction to Stochastic Processes                               | 3   |
| STATS 218                                  | Introduction to Stochastic Processes                               | 3   |
| <b>Mathematical Finances</b>               |  |     |
| STATS 240                                  | Statistical Methods in Finance                                     | 3-4 |
| STATS 243                                  | Financial Models and Statistical Methods in Active Risk Management | 3-4 |
| STATS 250                                  | Mathematical Finance   | 3   |

## Master of Science in Statistics

The department requires that a master's student take 45 units of work from offerings in the Department of Statistics (<http://explorecourses.stanford.edu/search?view=catalog&filter-coursestatus-Active=on&page=0&catalog=&academicYear=&q=STATS&collapse=>) or from authorized courses in other departments. With the advice of the master's program advisers, each student selects his or her own set of electives.

All requirements for the Statistics master's degree, including the coterminal master's degree, must be completed within three years of their first quarter of graduate standing. Ordinarily, four or five quarters are needed to complete all requirements. Honors Cooperative students must finish within five years.

Units for a given course may not be counted to meet the requirements of more than one degree, with the exception that up to 45 units of a Stanford M.A. or M.S. degree may be applied to the residency requirement for the Ph.D., D.M.A. or Engineer degrees. (GAP 3.2 (<http://gap.stanford.edu/3-2.html>))

## University Coterminal Requirements

Coterminal master's degree candidates are expected to complete all master's degree requirements as described in this bulletin. University requirements for the coterminal master's degree are described in the "Coterminal Master's Program (<http://exploreddegrees.stanford.edu/cotermdegrees>)" section. University requirements for the master's degree are described in the "Graduate Degrees (<http://exploreddegrees.stanford.edu/graduatedegrees/#masterstext>)" section of this bulletin.

After accepting admission to this coterminal master's degree program, students may request transfer of courses from the undergraduate to the graduate career to satisfy requirements for the master's degree. Transfer of courses to the graduate career requires review and approval of both the undergraduate and graduate programs on a case by case basis.

In this master's program, courses taken three quarters prior to the first graduate quarter, or later, are eligible for consideration for transfer to the graduate career. No courses taken prior to the first quarter of the sophomore year may be used to meet master's degree requirements.

Course transfers are not possible after the bachelor's degree has been conferred.

The University requires that the graduate adviser be assigned in the student's first graduate quarter even though the undergraduate career may still be open. The University also requires that the Master's Degree Program Proposal be completed by the student and approved by the department by the end of the student's first graduate quarter.

Students must submit a completed Coterminal Course Approval Form ([http://studentaffairs.stanford.edu/sites/default/files/registrar/files/Coterm\\_Course\\_Approval.pdf](http://studentaffairs.stanford.edu/sites/default/files/registrar/files/Coterm_Course_Approval.pdf)) with their Application for Admission to Coterminal Master's Program (<http://registrar.stanford.edu/pdf/CotermApplic.pdf>) indicating which courses must be transferred from the student's undergraduate to graduate career

For further information about the Statistics master's degree program requirements, see the department web site (<https://statistics.stanford.edu/academics/ms-statistics>). (<https://statistics.stanford.edu/masters-program-proposal-form>)

Students must earn a 3.0 GPA in the following M.S. degree requirements:

### 1. Statistics core courses (must complete all four courses):

|           |                                       | Units |
|-----------|---------------------------------------|-------|
| STATS 116 | Theory of Probability                 | 3-5   |
| STATS 191 | Introduction to Applied Statistics    | 3-4   |
| STATS 200 | Introduction to Statistical Inference | 3     |
| STATS 217 | Introduction to Stochastic Processes  | 2-3   |

All must be taken for a letter grade. Students with prior background may replace each course with a more advanced course from the same area. Courses previously taken may be waived by the adviser, in which case they must be replaced by other graduate courses offered by the department.

### 2. Linear Algebra Mathematics requirement:

|                              |                                  | Units |
|------------------------------|----------------------------------|-------|
| Select one of the following: |                                  |       |
| MATH 104                     | Applied Matrix Theory            | 3     |
| MATH 113                     | Linear Algebra and Matrix Theory | 3     |
| MATH 115                     | Functions of a Real Variable     | 3     |
| MATH 171                     | Fundamental Concepts of Analysis | 3     |

All must be taken for a letter grade. Substitution of other courses in Mathematics and Computer Science may be made with consent of the adviser.

### 3. Programming requirement:

|                              |  | Units |
|------------------------------|--|-------|
| Select one of the following: |  |       |
| CS 106A                      | Programming Methodology                | 3     |
| CS 106B                      | Programming Abstractions               | 3     |
| CS 106X                      | Programming Abstractions (Accelerated) | 3     |
| CME 108                      | Introduction to Scientific Computing   | 3     |

All must be taken for a letter grade. Substitution of other courses in Mathematics and Computer Science may be made with consent of the adviser.

### 4. Additional Statistics Courses

At least four additional Statistics courses must be taken from graduate offerings in the department (STATS 202 through 390). All must be taken for a letter grade, if offered. Students cannot count more than a total 6 units of the following toward the master's degree requirements:

|            |                           | Units |
|------------|---------------------------|-------|
| STATS 260A | Workshop in Biostatistics | 1-2   |

|            |                                       |      |
|------------|---------------------------------------|------|
| STATS 260B | Workshop in Biostatistics             | 1-2  |
| STATS 260C | Workshop in Biostatistics             | 1-2  |
| STATS 298  | Industrial Research for Statisticians | 1    |
| STATS 299  | Independent Study                     | 1-10 |
| STATS 390  | Consulting Workshop                   | 1    |

## 5. Elective Courses

Additional elective units to complete the requirements may be chosen from the list available from the department web site (<https://statistics.stanford.edu/academics/ms-statistics-elective-courses>). Other graduate courses (200 or above) may be authorized by the adviser if they provide skills relevant to statistics or deal primarily with an application of statistics or probability and do not overlap courses in the student's program. There is sufficient flexibility to accommodate students with interests in applications to business, computing, economics, engineering, health, operations research, and biological and social sciences.

Courses below 200 level are not acceptable, with the following exceptions:

|           |   | Units |
|-----------|---|-------|
| STATS 116 | Theory of Probability   | 3-5   |
| STATS 191 | Introduction to Applied Statistics                              | 3-4   |
| MATH 104  | Applied Matrix Theory   | 3     |
| MATH 113  | Linear Algebra and Matrix Theory                                | 3     |
| MATH 115  | Functions of a Real Variable                                    | 3     |
| MATH 171  | Fundamental Concepts of Analysis                                | 3     |
| MATH 180  | Introduction to Financial Mathematics                           | 3     |
| CS 106A   | Programming Methodology   | 3-5   |
| CS 106B   | Programming Abstractions  | 3-5   |
| CS 106X   | Programming Abstractions (Accelerated)                          | 3-5   |
| CS 140    | Operating Systems and Systems Programming                       | 3-4   |
| CS 142    | Web Applications  | 3     |
| CS 143    | Compilers   | 3-4   |
| CS 144    | Introduction to Computer Networking                             | 3-4   |
| CS 145    | Introduction to Databases                                       | 3-4   |
| CS 147    | Introduction to Human-Computer Interaction Design               | 3-5   |
| CS 148    | Introduction to Computer Graphics and Imaging                   | 3-4   |
| CS 149    | Parallel Computing  | 3-4   |
| CS 154    | Introduction to Automata and Complexity Theory                  | 3-4   |
| CS 155    | Computer and Network Security                                   | 3     |
| CS 157    | Logic and Automated Reasoning                                   | 3     |
| CS 161    | Design and Analysis of Algorithms                               | 3-5   |
| CS 170    | Stanford Laptop Orchestra: Composition, Coding, and Performance | 1-5   |
| CS 181    | Computers, Ethics, and Public Policy                            | 4     |

**At most, one of these courses may be counted:**

|           |                                    | Units |
|-----------|------------------------------------|-------|
| MATH 104  | Applied Matrix Theory              | 3     |
| MATH 113  | Linear Algebra and Matrix Theory   | 3     |
| MATH 151  | Introduction to Probability Theory | 3     |
| STATS 116 | Theory of Probability              | 3-5   |

## 6. Master's Degree Program Proposal Form (degree milestone)

This form (<https://statistics.stanford.edu/masters-program-proposal-form>) is to be submitted by the student to the major department's student services administrator prior to the end of the first quarter of enrollment in the

program. A revised program proposal must be submitted if your degree plans change.

There is no thesis requirement.

Students with a strong mathematical background who may wish to go on to a Ph.D. in Statistics should consider applying to the Ph.D. program.

## Master of Science in Statistics: Data Science (subplan)

The Department of Statistics and ICME have collaborated on a new specialization/subplan for the Master in Science degree focusing on big data in engineering and applied sciences. Students in the program will develop strong mathematical, statistical, computational, and programming skills through the ICME M.S. requirements and will gain a fundamental data science education by focusing 18 units of elective courses in the area of data science and related courses. Upon completion of the M.S. in Statistics with a specialization/subplan in Data Science, students will be prepared to continue on to their Ph.D. in Computer Science, ICME, or as a data science professional in industry.

The M.S. in Data Science specialization/subplan is overseen by a steering committee comprised of ICME and Statistics faculty members. Current members are Professors Guenther Walther, Trevor Hastie, Emmanuel Candes, and Margot Gerritsen.

Applicants will apply to the M.S. program in Statistics and declare their preference for the Data Science subplan within the application ("Department Specialization" option). Selection of the students is made by the Statistics admission committee, which has representation from the Data Science steering committee.

A Master's degree program proposal (<https://statistics.stanford.edu/stats-ds-program-proposal-form-pdf>), is to be submitted by the student to the major department's student services administrator prior to the end of the first quarter of enrollment in the program. A revised program proposal must be submitted if your degree plans change.

