STATISTICAL SCIENCE (STSCI)

STSCI 1380 - Data Science for All (4 Credits)
Crosslisted with CS 1380, ORIE 1380, INFO 1380
This course provides an introduction to data science using the statistical programming language R. We focus on building skills in inferential thinking and computational thinking, guided by the practical questions we seek to answer from data sets arising in medicine, economics and other social sciences. The course starts with essential R programming principles, and how to use R for data manipulation, visualization, and sampling. These techniques are then used to summarize and visualize real data sets, draw meaningful conclusions from those data, and assess the uncertainty surrounding those conclusions. Throughout the process, students will learn to develop hypotheses about their data, and use simulations and statistical techniques to test these hypotheses. The course also covers how to use the Tidyverse open-source R packages to clean and organize complex data sets, and create high quality graphics for data visualization.
Distribution Requirements: (DLG-AG, OPHLS-AG), (SDS-AS), (STA-IL)
Last Four Terms Offered: Spring 2024, Spring 2023, Spring 2022, Winter 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 2000 - Essential Statistics and Data Science (3 Credits)
This course will cover essential tools for data collection and analysis in the modern age of big data. Students will learn foundational topics in statistics, including the notions of sampling, data summarization and visualization, and statistical inference. Students will also learn basics of R programming required for exploratory data analysis. In addition, modern techniques for simulation based statistical inference will be covered to complement classical statistical inference.
Enrollment Information: Enrollment limited to: Ed Equity students.
Last Four Terms Offered: Spring 2025
Learning Outcomes:

- Identify appropriate summarization, visualization and uncertainty quantification techniques for analyzing tabular data
- Conduct descriptive and inferential statistical analysis using R
- Interpret and communicate results of descriptive and inferential statistical analysis

Schedule of Classes (https://classes.cornell.edu/)

STSCI 2100 - Introductory Statistics and Data Science (4 Credits)
Crosslisted with ILRST 2100
Statistics is about understanding the world through data. We are surrounded by data, so there is a lot to understand. Covers data exploration and display, data gathering methods, probability, and statistical inference methods through contingency tables and linear regression. The emphasis is on thinking scientifically, understanding what is commonly done with data (and doing some of it for yourself), and laying a foundation for further study. Students learn to use statistical software and simulation tools to discover fundamental results. They use computers regularly; the test includes both multimedia materials and a software package. This course does not focus on data from any particular discipline, but will use real-world examples from a wide variety of disciplines and current events.
Forbidden Overlaps: AEM 2100, BTRY 3010, BTRY 6010, CRP 1200, ENGRD 2700, HADM 2010, HADM 2011, ILRST 2100, ILRST 6100, MATH 1710, PSYCH 2500, PUBPOL 2100, PUBPOL 2101, SOC 3010, STSCI 2100, STSCI 2150, STSCI 2200. In addition, no credit for MATH 1710 if taken after ECON 3130, ECON 3140, MATH 4720, or any other upper-level course focusing on the statistical sciences.
Distribution Requirements: (DLS-AG, MQL-AG, OPHLS-AG), (ICE-IL, STA-IL), (SDS-AS)
Last Four Terms Offered: Summer 2025, Spring 2025, Winter 2025, Fall 2024
Schedule of Classes (https://classes.cornell.edu/)

STSCI 2110 - Statistical Methods for the Social Sciences II (4 Credits)
Crosslisted with ILRST 2110
A second course in statistics that emphasizes applications to the social sciences. Topics include simple linear regression, multiple linear regression (theory, model building, and model diagnostics), and the analysis of variance. Computer packages are used extensively.
Prerequisites: AEM 2100, CRP 1200, ENGRD 2700, HADM 2010, ILRST 2100, MATH 1710, PUBPOL 2100, PUBPOL 2101, PSYCH 2500, SOC 3010, STSCI 2100, or STSCI 2150.
Forbidden Overlaps: BTRY 3020, ILRST 2110, STSCI 2110, STSCI 3200
Enrollment Information: Open to: undergraduate students.
Distribution Requirements: (DLS-AG, OPHLS-AG), (ICE-IL, STA-IL), (SDS-AS)
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Spring 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 2120 - R Programming for Data Science (2 Credits)
The course will cover the basics of R programming for reading and writing data, executing simple operations on data, performing basic analyses, and producing visual graphics. An emphasis will be placed on practical aspects and understanding of fundamental R objects and functions.