(Subplans are printed on the transcript and diploma.)

## Curriculum and Degree Requirements

The course work follows the requirements of the traditional ICME M.S. degree with additional restrictions placed on the general and focused electives. As defined in the general graduate student requirements, students must maintain a grade point average (GPA) of 3.0 or better and classes must be taken at the 200 level or higher. Students must complete 45 units of required coursework in Data Science.

Students must demonstrate breadth of knowledge in the field by completing the following core courses. Courses in this area must be taken for letter grade.

|         |                                     | Units |
|---------|-------------------------------------|-------|
| CME 302 | Numerical Linear Algebra            | 3     |
| CME 304 | Numerical Optimization              | 3     |
| CME 305 | Discrete Mathematics and Algorithms | 3     |

In addition to the three core courses, the students are required to take a course in Stochastics. They can take either CME 308 Stochastic Methods in Engineering or an equivalent course approved by the steering committee. Must be taken for a letter grade.

## Requirement 2: Advanced Scientific Programming and High Performance Computing Core (6 units)

To ensure that students have a strong foundation in programming, all students will be required to take 6 units of advanced programming, with at least 3 units in parallel computing. Courses in this area must be taken for letter grade.

| Approved Advanced Programming courses: (3 units)   |  | Units |
|--|--|-------|
| CME 212  | Advanced Programming for Scientists and Engineers              | 3     |
| CME 214  | Software Design in Modern Fortran for Scientists and Engineers | 3     |
| CS 107   | Computer Organization and Systems                              | 3-5   |
| CS 249B  | Large-scale Software Development                               | 3     |
| Approved Parallel Computing/HCP courses: (3 units) |  |       |
| CME 213  | Introduction to parallel computing using MPI, openMP, and CUDA | 3     |
| CME 342  | Parallel Methods in Numerical Analysis                         | 3     |
| CS 149   | Parallel Computing   | 3-4   |
| CS 315A  | Parallel Computer Architecture and Programming                 | 3     |
| CS 316   | Advanced Multi-Core Systems                                    | 3     |

Students who do not start the program with a strong computational and/or programming background will take an extra 3 units to prepare themselves by, for example, taking CME 211 Software Development for Scientists and Engineers or an equivalent course, such as CS106A/B/X. For Data Science track students, the 1-unit course in MapReduce offered by ICME annually is also highly recommended.

## Requirement 3: Statistics Core (12 units)

Courses in this area must be taken for letter grade.

The curriculum for the Data Science track requires 12 units of focused coursework in Statistics consisting of the following courses:

|            |  | Units |
|------------|--|-------|
| STATS 200  | Introduction to Statistical Inference                      | 3     |
| STATS 203  | Introduction to Regression Models and Analysis of Variance | 3     |
| or         |  |       |
| STATS 305  | Introduction to Statistical Modeling                       | 3     |
| STATS 315A | Modern Applied Statistics: Learning                        | 2-3   |
| STATS 315B | Modern Applied Statistics: Data Mining                     | 2-3   |

or equivalent courses as approved by the steering committee.

Of the following 15 units in Requirements **Four** and **Five** combined, 6 units must be taken for a letter grade.

## Requirement 4: Domain Specialization or preparatory courses (9 units)

Three courses in specialized areas. One or two of these courses may be used by the students that enter the program with insufficient linear algebra or programming experience to prepare for the core requirements in the MS track.

Specialized courses include courses that further deepen the data science core. Some possibilities include:

|         |  | Units |
|---------|--|-------|
| CS 347  | Parallel and Distributed Data Management | 3     |
| CS 448  | Topics in Computer Graphics              | 3-4   |
| CS 224W | Social Information and Network Analysis  | 3-4   |

|              |  |     |
|--------------|--|-----|
| STATS 366    | Modern Statistics for Modern Biology                               | 3   |
| BIOS 221     | Modern Statistics for Modern Biology                               |     |
| PSYCH 204A   | Human Neuroimaging Methods   | 3   |
| PSYCH 303    | Human and Machine Hearing  | 3   |
| OIT 367      | Business Intelligence from Big Data                                | 4   |
| BIOMEDIN 215 | Data Driven Medicine   | 3   |
| ENERGY 240   | Geostatistics  | 2-3 |
| BIOE 214     | Representations and Algorithms for Computational Molecular Biology | 3-4 |

## Requirement 5: Practical component (6 units)

The students need 6 units of practical component that may include any combination of:

- Capstone project, supervised by a faculty member and approved by the steering committee: the capstone project should be computational in nature; students should submit a one-page proposal, supported by the faculty member, to the steering committee (gwalther@stanford.edu) for approval.
- Clinics, such as the Stanford Data Science Challenge Lab ENGR 250 Data Challenge Laboratory and Data Science Impact Lab ENGR 350 Data Impact Laboratory
- Other courses that have a strong hands-on and practical component, such as STATS 390 Statistical Consulting (up to 3 units).

## Doctor of Philosophy in Statistics

The department looks for students who wish to prepare for research careers in statistics or probability, either applied or theoretical. Advanced undergraduate or master's level work in mathematics and statistics provides a good background for the doctoral program. Quantitatively oriented students with degrees in other scientific fields are also encouraged to apply for admission. The program normally takes five years to complete.

## Program Summary

|                         |  | Units |
|-------------------------|--|-------|
| First-year core program |  |       |
| STATS 300               | Advanced Topics in Statistics (offered Summer Quarter) | 2-3   |
| STATS 300A              | Theory of Statistics                                   | 2-3   |
| STATS 300B              | Theory of Statistics                                   | 2-4   |
| STATS 300C              | Theory of Statistics                                   | 2-4   |
| STATS 305               | Introduction to Statistical Modeling                   | 3     |
| STATS 306A              | Methods for Applied Statistics                         | 3     |
| STATS 306B              | Methods for Applied Statistics: Unsupervised Learning  | 2-3   |
| STATS 310A              | Theory of Probability                                  | 2-4   |
| STATS 310B              | Theory of Probability                                  | 2-3   |
| STATS 310C              | Theory of Probability                                  | 2-4   |

- Pass two of three parts of the qualifying examinations (end of first year); breadth requirement (second and third year); successfully complete the thesis proposal meeting (before end of third year); pass the University oral examination (fourth or fifth year); dissertation (fifth year).
- In addition, students are required to take nine units of advanced topics courses offered by the department. Recommended courses include the following:

|            |  | <b>Units</b> |
|------------|--|--------------|
| STATS 314A | Advanced Statistical Theory                              | 3            |
| STATS 314B | Topics in Minimax Inference of Nonparametric Functionals | 3            |
| STATS 315A | Modern Applied Statistics: Learning                      | 2-3          |
| STATS 315B | Modern Applied Statistics: Data Mining                   | 2-3          |
| STATS 317  | Stochastic Processes                                     | 3            |
| STATS 318  | Modern Markov Chains                                     | 3            |
| STATS 330  | An Introduction to Compressed Sensing                    | 3            |
| STATS 370  | A Course in Bayesian Statistics                          | 3            |
| STATS 376A | Information Theory                                       | 3            |
| STATS 376B | Network Information Theory                               | 3            |
| EE 364A    | Convex Optimization I                                    | 3            |

- Complete a minimum of three units of STATS 390 Consulting Workshop, taking it at least twice.
- Take STATS 319 Literature of Statistics once per year after passing the Qualifying Exam until the year after passing the dissertation proposal meeting.

## First-Year Core Courses

- STATS 300 Advanced Topics in Statistics systematically surveys the ideas of estimation and of hypothesis testing for parametric and nonparametric models involving small and large samples.
- STATS 305 Introduction to Statistical Modeling is concerned with linear regression and the analysis of variance.
- STATS 306A Methods for Applied Statistics and STATS 306B Methods for Applied Statistics: Unsupervised Learning survey a large number of modeling techniques, related to but going beyond the linear models of STATS 305 Introduction to Statistical Modeling.
- STATS 310A Theory of Probability, STATS 310B Theory of Probability, and STATS 310C Theory of Probability are measure-theoretic courses in probability theory, beginning with basic concepts of the law of large numbers and martingale theory.
  - Students who do not have enough mathematics background can take STATS 310A,B,C after their first year but need to have their first-year program approved by the Ph.D. program adviser.

## Qualifying Examinations

These are intended to test the student's level of knowledge when the first-year program, common to all students, has been completed. There are separate examinations in the three core subjects of statistical theory and methods, applied statistics, and probability theory, and all are typically taken during the summer between the student's first and second years. Students are expected to show acceptable performance in two examinations. Letter grades are not given. After passing the qualifying exams, students will file for Ph.D. candidacy, a University milestone.

## Breadth Requirement

Students are required to take 15 units of coursework outside of the department and are advised to choose an area of concentration in a specific scientific field of statistical applications approved by the Ph.D. program adviser.

Popular areas with suggested course options include:

## Computational Biology and Statistical Genomics

Students are expected to take 9 units of graduate courses in genetics or neurosciences (imaging), such as GENE 203/BIO 203 (Advanced

Genetics), as well as 9 units of classes in Statistical Genetics or Bioinformatics:

|   |   | <b>Units</b> |
|---|---|--------------|
| Courses can be chosen from the following list: <sup>1</sup> |   |              |
| STATS 345   | Statistical and Machine Learning Methods for Genomics | 3            |
| STATS 366   | Modern Statistics for Modern Biology                  | 3            |
| STATS 367   | Statistical Models in Genetics                        | 3            |

<sup>1</sup> The following courses are not offered this year but may be used by students who completed them in fulfillment of this requirement: STATS 345, STATS 367.

## Machine Learning

Courses can be chosen from the following list:

| <b>Statistical Learning</b>                                   |  |     |
|---|--|-----|
| STATS 315A  | Modern Applied Statistics: Learning                | 2-3 |
| STATS 315B  | Modern Applied Statistics: Data Mining             | 2-3 |
| <b>Data Bases</b>   |  |     |
| CS 245  | Database Systems Principles                        | 3   |
| CS 346  | Database System Implementation                     | 3-5 |
| CS 347  | Parallel and Distributed Data Management           | 3   |
| <b>Probabilistic Methods in AI</b>                            |  |     |
| CS 221  | Artificial Intelligence: Principles and Techniques | 3-4 |
| CS 354  | Topics in Circuit Complexity                       | 3   |
| <b>Statistical Learning Theory and Pattern Classification</b> |  |     |
| CS 229  | Machine Learning                                   | 3-4 |

## Applied Probability

Students are expected to take 15 units of graduate courses in some of the following areas:

| <b>Control and Stochastic Calculus</b> |   |   |
|--|---|---|
| MSE 322                                | Stochastic Calculus and Control                       | 3 |
| MSE 351                                | Dynamic Programming and Stochastic Control            | 3 |
| MATH 237                               | Default and Systemic Risk                             | 3 |
| <b>Finance</b>                         |   |   |
| STATS 250                              | Mathematical Finance                                  | 3 |
| FINANCE 622                            | Dynamic Asset Pricing Theory                          | 4 |
| MATH 236                               | Introduction to Stochastic Differential Equations     | 3 |
| <b>Information Theory</b>              |   |   |
| EE 376A                                | Information Theory                                    | 3 |
| EE 376B                                | Network Information Theory                            | 3 |
| <b>Monte Carlo</b> <sup>1</sup>        |   |   |
| STATS 318                              | Modern Markov Chains                                  | 3 |
| STATS 345                              | Statistical and Machine Learning Methods for Genomics | 3 |
| STATS 362                              | Topic: Monte Carlo                                    | 3 |
| <b>Queueing Theory</b>                 |   |   |
| MSE 335                                | Queueing and Scheduling in Processing Networks        | 3 |
| <b>Stochastic Processes</b>            |   |   |
| STATS 317                              | Stochastic Processes                                  | 3 |
| MATH 234                               | Large Deviations Theory                               | 3 |

- <sup>1</sup> The following courses are not offered this year but may be used by students who completed them in fulfillment of this requirement: STATS 318, STATS 345, STATS 362.

## Earth Science Statistics

Students are expected to take:

|           |                                  | Units |
|-----------|----------------------------------|-------|
| STATS 313 | Introduction to Graphical Models | 3     |
| STATS 317 | Stochastic Processes             | 3     |
| STATS 318 | Modern Markov Chains             | 3     |

In addition, students are expected to take three courses from the GS or Geophysics departments, such as GEOPHYS 210.

## Social and Behavioral Sciences

Students are expected to take three advanced courses from the department with an applied orientation such as:

|   |   | Units |
|---|---|-------|
| Courses can be chosen from the following list: <sup>1</sup> |   |       |
| STATS 261/262   | Intermediate Biostatistics: Analysis of Discrete Data | 3     |
| STATS 324   | Multivariate Analysis                                 | 2-3   |

- <sup>1</sup> The following courses are not offered this year but may be used by students who completed them in fulfillment of this requirement: STATS 343, 354.

In addition, students must complete at least three advanced quantitative courses from departments such as Anthropology, Economics, Political Science, Psychology, and Sociology, and the schools of Education, Business, or Medicine.

## Dissertation Reading Committee, Dissertation Proposal Meeting and University Oral Examinations

The dissertation reading committee consists of the student's adviser plus two faculty readers, all of whom are responsible for reading and approving the full dissertation.

The dissertation proposal meeting is intended to demonstrate students' depth in some areas of statistics, and to examine the general plan for their research. It also confirms that students have chosen a Ph.D. faculty adviser and have started to work with that adviser on a research topic. In the meeting, they will give a short presentation and discuss their ideas for completing a Ph.D. thesis, with a committee consisting of the dissertation reading committee plus a fourth member. The meeting must be successfully completed before the end of their third year. "Successful completion" means that the general research plan is sound and has a reasonable chance of success. If they do not successfully complete the meeting to the satisfaction of the committee, then the meeting must be repeated. Repeated failure can lead to a loss of financial support.

The oral examination/dissertation defense is scheduled when the student has finished their dissertation and is in the process of completing their final draft. The oral exam consists of a 40-minute presentation on the thesis topic, followed by a question period. The questions relate both to the student's presentation and also explore the student's familiarity with broader statistical topics related to the thesis research. The oral examination is normally completed within the last few months of the student's Ph.D. period. The examining committee usually consists of the dissertation proposal meeting committee and a fifth faculty member from outside the

department. Four out of five passing votes are required and no grades are given. Nearly all students can expect to pass this examination, although it is common for specific recommendations to be made regarding completion of the thesis.

For further information on University oral examinations and committees, see the Graduate Academic Policies and Procedures (GAP) Handbook, section 4.7 (<http://gap.stanford.edu/4-7.html>) or the "University Oral Examination (<http://exploreddegrees.stanford.edu/graduatedegrees/#doctoraltext>)" section of this bulletin.

## Doctoral and Research Advisers

From the student's arrival until the selection of a research adviser, the student's academic progress is monitored by the department Director of Graduate Studies. Each student should meet at least once a quarter with the Doctoral Adviser to discuss their academic plans and their progress towards choosing a thesis adviser.

## Financial Support

Students accepted to the Ph.D. program are offered financial support. All tuition expenses are paid and there is a fixed monthly stipend determined to be sufficient to pay living expenses. Financial support can be continued for five years, department resources permitting, for students in good standing. The resources for student financial support derive from funds made available for student teaching and research assistantships. Students receive both a teaching and research assignment each quarter which, together, do not exceed 20 hours. Students are encouraged to apply for outside scholarships, fellowships, and other forms of financial support.

## Ph.D. Minor in Statistics

Students must complete 30 total units for the Ph.D. minor. 20 units must be from Statistics courses numbered 300 and above and taken for letter grades. The remaining 10 units can be from Statistics courses numbered 200 and above, and may be taken for credit. The selection of courses must be approved by the Director of Graduate Studies. The Application for the Ph.D. Minor form must be approved by both the student's Ph.D. department and the Statistics department.

For further information about the Statistics Ph.D. degree program requirements, see the department web site (<https://statistics.stanford.edu/academics/doctoral-program>).

*Emeriti:* Theodore W. Anderson, Jerome H. Friedman, Ingram Olkin, Charles Stein, Paul Switzer

*Chair:* Guenther Walther (Aut), Emmanuel Candés (Win, Spr, Sum)

*Professors:* Emmanuel Candés, Sourav Chatterjee, Amir Dembo, Persi Diaconis, David L. Donoho, Bradley Efron, Trevor J. Hastie, Susan P. Holmes, Iain M. Johnstone, Tze L. Lai, Art Owen, Joseph P. Romano, David O. Siegmund, Jonathan Taylor, Robert J. Tibshirani, Guenther Walther, Wing H. Wong

*Associate Professor:* Andrea Montanari

*Assistant Professors:* John Duchi, Lester Mackey, Balakanapathy Rajaratnam

*Courtesy Professors:* John Ioannidis, Philip W. Lavori, Richard A. Olshen

*Courtesy Associate Professors:* Simon Jackman (on leave), David Rogosa, Chiara Sabatti, Hua Tang

*Courtesy Assistant Professors:* Mike Baiocchi, Percy Liang

*Consulting Professor:* John Chambers

*Stein Fellows:* Rajarshi Mukherjee, Rachel Wang, Lucy Xia

## Courses

### **STATS 42Q. Undergraduate Admissions to Selective Universities - a Statistical Perspective. 2 Units.**

The goal is the building of a statistical model, based on applicant data, for predicting admission to selective universities. The model will consider factors such as gender, ethnicity, legacy status, public-private schooling, test scores, effects of early action, and athletics. Common misconceptions and statistical pitfalls are investigated. The applicant data are not those associated with any specific university.

### **STATS 48N. Riding the Data Wave. 3 Units.**

Imagine collecting a bit of your saliva and sending it in to one of the personalized genomics company: for very little money you will get back information about hundreds of thousands of variable sites in your genome. Records of exposure to a variety of chemicals in the areas you have lived are only a few clicks away on the web; as are thousands of studies and informal reports on the effects of different diets, to which you can compare your own. What does this all mean for you? Never before in history humans have recorded so much information about themselves and the world that surrounds them. Nor has this data been so readily available to the lay person. Expression as "data deluge" are used to describe such wealth as well as the loss of proper bearings that it often generates. How to summarize all this information in a useful way? How to boil down millions of numbers to just a meaningful few? How to convey the gist of the story in a picture without misleading oversimplifications? To answer these questions we need to consider the use of the data, appreciate the diversity that they represent, and understand how people instinctively interpret numbers and pictures. During each week, we will consider a different data set to be summarized with a different goal. We will review analysis of similar problems carried out in the past and explore if and how the same tools can be useful today. We will pay attention to contemporary media (newspapers, blogs, etc.) to identify settings similar to the ones we are examining and critique the displays and summaries there documented. Taking an experimental approach, we will evaluate the effectiveness of different data summaries in conveying the desired information by testing them on subsets of the enrolled students.