Distribution Requirements: (STA-IL)
Last Four Terms Offered: Fall 2024, Fall 2023
Learning Outcomes:

- Describe and operate on basic R data structures such as vectors, matrices, and data frames.
- Access and get help from R documentation.
- Read and write data from and to CSV and RData formats.
- Visualize data using basic R graphics capabilities.
- Calculate statistics from data and print results in a readable format.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 2130 - Applied Regression Analysis (2 Credits)
Crosslisted with ILRST 2130
This seven week, two-credit class will cover the regression requirements, hypothesis tests, and interpretation of results. Students will learn to identify the data necessary to perform a regression analysis, evaluate the conditions, and apply the statistical tests. Interpretation of overall results will be made. Independent/group projects by each student will be done. These will consist of identifying an issue of interest, a relevant data set, and analysis using the regression methods. Presentation of the results in verbal and written form will be required. Recommended for students who want to develop applied analysis skills.
Prerequisites: AEM 2100,CRP 1200, ENGRD 2700, HADM 2010, ILRST 2100, MATH 1710, PUBPOL 2100, PUBPOL 2101, PSYCH 2500, SOC 3010, STSCI 2100, or STSCI 2150.
Distribution Requirements: (ICE-IL, STA-IL)
Last Four Terms Offered: Fall 2024, Fall 2023, Spring 2023, Fall 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 2150 - Introductory Statistics for Biology (4 Credits)
This course provides an introduction to data analysis and statistical inference illustrated with biological applications. The computer labs will teach graphical analysis and statistical computation using R. Topics include graphical display, populations and sampling, probability distributions, expectation and variance, estimation, testing, correlation, regression, contingency tables, and the design of experiments. Emphasis is on concepts and the careful modeling of biological data, so that statistical methods are applied properly, pitfalls are avoided, and sound conclusions are reached.
Forbidden Overlaps: AEM 2100, BTRY 3010, BTRY 6010, CRP 1200, ENGRD 2700, HADM 2010, HADM 2011, ILRST 2100, ILRST 6100, MATH 1710, PSYCH 2500, PUBPOL 2100, PUBPOL 2101, SOC 3010, STSCI 2100, STSCI 2150, STSCI 2200. In addition, no credit for MATH 1710 if taken after ECON 3130, ECON 3140, MATH 4720, or any other upper-level course focusing on the statistical sciences.
Distribution Requirements: (DLS-AG, MQL-AG, OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 2200 - Statistics I (4 Credits)
Crosslisted with BTRY 3010
Students will be able to perform a variety of basic statistical analyses including: t-tests, two-sample t-tests, tests for categorical data, and linear regression.
Prerequisites: MATH 1110 or equivalent.
Forbidden Overlaps: AEM 2100, BTRY 3010, BTRY 6010, CRP 1200, ENGRD 2700, HADM 2010, HADM 2011, ILRST 2100, ILRST 6100, MATH 1710, PSYCH 2500, PUBPOL 2100, PUBPOL 2101, SOC 3010, STSCI 2100, STSCI 2150, STSCI 2200. In addition, no credit for MATH 1710 if taken after ECON 3130, ECON 3140, MATH 4720, or any other upper-level course focusing on the statistical sciences.
Distribution Requirements: (DLS-AG, MQL-AG, OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Learning Outcomes:
- Students will be able to design an experiment using randomization techniques.
- Students will be able to use R Markdown for reproducible research.
- Students will be able to produce effective graphical summaries of collected data.
- Students will learn how sampling distributions are determined and utilized for statistical analysis.
- Students will understand why some estimators are more desirable than others.
- Students will be able to perform a variety of basic statistical analyses including: t-tests, ANOVA, two-sample t-tests, tests for categorical data, linear regression, and multiple linear regression.
- Students will be able to assess the quality of a statistical analysis.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 2220 - R Programming for Data Science II (2 Credits)
Statistics courses usually use clean and well-behaved data, this leaves many unprepared for the messiness and chaos of data in the real world. This course will follow on from STSCI 2120 and cover more advanced data wrangling topics including how to tidy data using the tidyverse R packages to better facilitate data analysis. This includes string processing with regular expressions, manipulating date and time data, web scraping, and text mining. Data visualization topics will cover visualization principles, the use of ggplot2 to create custom plots, and how to communicate data-driven findings.
Prerequisites: STSCI 2120 or permission of instructor.
Distribution Requirements: (STA-IL)
Learning Outcomes:
- Demonstrate ability to combine and tidy data using the tidyverse R package.
- Produce professional and informative data visualizations using the ggplot2 R package.
- Create reports to document data analysis and communicate findings using RMarkdown.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 3040 - R Programming for Data Science (4 Credits)
Statistics courses usually use clean and well-behaved data, this leaves many unprepared for the messiness and chaos of data in the real world. This course aims to prepare students for dealing with data using the R programming language. The introduction will overview the basic R syntax, foundational R programming concepts such as data types, vectors arithmetic, and indexing, and importing data into R from different file formats. The data wrangling topics include how to tidy data using the tidy verse to better facilitate analysis, string processing with regular expressions and with dates and times as file formats, web scraping, and text mining. Data visualization topics will cover visualization principles, the use of ggplot2 to create custom plots, and how to communicate data-driven findings.
Prerequisites: ECON 3110/STSCI 3110, ENGRD 2700.
Forbidden Overlaps: AEM 2850, GDEV 4290, GDEV 5290, NTRES 6100, STSCI 3040, STSCI 5040
Distribution Requirements: (STA-IL)
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Learning Outcomes:
- Learn basic R syntax, foundational R programming concepts such as data types, vectors arithmetic, and indexing, and importing data into R from different file formats.
- Learn data wrangling topics include how to tidy data using the tidy verse.
- Produce professional and informative data visualizations.
- Use R Markdown to create reports to document data analysis and communicate findings.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 3080 - Probability Models and Inference (4 Credits)
Crosslisted with BTRY 3080, ILRST 3080
This course provides an introduction to probability and parametric inference. Topics include: random variables, standard distributions, the law of large numbers, the central limit theorem, likelihood-based estimation, the method of moments, sampling distributions and confidence intervals.
Prerequisites: STSCI 2150 or STSCI 2200, MATH 1120 and MATH 2220 or their equivalents.
Forbidden Overlaps: BTRY 3080, ECON 3110, ECON 3130, ILRST 3080, ILRST 3110, MATH 4710, STSCI 3080, STSCI 3110
Distribution Requirements: (DLS-AG, OPHLS-AG), (ICE-IL), (SDS-AS)
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Learning Outcomes:
- Students will be able to manipulate random variables and their distributions using differential and integral calculus.
- Students will be able to derive properties of standard probability.
- Students will be able to derive maximum likelihood estimators for standard probability distributions and discuss their properties.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 3090 - Financial Math for Actuarial Science (4 Credits)
Crosslisted with BTRY 3090
This course will cover financial mathematics and financial instruments relevant to exam FM offered by the Society of Actuaries. Topics on the present and accumulated value of future cash flows will be covered including the measurement of simple and compound interest, annuities, yield rates, amortization schedules, bonds.
Prerequisites: BTRY 3080 or permission from the instructor.
Distribution Requirements: (SMR-AS)
Last Four Terms Offered: Fall 2023, Fall 2022, Spring 2020, Spring 2018
Learning Outcomes:
- Apply the principles of financial mathematics.
- Understand the fundamentals of financial instruments.
- Have a strong background for studying for the SOA FM exam.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 3100 - Statistical Sampling (4 Credits)
Crosslisted with BTRY 3100, ILRST 3100
Theory and application of statistical sampling, especially in regard to sample design, cost, estimation of population quantities, and error estimation. Assessment of nonsampling errors. Discussion of applications to social and biological sciences and to business problems.
Prerequisites: STSCI 2150 or STSCI 2200/BTRY 3010 or equivalent, STSCI 3200/BTRY 3020 or BTRY 6020.
Distribution Requirements: (DLS-AG, OPHLS-AG), (ICE-IL), (SDS-AS)
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 3110 - Applied Probability and Statistics (4 Credits)
Crosslisted with ECON 3110, ILRST 3110
This course provides an introduction to probability and parametric inference. Topics include: random variables, standard distributions, the law of large numbers, the central limit theorem, likelihood-based estimation, sampling distributions and hypothesis testing.
Prerequisites: MATH 1106 or MATH 1110.
Forbidden Overlaps: BTRY 3080, ECON 3110, ECON 3130, ILRST 3080, ILRST 3110, MATH 4710, STSCI 3080, STSCI 3110
Enrollment Information: Open to: undergraduate students.