### **STATS 50. Mathematics of Sports. 3 Units.**

The use of mathematics, statistics, and probability in the analysis of sports performance, sports records, and strategy. Topics include mathematical analysis of the physics of sports and the determinations of optimal strategies. New diagnostic statistics and strategies for each sport. Corequisite: STATS 60, 110 or 116. Same as: MCS 100

### **STATS 60. Introduction to Statistical Methods: Precalculus. 5 Units.**

Techniques for organizing data, computing, and interpreting measures of central tendency, variability, and association. Estimation, confidence intervals, tests of hypotheses, t-tests, correlation, and regression. Possible topics: analysis of variance and chi-square tests, computer statistical packages. Same as: PSYCH 10, STATS 160

### **STATS 90. Mathematics in the Real World. 3 Units.**

Introduction to non-calculus applications of mathematical ideas and principles in real-world problems. Topics include probability and counting, basic statistical concepts, geometric series. Applications include insurance, gambler's ruin, false positives in disease testing, present value of money, and mortgages. No knowledge of calculus required. Enrollment limited to students who do not have Stanford credit for a high school or college course in calculus or statistics. Same as: MATH 16

### **STATS 110. Statistical Methods in Engineering and the Physical Sciences. 4-5 Units.**

Introduction to statistics for engineers and physical scientists. Topics: descriptive statistics, probability, interval estimation, tests of hypotheses, nonparametric methods, linear regression, analysis of variance, elementary experimental design. Prerequisite: one year of calculus.

### **STATS 116. Theory of Probability. 3-5 Units.**

Probability spaces as models for phenomena with statistical regularity. Discrete spaces (binomial, hypergeometric, Poisson). Continuous spaces (normal, exponential) and densities. Random variables, expectation, independence, conditional probability. Introduction to the laws of large numbers and central limit theorem. Prerequisites: MATH 52 and familiarity with infinite series, or equivalent.

### **STATS 141. Biostatistics. 3-5 Units.**

Introductory statistical methods for biological data: describing data (numerical and graphical summaries); introduction to probability; and statistical inference (hypothesis tests and confidence intervals). Intermediate statistical methods: comparing groups (analysis of variance); analyzing associations (linear and logistic regression); and methods for categorical data (contingency tables and odds ratio). Course content integrated with statistical computing in R. Same as: BIO 141

### **STATS 155. Statistical Methods in Computational Genetics. 3 Units.**

The computational methods necessary for the construction and evaluation of sequence alignments and phylogenies built from molecular data and genetic data such as micro-arrays and data base searches. How to formulate biological problems in an algorithmic decomposed form, and building blocks common to many problems such as Markovian models, multivariate analyses. Some software covered in labs (Python, Biopython, XGobi, MrBayes, HMMER, Probe). Prerequisites: knowledge of probability equivalent to STATS 116, STATS 202 and one class in computing at the CS 106 level. Writing intensive course for undergraduates only. Instructor consent required. (WIM).

### **STATS 160. Introduction to Statistical Methods: Precalculus. 5 Units.**

Techniques for organizing data, computing, and interpreting measures of central tendency, variability, and association. Estimation, confidence intervals, tests of hypotheses, t-tests, correlation, and regression. Possible topics: analysis of variance and chi-square tests, computer statistical packages. Same as: PSYCH 10, STATS 60

### **STATS 167. Probability: Ten Great Ideas About Chance. 4 Units.**

Foundational approaches to thinking about chance in matters such as gambling, the law, and everyday affairs. Topics include: chance and decisions; the mathematics of chance; frequencies, symmetry, and chance; Bayes great idea; chance and psychology; misuses of chance; and harnessing chance. Emphasis is on the philosophical underpinnings and problems. Prerequisite: exposure to probability or a first course in statistics at the level of STATS 60 or 116. Same as: PHIL 166, PHIL 266, STATS 267

### **STATS 191. Introduction to Applied Statistics. 3-4 Units.**

Statistical tools for modern data analysis. Topics include regression and prediction, elements of the analysis of variance, bootstrap, and cross-validation. Emphasis is on conceptual rather than theoretical understanding. Applications to social/biological sciences. Student assignments/projects require use of the software package R. Recommended: 60, 110, or 141.

**STATS 195. Introduction to R. 1 Unit.**

This short course runs for the first four weeks of the quarter and is offered in fall and spring. It is recommended for students who want to use R in statistics, science, or engineering courses and for students who want to learn the basics of R programming. The goal of the short course is to familiarize students with R's tools for scientific computing. Lectures will be interactive with a focus on learning by example, and assignments will be application-driven. No prior programming experience is needed. Topics covered include basic data structures, File I/O, graphs, control structures, etc., and some useful packages in R.

Same as: CME 195

**STATS 196A. Multilevel Modeling Using R. 1 Unit.**

Multilevel data analysis examples using R. Topics include: two-level nested data, growth curve modeling, generalized linear models for counts and categorical data, nonlinear models, three-level analyses. For more information, see course website: <http://rogosateaching.com/stat196/>.

Same as: EDUC 401D

**STATS 198. Practical Training. 1 Unit.**

For students majoring in Mathematical and Computational Science only. Students obtain employment in a relevant industrial or research activity to enhance their professional experience.

**STATS 199. Independent Study. 1-15 Unit.**

For undergraduates.

**STATS 200. Introduction to Statistical Inference. 3 Units.**

Modern statistical concepts and procedures derived from a mathematical framework. Statistical inference, decision theory; point and interval estimation, tests of hypotheses; Neyman-Pearson theory. Bayesian analysis; maximum likelihood, large sample theory. Prerequisite: 116.

**STATS 201. Design and Analysis of Experiments. 3-5 Units.**

Theory and applications. Factors that affect response. Optimum levels of parameters. How to balance theory and practical design techniques. Prerequisites: basic statistics and probability theory.

**STATS 202. Data Mining and Analysis. 3 Units.**

Data mining is used to discover patterns and relationships in data. Emphasis is on large complex data sets such as those in very large databases or through web mining. Topics: decision trees, association rules, clustering, case based methods, and data visualization. Prereqs: Introductory courses in statistics or probability (e.g., Stats 60), linear algebra (e.g., Math 51), and computer programming (e.g., CS 105).

**STATS 203. Introduction to Regression Models and Analysis of Variance. 3 Units.**

Modeling and interpretation of observational and experimental data using linear and nonlinear regression methods. Model building and selection methods. Multivariable analysis. Fixed and random effects models. Experimental design. Pre- or corequisite: 200.

**STATS 204. Sampling. 3 Units.**

How best to take data and where to sample it. Examples include surveys and sampling from data warehouses. Emphasis is on methods for finite populations. Topics: simple random sampling, stratified sampling, cluster sampling, ratio and regression estimators, two stage sampling.

**STATS 205. Introduction to Nonparametric Statistics. 3 Units.**

Nonparametric analogs of the one- and two-sample *t*-tests and analysis of variance; the sign test, median test, Wilcoxon's tests, and the Kruskal-Wallis and Friedman tests, tests of independence. Nonparametric regression and nonparametric density estimation, modern nonparametric techniques, nonparametric confidence interval estimates.

**STATS 206. Applied Multivariate Analysis. 3 Units.**

Introduction to the statistical analysis of several quantitative measurements on each observational unit. Emphasis is on concepts, computer-intensive methods. Examples from economics, education, geology, psychology. Topics: multiple regression, multivariate analysis of variance, principal components, factor analysis, canonical correlations, multidimensional scaling, clustering. Pre- or corequisite: 200.

**STATS 207. Introduction to Time Series Analysis. 3 Units.**

Time series models used in economics and engineering. Trend fitting, autoregressive and moving average models and spectral analysis, Kalman filtering, and state-space models. Seasonality, transformations, and introduction to financial time series. Prerequisite: basic course in Statistics at the level of 200.

**STATS 208. Introduction to the Bootstrap. 3 Units.**

The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates. By substituting computation in place of mathematical formulas, it permits the statistical analysis of complicated estimators. Topics: nonparametric assessment of standard errors, biases, and confidence intervals; related resampling methods including the jackknife, cross-validation, and permutation tests. Theory and applications. Prerequisite: course in statistics or probability.

**STATS 209. Statistical Methods for Group Comparisons and Causal Inference. 3 Units.**

Critical examination of statistical methods in social science and life sciences applications, especially for cause and effect determinations. Topics: mediating and moderating variables, potential outcomes framework, encouragement designs, multilevel models, matching and propensity score methods, analysis of covariance, instrumental variables, compliance, path analysis and graphical models, group comparisons with longitudinal data. See <http://rogosateaching.com/stat209/>. Prerequisite: intermediate-level statistical methods.

Same as: EDUC 260A, HRP 239

**STATS 211. Meta-research: Appraising Research Findings, Bias, and Meta-analysis. 3 Units.**

Open to graduate, medical, and undergraduate students. Appraisal of the quality and credibility of research findings; evaluation of sources of bias. Meta-analysis as a quantitative (statistical) method for combining results of independent studies. Examples from medicine, epidemiology, genomics, ecology, social/behavioral sciences, education. Collaborative analyses. Project involving generation of a meta-research project or reworking and evaluation of an existing published meta-analysis. Prerequisite: knowledge of basic statistics.

Same as: CHPR 206, HRP 206, MED 206

**STATS 212. Applied Statistics with SAS. 3 Units.**

Data analysis and implementation of statistical tools in SAS. Topics: reading in and describing data, categorical data, dates and longitudinal data, correlation and regression, nonparametric comparisons, ANOVA, multiple regression, multivariate data analysis, using arrays and macros in SAS. Prerequisite: statistical techniques at the level of STATS 191 or 203; knowledge of SAS not required.

**STATS 213. Introduction to Graphical Models. 3 Units.**

Multivariate Normal Distribution and Inference, Wishart distributions, graph theory, probabilistic Markov models, pairwise and global Markov property, decomposable graph, Markov equivalence, MLE for DAG models and undirected graphical models, Bayesian inference for DAG models and undirected graphical models. Prerequisites: STATS 217, STATS 200 (preferably STATS 300A), MATH 104 or equivalent class in linear algebra. Same as: STATS 313

**STATS 215. Statistical Models in Biology. 3 Units.**

Poisson and renewal processes, Markov chains in discrete and continuous time, branching processes, diffusion. Applications to models of nucleotide evolution, recombination, the Wright-Fisher process, coalescence, genetic mapping, sequence analysis. Theoretical material approximately the same as in STATS 217, but emphasis is on examples drawn from applications in biology, especially genetics. Prerequisite: 116 or equivalent.