Distribution Requirements: (DLG-AG, MQL-AG, OPHLS-AG), (ICE-IL, STA-IL), (SDS-AS, SMR-AS)
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 3200 - Statistics II (4 Credits)
Crosslisted with BTRY 3020
Applies linear statistical methods to quantitative problems addressed in biological and environmental research. Methods include linear regression, inference, model assumption evaluation, the likelihood approach, matrix formulation, generalized linear models, single-factor and multifactor analysis of variance (ANOVA), and a brief foray into nonlinear modeling. Carries out applied analysis in a statistical computing environment.
Prerequisites: BTRY 3010 or equivalent.
Forbidden Overlaps: BTRY 3020, ILRST 2110, STSCI 2110, STSCI 3200
Distribution Requirements: (DLS-AG, OPHLS-AG)
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Learning Outcomes:
  • Students will be able to design a statistical experiment using randomization techniques.
  • Students will be able to analyze multivariate linear and nonlinear data that include quantitative and qualitative variables.
  • Students will be able to apply generalized linear model, generalized additive models, and mixed effects models to appropriately collected data.
  • Students will be able to formulate and evaluate parametric and nonparametric methods for determining model uncertainty.
  • Students will be able to employ matrix methods to effectively design and implement linear models.
  • Students will be able to assess the quality of a statistical analysis.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 3510 - Stochastic Processes for Decision-Making (4 Credits)
Crosslisted with ORIE 3510
Uses basic concepts and techniques of random processes to construct models for a variety of problems of practical interest. Topics include: the Poisson process, Markov chains, renewal theory, models for queuing, and reliability.
Prerequisites: ORIE 3500 or equivalent.
Distribution Requirements: (OPHLS-AG)
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 3600 - Integrated Ethics in Data Science (2 Credits)
Integrated Ethics in Data Science examines your responsibilities in data analysis. Our investigation starts with the aggregated impacts of data science on fairness, privacy, and justice outcomes for groups and individuals. Use of supplied data and applications are analyzed using agency, moral imagination, and virtue ethics. Responsible practices in data science, codes of conduct, and current regulations will be applied. Evaluation of the act of speaking up, supporting others and working with an ethics committee will develop professional skills. Case studies from legal issues, policy concerns, and industry practices provide problems to evaluate the individual choices that led to the results. Course success depends on your frequent written work, engagement in small group discussions, and planning for professional practice.
Distribution Requirements: (OCE-IL, STA-IL)
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Fall 2022
Learning Outcomes:
  • Identify ethical conflicts in data science practices.
  • Analyze the connection between individual ethical choices and aggregated outcome impacts.
  • Create a plan for individual moral awareness, habits, and virtue development.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 3740 - Data Mining and Machine Learning (4 Credits)
We start off with a detailed refresher for Linear Regression. We then turn to popular methods for classification including Logistic Regression and Discriminant Analysis. Finally, we consider more advanced topics which may include - depending on the audience - Resampling Methods, Tree-based Methods, or Support Vector Machines. The statistics software R is introduced and used for applications.
Prerequisites: CS 1112 or equivalent, MATH 2220, STSCI 3200, STSCI 3080 or MATH 4710.
Forbidden Overlaps: CS 3780, CS 5780, ECE 3200, ECE 5420, ORIE 3741, ORIE 5741, STSCI 3740, STSCI 5740
Distribution Requirements: (DLG-AG, OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 3900 - Causal Inference (3 Credits)
Crosslisted with INFO 3900, ILRST 3900
Causal claims are essential in both science and policy. Would a new experimental drug improve disease survival? Would a new advertisement cause higher sales? Would a person's income be higher if they finished college? These questions involve counterfactuals: outcomes that would be realized if a treatment were assigned differently. This course will define counterfactuals mathematically, formalize conceptual assumptions that link empirical evidence to causal conclusions, and engage with statistical methods for estimation. Students will enter the course with knowledge of statistical inference: how to assess if a variable is associated with an outcome. Students will emerge from the course with knowledge of causal inference: how to assess whether an intervention to change that input would lead to a change in the outcome.
Prerequisites: STSCI 2100 or PSYCH 2500 or SOC 3010 or ECON 3110 or equivalent.