**STATS 216. Introduction to Statistical Learning. 3 Units.**

Overview of supervised learning, with a focus on regression and classification methods. Syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis; cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines; Some unsupervised learning: principal components and clustering (k-means and hierarchical). Computing is done in R, through tutorial sessions and homework assignments. This math-light course is offered via video segments (MOOC style), and in-class problem solving sessions. Prereqs: Introductory courses in statistics or probability (e.g., Stats 60), linear algebra (e.g., Math 51), and computer programming (e.g., CS 105).

**STATS 216V. Introduction to Statistical Learning. 3 Units.**

Overview of supervised learning, with a focus on regression and classification methods. Syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis; cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines; Some unsupervised learning: principal components and clustering (k-means and hierarchical). Computing is done in R, through tutorial sessions and homework assignments. This math-light course is offered remotely only via video segments (MOOC style). TAs will host remote weekly office hours using an online platform such as Google Hangout or BlueJeans. There are four homework assignments, a midterm, and final exam. Prereqs: Introductory courses in statistics or probability (e.g., Stats 60), linear algebra (e.g., Math 51), and computer programming (e.g., CS 105).

**STATS 217. Introduction to Stochastic Processes. 2-3 Units.**

Discrete and continuous time Markov chains, poisson processes, random walks, branching processes, first passage times, recurrence and transience, stationary distributions. Non-Statistics masters students may want to consider taking STATS 215 instead. Prerequisite: STATS 116 or consent of instructor.

**STATS 218. Introduction to Stochastic Processes. 3 Units.**

Renewal theory, Brownian motion, Gaussian processes, second order processes, martingales.

**STATS 219. Stochastic Processes. 3 Units.**

Introduction to measure theory,  $L_p$  spaces and Hilbert spaces. Random variables, expectation, conditional expectation, conditional distribution. Uniform integrability, almost sure and  $L_p$  convergence. Stochastic processes: definition, stationarity, sample path continuity. Examples: random walk, Markov chains, Gaussian processes, Poisson processes, Martingales. Construction and basic properties of Brownian motion. Prerequisite: STATS 116 or MATH 151 or equivalent. Recommended: MATH 115 or equivalent.

Same as: MATH 136

**STATS 221. Introduction to Mathematical Finance. 3-4 Units.**

Interest rate and discounted value. Financial derivatives, hedging, and risk management. Stochastic models of financial markets, introduction to Ito calculus and stochastic differential equations. Black-Scholes pricing of European options. Optimal stopping and American options. Prerequisites: MATH 53, STATS 116, or equivalents.

**STATS 222. Statistical Methods for Longitudinal Research. 2-3 Units.**

Research designs and statistical procedures for time-ordered (repeated-measures) data. The analysis of longitudinal panel data is central to empirical research on learning, development, aging, and the effects of interventions. Topics include: measurement of change, growth curve models, analysis of durations including survival analysis, experimental and non-experimental group comparisons, reciprocal effects, stability. See <http://rogosateaching.com/stat222/>. Prerequisite: intermediate statistical methods.

Same as: EDUC 351A

**STATS 229. Machine Learning. 3-4 Units.**

Topics: statistical pattern recognition, linear and non-linear regression, non-parametric methods, exponential family, GLMs, support vector machines, kernel methods, model/feature selection, learning theory, VC dimension, clustering, density estimation, EM, dimensionality reduction, ICA, PCA, reinforcement learning and adaptive control, Markov decision processes, approximate dynamic programming, and policy search. Prerequisites: linear algebra, and basic probability and statistics.

Same as: CS 229

**STATS 231. Statistical Learning Theory. 3 Units.**

(Same as STATS 231) How do we formalize what it means for an algorithm to learn from data? This course focuses on developing mathematical tools for answering this question. We will present various common learning algorithms and prove theoretical guarantees about them. Topics include online learning, kernel methods, generalization bounds (uniform convergence), and spectral methods. Prerequisites: A solid background in linear algebra and probability theory, statistics and machine learning (STATS 315A or CS 229). Convex optimization (EE 364a) is helpful but not required.

Same as: CS 229T

**STATS 237. Theory of Investment Portfolios and Derivative Securities. 3 Units.**

Asset returns and their volatilities. Markowitz; portfolio theory, capital asset pricing model, multifactor pricing models. Measures of market risk. Financial derivatives and hedging. Black-Scholes; Scholes pricing of European options. Valuation of American options. Implied volatility and the Greeks. Prerequisite: STATS 116 or equivalent.

**STATS 238. The Future of Finance. 2 Units.**

If you are interested in a career in finance or that touches finance (computational science, economics, public policy, legal, regulatory, corporate, other), this course will give you a useful perspective. We will take on hot topics in the current landscape of the global markets as the world continues to evolve from the financial crisis. We will discuss the sweeping change underway at the policy level by regulators and legislators around the world and how this is changing business models for existing players and attracting new players to finance. The course will include guest-lecturer perspectives on where the greatest opportunities exist for students entering or touching the world of finance today including new and disruptive players in fin tech, crowd financing, block chain, robo advising, algorithmic trading, big data and other areas. New challenges such as cyber and financial warfare threats also will be addressed. While derivatives and other quantitative concepts will be handled in a non-technical way, some knowledge of finance and the capital markets is presumed. Elements used in grading: Class Participation, Attendance, Final Paper. Consent Application: To apply for this course, students must complete and email to the instructors the Consent Application Form, which will be made available on the Public Policy Program's website prior to the beginning of Winter Quarter. See Consent Application Form for submission deadline. (Cross-listed as ECON252/152, PUBLPOL364, STATS238, LAW 564.).

Same as: ECON 152, ECON 252, PUBLPOL 364

**STATS 239. Mathematical and Computational Finance Seminar. 1 Unit.**

Same as: CME 242

**STATS 239A. Workshop in Quantitative Finance. 1 Unit.**

Topics of current interest.

**STATS 239B. Workshop in Quantitative Finance. 1 Unit.**

Topics of current interest. May be repeated for credit.

Same as: CME 239B

**STATS 240. Statistical Methods in Finance. 3-4 Units.**

(SCPD students register for 240P.) Regression analysis and applications to investment models. Principal components and multivariate analysis. Likelihood inference and Bayesian methods. Financial time series. Estimation and modeling of volatilities. Statistical methods for portfolio management. Prerequisite: STATS 200 or equivalent.

**STATS 240P. Statistical Methods in Finance. 3 Units.**

For SCPD students; see 240.

**STATS 241. Data-driven Financial and Risk Econometrics. 3-4 Units.**

(SCPD students register for 241P) Substantive and empirical modeling approaches in options, interest rate, and credit markets. Nonlinear least squares, logistic regression and generalized linear models. Nonparametric regression and model selection. Multivariate time series modeling and forecasting. Vector autoregressive models and cointegration. Risk measures, models and analytics. Prerequisite or corequisite: STATS 240 or equivalent.

**STATS 241P. Data-driven Financial and Risk Econometrics. 3 Units.**

For SCPD students; see STATS241.

**STATS 242. Algorithmic Trading and Quantitative Strategies. 3 Units.**

An introduction to financial trading strategies based on methods of statistical arbitrage that can be automated. Methodologies related to high frequency data and stylized facts on asset returns; models of order book dynamics and order placement, dynamic trade planning with feedback; momentum strategies, pairs trading. Emphasis on developing and implementing models that reflect the market and behavioral patterns. Prerequisite: STATS 240 or equivalent.

**STATS 243. Financial Models and Statistical Methods in Active Risk Management. 2-4 Units.**

(SCPD students register for 243P.) Market risk and credit risk, credit markets. Back testing, stress testing and Monte Carlo methods. Logistic regression, generalized linear models and generalized mixed models. Loan prepayment and default as competing risks. Survival and hazard functions, correlated default intensities, frailty and contagion. Risk surveillance, early warning and adaptive control methodologies. Banking and bank regulation, asset and liability management. Prerequisite: STATS 240 or equivalent. Same as: CME 243

**STATS 243P. Financial Models and Statistical Methods in Risk Management. 3 Units.**

For SCPD students; see STATS243.

**STATS 244. Quantitative Trading: Algorithms, Data, and Optimization. 2-4 Units.**

Statistical trading rules and performances evaluation. Active portfolio management and dynamic investment strategies. Data analytics and models of transactions data. Limit order book dynamics in electronic exchanges. Algorithmic trading, informatics, and optimal execution. Market making and inventory control. Risk management and regulatory issues. Prerequisites: STATS 240 or equivalent.

**STATS 245. Data, Models, and Decision Analytics. 3 Units.**

(SCPD students register for 245P) Statistical models and decision theory. Online A/B testing, comparative effective studies of medical treatments. Introduction to recommender systems in online services, personalized medicine and marketing. Prerequisite or corequisite: STATS 202, or CS 229, or CME 250, or equivalent.

**STATS 245P. Data, Models, and Decision Analytics. 3 Units.**

For SCPD students; see STATS245.

**STATS 250. Mathematical Finance. 3 Units.**

Stochastic models of financial markets. Forward and futures contracts. European options and equivalent martingale measures. Hedging strategies and management of risk. Term structure models and interest rate derivatives. Optimal stopping and American options. Corequisites: MATH 236 and 227 or equivalent. Same as: MATH 238

**STATS 253. Analysis of Spatial and Temporal Data. 3 Units.**

A unified treatment of methods for spatial data, time series, and other correlated data from the perspective of regression with correlated errors. Two main paradigms for dealing with autocorrelation: covariance modeling (kriging) and autoregressive processes. Bayesian methods. Prerequisites: applied linear algebra (MATH 103 or equivalent), statistical estimation (STATS 200 or CS 229), and linear regression (STATS 203 or equivalent).

**STATS 260A. Workshop in Biostatistics. 1-2 Unit.**

Applications of statistical techniques to current problems in medical science. To receive credit for one or two units, a student must attend every workshop. To receive two units, in addition to attending every workshop, the student is required to write an acceptable one page summary of two of the workshops, with choices made by the student.

Same as: HRP 260A

**STATS 260B. Workshop in Biostatistics. 1-2 Unit.**

Applications of statistical techniques to current problems in medical science. To receive credit for one or two units, a student must attend every workshop. To receive two units, in addition to attending every workshop, the student is required to write an acceptable one page summary of two of the workshops, with choices made by the student.

Same as: HRP 260B

**STATS 260C. Workshop in Biostatistics. 1-2 Unit.**

Applications of statistical techniques to current problems in medical science. To receive credit for one or two units, a student must attend every workshop. To receive two units, in addition to attending every workshop, the student is required to write an acceptable one page summary of two of the workshops, with choices made by the student.

Same as: HRP 260C

**STATS 261. Intermediate Biostatistics: Analysis of Discrete Data. 3 Units.**

Methods for analyzing data from case-control and cross-sectional studies: the 2x2 table, chi-square test, Fisher's exact test, odds ratios, Mantel-Haenzel methods, stratification, tests for matched data, logistic regression, conditional logistic regression. Emphasis is on data analysis in SAS. Special topics: cross-fold validation and bootstrap inference.

Same as: BIOMEDIN 233, HRP 261

**STATS 262. Intermediate Biostatistics: Regression, Prediction, Survival Analysis. 3 Units.**

Methods for analyzing longitudinal data. Topics include Kaplan-Meier methods, Cox regression, hazard ratios, time-dependent variables, longitudinal data structures, profile plots, missing data, modeling change, MANOVA, repeated-measures ANOVA, GEE, and mixed models. Emphasis is on practical applications. Prerequisites: basic ANOVA and linear regression.

Same as: HRP 262

**STATS 263. Design of Experiments. 3 Units.**

Experiments vs observation. Confounding. Randomization. ANOVA. Blocking. Latin squares. Factorials and fractional factorials. Split plot. Response surfaces. Mixture designs. Optimal design. Central composite. Box-Behnken. Taguchi methods. Computer experiments and space filling designs. Prerequisites: probability at STATS 116 level or higher, and at least one course in linear models.

Same as: STATS 363

**STATS 266. Advanced Statistical Methods for Observational Studies. 2-3 Units.**

Design principles and statistical methods for observational studies, particularly for cause and effect determinations. Topics include: matching methods, sensitivity analysis, instrumental variables, graphical models, marginal structural models. 3 unit registration requires a small project and presentation. Computing is in R. Pre-requisites: HRP 261 and 262 or STAT 209 (HRP 239), or equivalent. See <http://rogosateaching.com/somgen290/>. Same as: CHPR 290, EDUC 260B

**STATS 267. Probability: Ten Great Ideas About Chance. 4 Units.**

Foundational approaches to thinking about chance in matters such as gambling, the law, and everyday affairs. Topics include: chance and decisions; the mathematics of chance; frequencies, symmetry, and chance; Bayes great idea; chance and psychology; misuses of chance; and harnessing chance. Emphasis is on the philosophical underpinnings and problems. Prerequisite: exposure to probability or a first course in statistics at the level of STATS 60 or 116.

Same as: PHIL 166, PHIL 266, STATS 167

**STATS 270. A Course in Bayesian Statistics. 3 Units.**

Advanced-level Bayesian statistics. Topics: Discussion of the mathematical and theoretical foundation for Bayesian inferential procedures. Examination of the construction of priors and the asymptotic properties of likelihoods and posterior densities. Discussion including but not limited to the case of finite dimensional parameter space. Prerequisite: familiarity with standard probability and multivariate distribution theory.

Same as: STATS 370

**STATS 290. Paradigms for Computing with Data. 3 Units.**

Advanced programming and computing techniques to support projects in data analysis and related research. For Statistics graduate students and others whose research involves data analysis and development of associated computational software. Prerequisites: Programming experience including familiarity with R; computing at least at the level of CS 106; statistics at the level of STATS 110 or 141.

**STATS 298. Industrial Research for Statisticians. 1 Unit.**

Masters-level research as in 299, but with the approval and supervision of a faculty adviser, it must be conducted for an off-campus employer. Students must submit a written final report upon completion of the internship in order to receive credit. Repeatable for credit. Prerequisite: enrollment in Statistics M.S. program.

**STATS 299. Independent Study. 1-10 Unit.**

For Statistics M.S. students only. Reading or research program under the supervision of a Statistics faculty member. May be repeated for credit.

**STATS 300. Advanced Topics in Statistics. 2-3 Units.**

Topic: Exploratory Multivariate Data Analysis. Describing and visualizing data with principal component analysis (PCA) for continuous data, correspondence analysis (CA) for contingency tables, multiple correspondence analysis (MCA) for categorical data, factorial analysis for mixed data (FAMD) for both continuous and categorical data, and multiple factor analysis (MFA) for data structured into groups of variables. Studying and visualization of the correlation between groups of variables with the RV coefficient. Performing PCA with missing values, matrix completion of continuous and categorical data with principal components. Examples from sensory analysis, public health, genetics. All the analysis will be performed with R.

**STATS 300A. Theory of Statistics. 2-3 Units.**

Finite sample optimality of statistical procedures; Decision theory: loss, risk, admissibility; Principles of data reduction: sufficiency, ancillarity, completeness; Statistical models: exponential families, group families, nonparametric families; Point estimation: optimal unbiased and equivariant estimation, Bayes estimation, minimax estimation; Hypothesis testing and confidence intervals: uniformly most powerful tests, uniformly most accurate confidence intervals, optimal unbiased and invariant tests. Prerequisites: Real analysis, introductory probability (at the level of STATS 116), and introductory statistics.

**STATS 300B. Theory of Statistics. 2-4 Units.**

Elementary decision theory; loss and risk functions, Bayes estimation; UMVU estimator, minimax estimators, shrinkage estimators. Hypothesis testing and confidence intervals: Neyman-Pearson theory; UMP tests and uniformly most accurate confidence intervals; use of unbiasedness and invariance to eliminate nuisance parameters. Large sample theory: basic convergence concepts; robustness; efficiency; contiguity, locally asymptotically normal experiments; convolution theorem; asymptotically UMP and maximin tests. Asymptotic theory of likelihood ratio and score tests. Rank permutation and randomization tests; jackknife, bootstrap, subsampling and other resampling methods. Further topics: sequential analysis, optimal experimental design, empirical processes with applications to statistics, Edgeworth expansions, density estimation, time series.

**STATS 300C. Theory of Statistics. 2-4 Units.**

Decision theory formulation of statistical problems. Minimax, admissible procedures. Complete class theorems ("all" minimax or admissible procedures are "Bayes"), Bayes procedures, conjugate priors, hierarchical models. Bayesian non parametrics: diaichlet, tail free, polya trees, bayesian sieves. Inconsistency of bayes rules.

**STATS 302. Qualifying Exams Workshop. 3 Units.**

Prepares Statistics Ph.D. students for the qualifying exams by reviewing relevantnncourse topics and problem solving strategies.

**STATS 303. PhD First Year Student Workshop. 1 Unit.**

For Statistics First Year PhD students only. Discussion of relevant topics in first year student courses, consultation with PhD advisor.

**STATS 305. Introduction to Statistical Modeling. 3 Units.**

Review of univariate regression. Multiple regression. Geometry, subspaces, orthogonality, projections, normal equations, rank deficiency, estimable functions and Gauss-Markov theorem. Computation via QR decomposition, Gramm-Schmidt orthogonalization and the SVD. Interpreting coefficients, collinearity, graphical displays. Fits and the Hat matrix, leverage & influence, diagnostics, weighted least squares and resistance. Model selection, Cp/Aic and crossvalidation, stepwise, lasso. Basis expansions, splines. Multivariate normal distribution theory. ANOVA: Sources of measurements, fixed and random effects, randomization. Emphasis on problem sets involving substantive computations with data sets. Prerequisites: consent of instructor, 116, 200, applied statistics course, CS 106A, MATH 114.

**STATS 306A. Methods for Applied Statistics. 3 Units.**

Regression modeling extended to categorical data. Logistic regression. Loglinear models. Generalized linear models. Discriminant analysis. Categorical data models from information retrieval and Internet modeling. Prerequisite: 305 or equivalent.

**STATS 306B. Methods for Applied Statistics: Unsupervised Learning. 2-3 Units.**

Unsupervised learning techniques in statistics, machine learning, and data mining.

**STATS 310A. Theory of Probability. 2-4 Units.**

Mathematical tools: sigma algebras, measure theory, connections between coin tossing and Lebesgue measure, basic convergence theorems. Probability: independence, Borel-Cantelli lemmas, almost sure and  $L_p$  convergence, weak and strong laws of large numbers. Large deviations. Weak convergence; central limit theorems; Poisson convergence; Stein's method. Prerequisites: 116, MATH 171. Same as: MATH 230A

**STATS 310B. Theory of Probability. 2-3 Units.**

Conditional expectations, discrete time martingales, stopping times, uniform integrability, applications to 0-1 laws, Radon-Nikodym Theorem, ruin problems, etc. Other topics as time allows selected from (i) local limit theorems, (ii) renewal theory, (iii) discrete time Markov chains, (iv) random walk theory, (v) ergodic theory. Prerequisite: 310A or MATH 230A. Same as: MATH 230B

**STATS 310C. Theory of Probability. 2-4 Units.**

Continuous time stochastic processes: martingales, Brownian motion, stationary independent increments, Markov jump processes and Gaussian processes. Invariance principle, random walks, LIL and functional CLT. Markov and strong Markov property. Infinitely divisible laws. Some ergodic theory. Prerequisite: 310B or MATH 230B. Same as: MATH 230C

**STATS 311. Information Theory and Statistics. 3 Units.**

Information theoretic techniques in probability and statistics. Fano, Assouad, and Le Cam methods for optimality guarantees in estimation. Large deviations and concentration inequalities (Sanov's theorem, hypothesis testing, then entropy method, concentration of measure). Approximation of (Bayes) optimal procedures, surrogate risks,  $f$ -divergences. Penalized estimators and minimum description length. Online game playing, gambling, no-regret learning. Prerequisites: EE 376A (or equivalent) or STATS 300A. Same as: EE 377

**STATS 312. Statistical Methods in Neuroscience. 3 Units.**

The goal is to discuss statistical methods for neuroscience in their natural habitat: the research questions, measurement technologies and experiment designs used in modern neuroscience. We will emphasize both the choice and quality of the methods, as well as the reporting, interpretation and visualization of results. Likely topics include preprocessing and signal extraction for single-neuron and neuroimaging technologies, statistical models for single response, encoding and decoding models, multiple-responses and parametric maps, and testing. Participation includes analyzing methods and real data, discussing papers in class, and a final project. Requirements: we will assume familiarity with linear models, likelihoods etc. Students who have not taken graduate level statistics courses are required to contact the instructor. Background in neuroscience is not assumed.

**STATS 313. Introduction to Graphical Models. 3 Units.**

Multivariate Normal Distribution and Inference, Wishart distributions, graph theory, probabilistic Markov models, pairwise and global Markov property, decomposable graph, Markov equivalence, MLE for DAG models and undirected graphical models, Bayesian inference for DAG models and undirected graphical models. Prerequisites: STATS 217, STATS 200 (preferably STATS 300A), MATH 104 or equivalent class in linear algebra. Same as: STATS 213

**STATS 314A. Advanced Statistical Theory. 3 Units.**

Covers a range of topics, including: empirical processes, asymptotic efficiency, uniform convergence of measures, contiguity, resampling methods, Edgeworth expansions.

**STATS 314B. Topics in Minimax Inference of Nonparametric Functionals. 3 Units.**

Topics in the estimation of various functionals of underlying distribution for nonparametric problems. Development of ideas of higher order influence functions that extend the theory of classical first order semiparametric theory. Topics on adaptive estimation and adaptive confidence sets construction. Understanding results from wavelet theory and higher order  $U$ -statistics.

**STATS 315A. Modern Applied Statistics: Learning. 2-3 Units.**

Overview of supervised learning. Linear regression and related methods. Model selection, least angle regression and the lasso, stepwise methods. Classification. Linear discriminant analysis, logistic regression, and support vector machines (SVMs). Basis expansions, splines and regularization. Kernel methods. Generalized additive models. Kernel smoothing. Gaussian mixtures and the EM algorithm. Model assessment and selection: crossvalidation and the bootstrap. Pathwise coordinate descent. Sparse graphical models. Prerequisites: STATS 305, 306A,B or consent of instructor.

**STATS 315B. Modern Applied Statistics: Data Mining. 2-3 Units.**

Two-part sequence. New techniques for predictive and descriptive learning using ideas that bridge gaps among statistics, computer science, and artificial intelligence. Emphasis is on statistical aspects of their application and integration with more standard statistical methodology. Predictive learning refers to estimating models from data with the goal of predicting future outcomes, in particular, regression and classification models. Descriptive learning is used to discover general patterns and relationships in data without a predictive goal, viewed from a statistical perspective as computer automated exploratory analysis of large complex data sets.

**STATS 316. Stochastic Processes on Graphs. 1-3 Unit.**

Local weak convergence, Gibbs measures on trees, cavity method, and replica symmetry breaking. Examples include random  $k$ -satisfiability, the assignment problem, spin glasses, and neural networks. Prerequisite: 310A or equivalent.

**STATS 317. Stochastic Processes. 3 Units.**

Semimartingales, stochastic integration, Ito's formula, Girsanov's theorem. Gaussian and related processes. Stationary/isotropic processes. Integral geometry and geometric probability. Maxima of random fields and applications to spatial statistics and imaging.

**STATS 318. Modern Markov Chains. 3 Units.**

Tools for understanding Markov chains as they arise in applications. Random walk on graphs, reversible Markov chains, Metropolis algorithm, Gibbs sampler, hybrid Monte Carlo, auxiliary variables, hit and run, Swedson-Wong algorithms, geometric theory, Poincare-Nash-Cheeger-Log-Sobolov inequalities. Comparison techniques, coupling, stationary times, Harris recurrence, central limit theorems, and large deviations.

**STATS 319. Literature of Statistics. 1-3 Unit.**

Literature study of topics in statistics and probability culminating in oral and written reports. May be repeated for credit.

**STATS 320. Heterogeneous Data with Kernels. 3 Units.**

Mathematical and computational methods necessary to understanding analysis of heterogeneous data using generalized inner products and Kernels. For areas that need to integrate data from various sources, biology, environmental and chemical engineering, molecular biology, bioinformatics. Topics: Distances, inner products and duality. Multivariate projections. Complex heterogeneous data structures (networks, trees, categorical as well as multivariate continuous data). Canonical correlation analysis, canonical correspondence analysis. Kernel methods in Statistics. Representer theorem. Kernels on graphs. Kernel versions of standard statistical procedures. Data cubes and tensor methods.

**STATS 321. Modern Applied Statistics: Transposable Data. 2-3 Units.**

Topics: clustering, biclustering, and spectral clustering. Data analysis using the singular value decomposition, nonnegative decomposition, and generalizations. Plaid model, aspect model, and additive clustering. Correspondence analysis, Rasch model, and independent component analysis. Page rank, hubs, and authorities. Probabilistic latent semantic indexing. Recommender systems. Applications to genomics and information retrieval. Prerequisites: 315A,B, 305/306A,B, or consent of instructor.

**STATS 322. Function Estimation in White Noise. 2-3 Units.**

Gaussian white noise model sequence space form. Hyperrectangles, quadratic convexity, and Pinsker's theorem. Minimax estimation on  $L_p$  balls and Besov spaces. Role of wavelets and unconditional bases. Linear and threshold estimators. Oracle inequalities. Optimal recovery and universal thresholding. Stein's unbiased risk estimator and threshold choice. Complexity penalized model selection. Connecting fast wavelet algorithms and theory. Beyond orthogonal bases.

**STATS 324. Multivariate Analysis. 2-3 Units.**

Classic multivariate statistics: properties of the multivariate normal distribution, determinants, volumes, projections, matrix square roots, the singular value decomposition; Wishart distributions, Hotelling's T-square; principal components, canonical correlations, Fisher's discriminant, the Cauchy projection formula.

**STATS 325. Multivariate Analysis and Random Matrices in Statistics. 2-3 Units.**

Topics on Multivariate Analysis and Random Matrices in Statistics (full description TBA).

**STATS 329. Large-Scale Simultaneous Inference. 1-3 Unit.**

Estimation, testing, and prediction for microarray-like data. Modern scientific technologies, typified by microarrays and imaging devices, produce inference problems with thousands of parallel cases to consider simultaneously. Topics: empirical Bayes techniques, James-Stein estimation, large-scale simultaneous testing, false discovery rates, local fdr, proper choice of null hypothesis (theoretical, permutation, empirical nulls), power, effects of correlation on tests and estimation accuracy, prediction methods, related sets of cases ("enrichment"), effect size estimation. Theory and methods illustrated on a variety of large-scale data sets.

**STATS 330. An Introduction to Compressed Sensing. 3 Units.**

Compressed sensing is a new data acquisition theory asserting that one can design nonadaptive sampling techniques that condense the information in a compressible signal into a small amount of data. This revelation may change the way engineers think about signal acquisition. Course covers fundamental theoretical ideas, numerical methods in large-scale convex optimization, hardware implementations, connections with statistical estimation in high dimensions, and extensions such as recovery of data matrices from few entries (famous Netflix Prize). Same as: CME 362

**STATS 331. Survival Analysis. 2 Units.**

The course introduces basic concepts, theoretical basis and statistical methods associated with survival data. Topics include censoring, Kaplan-Meier estimation, logrank test, proportional hazards regression, accelerated failure time model, multivariate failure time analysis and competing risks. The traditional counting process/martingale methods as well as modern empirical process methods will be covered. Prerequisite: Understanding of basic probability theory and statistical inference methods.

**STATS 333. Modern Spectral Analysis. 3 Units.**

Traditional spectral analysis encompassed Fourier methods and their elaborations, under the assumption of a simple superposition of sinusoids, independent of time. This enables development of efficient and effective computational schemes, such as the FFT. Since many systems change in time, it becomes of interest to generalize classical spectral analysis to the time-varying setting. In addition, classical methods suffer from resolution limits which we hope to surpass. In this topics course, we follow two threads. On the one hand, we consider the estimation of instantaneous frequencies and decomposition of source signals, which may be time-varying. The thread begins with the empirical mode decomposition (EMD) for non-stationary signal decomposition into intrinsic mode functions (IMF), introduced by N. Huang et al [1], together with its machinery of the sifting process and computation of the Hilbert spectrum, resulting in the so-called adaptive harmonic model (AHM). Next, this thread considers the wavelet synchrosqueezing transform (WSST) proposed by Daubechies et al [2], which attempts to estimate instantaneous frequencies (IF), via the frequency re-assignment (FRA) rule, that facilitates non-stationary signal decomposition. In reference [3], a real-time method is proposed for computing the FRA rule; and in reference [4], the exact number of AHM components is determined with more precise estimation of the IF, for more accurate extraction of the signal components and polynomial-like trend. Another thread, recent developments in optimization have been applied to obtain time-varying spectra or very high-resolution spectra; in particular, references [5]-[8] give examples of recent results where convex estimation is applied to obtain new and more highly resolved spectral estimates, some with time-varying structure.

**STATS 338. Topics in Biostatistics. 3 Units.**

Data monitoring and interim analysis of clinical trials. Design of Phase I, II, III trials. Survival analysis. Longitudinal data analysis.

**STATS 341. Applied Multivariate Statistics. 3 Units.**

Theory, computational aspects, and practice of a variety of important multivariate statistical tools for data analysis. Topics include classical multivariate Gaussian and undirected graphical models, graphical displays. PCA, SVD and generalizations including canonical correlation analysis, linear discriminant analysis, correspondence analysis, with focus on recent variants. Factor analysis and independent component analysis. Multidimensional scaling and its variants (e.g. Isomap, spectral clustering). Students are expected to program in R. Prerequisite: STATS 305 or equivalent.

**STATS 344. Introduction to Statistical Genetics. 3 Units.**

Statistical methods for analyzing human genetics studies of Mendelian disorders and common complex traits. Probable topics include: principles of population genetics; epidemiologic designs; familial aggregation; segregation analysis; linkage analysis; linkage-disequilibrium-based association mapping approaches; and genome-wide analysis based on high-throughput genotyping platforms. Prerequisite: STATS 116 or equivalent or consent of instructor.

Same as: GENE 244

**STATS 345. Statistical and Machine Learning Methods for Genomics. 3 Units.**

Introduction to statistical and computational methods for genomics. Sample topics include: expectation maximization, hidden Markov model, Markov chain Monte Carlo, ensemble learning, probabilistic graphical models, kernel methods and other modern machine learning paradigms. Rationales and techniques illustrated with existing implementations used in population genetics, disease association, and functional regulatory genomics studies. Instruction includes lectures and discussion of readings from primary literature. Homework and projects require implementing some of the algorithms and using existing toolkits for analysis of genomic datasets. Same as: BIO 268, BIOMEDIN 245, CS 373, GENE 245

**STATS 350. Topics in Probability Theory: Probabilistic Concepts in Statistical Physics and Information Theory. 1-3 Unit.**

Concentration of measure techniques. Mean field models for disordered systems: infinite size limit, computing the free energy, ultrametricity, dynamics. Interpolation techniques and infinite size limit in information theory and coding. May be repeated once for credit. Prerequisite: 310A or equivalent.

**STATS 351. Random Walks, Networks and Environment. 3 Units.**

Selected material about probability on trees and networks, random walk in random and non-random environments, percolation and related interacting particle systems. Prerequisite: Exposure to measure theoretic probability and to stochastic processes.

**STATS 351A. An Introduction to Random Matrix Theory. 3 Units.**

Patterns in the eigenvalue distribution of typical large matrices, which also show up in physics (energy distribution in scattering experiments), combinatorics (length of longest increasing subsequence), first passage percolation and number theory (zeros of the zeta function). Classical compact ensembles (random orthogonal matrices). The tools of determinantal point processes.

Same as: MATH 231A

**STATS 355. Observational Studies. 2-3 Units.**

This course will cover statistical methods for the design and analysis of observational studies. Topics for the course will include the potential outcomes framework for causal inference; randomized experiments; methods for controlling for observed confounders in observational studies; sensitivity analysis for hidden bias; instrumental variables; tests of hidden bias; coherence; and design of observational studies.

Same as: HRP 255

**STATS 360. Advanced Statistical Methods for Earth System Analysis. 3 Units.**

Introduction for graduate students to important issues in data analysis relevant to earth system studies. Emphasis on methodology, concepts and implementation (in R), rather than formal proofs. Likely topics include the bootstrap, non-parametric methods, regression in the presence of spatial and temporal correlation, extreme value analysis, time-series analysis, high-dimensional regressions and change-point models. Topics subject to change each year. Prerequisites: STATS 110 or equivalent.

Same as: ESS 260

**STATS 362. Topic: Monte Carlo. 3 Units.**

Random numbers and vectors: inversion, acceptance-rejection, copulas. Variance reduction: antithetics, stratification, control variates, importance sampling. MCMC: Markov chains, detailed balance, Metropolis-Hastings, random walk Metropolis, independence sampler, Gibbs sampling, slice sampler, hybrids of Gibbs and Metropolis, tempering. Sequential Monte Carlo. Quasi-Monte Carlo. Randomized quasi-Monte Carlo. Examples, problems and motivation from Bayesian statistics, machine learning, computational finance and graphics. May be repeat for credit.

**STATS 363. Design of Experiments. 3 Units.**

Experiments vs observation. Confounding. Randomization. ANOVA. Blocking. Latin squares. Factorials and fractional factorials. Split plot. Response surfaces. Mixture designs. Optimal design. Central composite. Box-Behnken. Taguchi methods. Computer experiments and space filling designs. Prerequisites: probability at STATS 116 level or higher, and at least one course in linear models.

Same as: STATS 263

**STATS 366. Modern Statistics for Modern Biology. 3 Units.**

Application based course in nonparametric statistics. Modern toolbox of visualization and statistical methods for the analysis of data, examples drawn from immunology, microbiology, cancer research and ecology. Methods covered include multivariate methods (PCA and extensions), sparse representations (trees, networks, contingency tables) as well as nonparametric testing (Bootstrap, permutation and Monte Carlo methods). Hands on, use R and cover many Bioconductor packages. Prerequisite: Minimal familiarity with computers. Instructor consent. Location: Li Ka Shing Center, room 120.

Same as: BIOS 221

**STATS 367. Statistical Models in Genetics. 3 Units.**

Statistical problems in association and linkage analysis of qualitative and quantitative traits in human and experimental populations; sequence alignment and analysis; population genetics/evolution (Wright-Fisher model, Kingman coalescent, models of nucleotide substitution); related computational algorithms. Prerequisites: knowledge of probability through elementary stochastic processes and statistics through likelihood theory.

**STATS 370. A Course in Bayesian Statistics. 3 Units.**

Advanced-level Bayesian statistics. Topics: Discussion of the mathematical and theoretical foundation for Bayesian inferential procedures. Examination of the construction of priors and the asymptotic properties of likelihoods and posterior densities. Discussion including but not limited to the case of finite dimensional parameter space. Prerequisite: familiarity with standard probability and multivariate distribution theory.

Same as: STATS 270

**STATS 374. Large Deviations Theory. 3 Units.**

Combinatorial estimates and the method of types. Large deviation probabilities for partial sums and for empirical distributions, Cramer's and Sanov's theorems and their Markov extensions. Applications in statistics, information theory, and statistical mechanics. Prerequisite: MATH 230A or STATS 310. Offered every 2-3 years.

Same as: MATH 234

**STATS 375. Inference in Graphical Models. 3 Units.**

Graphical models as a unifying framework for describing the statistical relationships between large sets of variables; computing the marginal distribution of one or a few such variables. Focus is on sparse graphical structures, low-complexity algorithms, and their analysis. Topics include: variational inference; message passing algorithms; belief propagation; generalized belief propagation; survey propagation. Analysis techniques: correlation decay; distributional recursions. Applications from engineering, computer science, and statistics. Prerequisite: EE 278, STATS 116, or CS 228. Recommended: EE 376A or STATS 217.

**STATS 376A. Information Theory. 3 Units.**

The fundamental ideas of information theory. Entropy and intrinsic randomness. Data compression to the entropy limit. Huffman coding. Arithmetic coding. Channel capacity, the communication limit. Gaussian channels. Kolmogorov complexity. Asymptotic equipartition property. Information theory and Kelly gambling. Applications to communication and data compression. Prerequisite: EE178 or STATS 116, or equivalent.

Same as: EE 376A

**STATS 376B. Network Information Theory. 3 Units.**

Network information theory deals with the fundamental limits on information flow in networks and the optimal coding schemes that achieve these limits. It aims to extend Shannon's point-to-point information theory and the Ford-Fulkerson max-flow min-cut theorem to networks with multiple sources and destinations. The course presents the basic results and tools in the field in a simple and unified manner. Topics covered include: multiple access channels, broadcast channels, interference channels, channels with state, distributed source coding, multiple description coding, network coding, relay channels, interactive communication, and noisy network coding. Prerequisites: EE376A.

Same as: EE 376B

**STATS 390. Consulting Workshop. 1-3 Unit.**

Skills required of practicing statistical consultants, including exposure to statistical applications. Students participate as consultants in the department's drop-in consulting service, analyze client data, and prepare formal written reports. Seminar provides supervised experience in short term consulting. May be repeated for credit. Prerequisites: course work in applied statistics or data analysis, and consent of instructor.

**STATS 396. Research Workshop in Computational Biology. 1-2 Unit.**

Applications of Computational Statistics and Data Mining to Biological Data. Attendance mandatory. Instructor approval required.

**STATS 397. PhD Oral Exam Workshop. 1 Unit.**

For Statistics PhD students defending their dissertation.

**STATS 398. Industrial Research for Statisticians. 1 Unit.**

Doctoral research as in 298, but must be conducted for an off-campus employer. Final report required. May be repeated for credit. Prerequisite: Statistics Ph.D. candidate.

**STATS 399. Research. 1-10 Unit.**

Research work as distinguished from independent study of nonresearch character listed in 199. May be repeated for credit.

**STATS 801. TGR Project. 0 Units.**

.

**STATS 802. TGR Dissertation. 0 Units.**

.