Distribution Requirements: (DLS-AG), (ICE-IL, STA-IL)
Last Four Terms Offered: Fall 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4030 - Linear Models with Matrices (4 Credits)
Crosslisted with BTRY 4030
The focus of this course is the theory and application of the general linear model expressed in its matrix form. Topics will include: least squares estimation, multiple linear regression, coding for categorical predictors, residual diagnostics, anova decomposition, polynomial regression, model selection techniques, random effects and mixed models, maximum likelihood estimation and distributional theory assuming normal errors. Homework assignments will involve computation using the R statistical package.
Prerequisites: STSCI 2150 or STSCI 2200/BTRY 3010, BTRY 3080, MATH 1920, MATH 2210 or their equivalents, STSCI 3200/BTRY 3020 or BTRY 6020.
Distribution Requirements: (DLS-AG, OPHLS-AG), (SDS-AS), (STA-IL)
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Learning Outcomes:
  • Students will be able to discuss the mathematical foundations of linear statistical models using matrix algebra.
  • Students will be able to use diagnostic measures to assess the validity of a given statistical model.
  • Students will be able to analyze data involving both fixed and random factors.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4050 - Modern Regression Models for Data Science (4 Credits)
Prerequisites: STSCI 3080, STSCI 3200 or equivalents.
Last Four Terms Offered: Spring 2024
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4060 - Python Programming and its Applications in Statistics (4 Credits)
The first part of the course teaches basic Python programming knowledge and skills, such as Python variables, data containers, language controls, functions, objects, class, data structures, regular expressions, graphics, GUI, Jupyter notebook, etc. The second part deals with Python application in statistics (e.g., 2D/3D data visualization and statistical analysis, using some important Python packages for statistical computing and machine learning, for example, Numpy, Scipy, Pandas, and Scikit-learn, etc.) Python-database integration (e.g., access, update and control an Oracle database), and Python web services (e.g., database-driven dynamic webpages using Python CGI scripts). These techniques are utilized in a comprehensive course project.
Prerequisites: basic programming skills (any language), one introductory Statistics course, SQL (Oracle preferred).
Distribution Requirements: (OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Fall 2024, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4090 - Theory of Statistics (4 Credits)
Crosslisted with BTRY 4090
Introduction to classical theory of parametric statistical inference that builds on the material covered in BTRY 3080. Topics include: sampling distributions, principles of data reduction, likelihood, parameter estimation, hypothesis testing, interval estimation, and basic asymptotic theory.
Prerequisites: BTRY 3080 or MATH 4710 or equivalent and at least one introductory statistics course.
Forbidden Overlaps: BTRY 4090, ECON 3130, MATH 4720, STSCI 4090
Distribution Requirements: (DLS-AG, OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Learning Outcomes:
  • Describe the general principles of statistical estimation and testing.
  • Design a statistical estimator in a principled way based on a description of a dataset.
  • Analyze the theoretical properties of an estimator and a hypothesis test.
  • Calculate and correctly interpret confidence intervals, p-values, statistical significance, and power.
  • Recognize the general principles underlying common statistical procedures.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4100 - Multivariate Analysis (4 Credits)
Crosslisted with ILRST 4100, BTRY 4100
This course is on the basics of multivariate statistical analysis. The focus ison the applied side, and the students will learn by examples of multiple real-life datasets. Studentswill learn to visualize the datasets and conduct simple statistical analysis using linear/nonlinearmethods. We will also cover web-scraping and data cleaning.
Prerequisites: STSCI 2100 or equivalent.
Distribution Requirements: (DLS-AG, OPHLS-AG), (ICE-IL), (SDS-AS)
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2016
Learning Outcomes:
  • Prepares the students for real-life multivariate data analysis. The students will get their hands dirty in messy datasets and learn that each dataset calls for its own approach of analysis. They will get more familiar with manipulate datasets in R, collaborate with others, and enhance their skills in creative thinking and presentation.
  • Students will be able to analyze multivariate data using modern statistical software.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4110 - Categorical Data (3 Credits)
Crosslisted with BTRY 4110, ILRST 4110
Categorical data analysis, including logistic regression, log-linear models, stratified tables, matched pairs analysis, polytomous response, and ordinal data. Applications in biological, biomedical and social sciences.
Prerequisites: BTRY 3020, BTRY 6020, or equivalent with BTRY 3080 or MATH 4710 also highly recommended.
Distribution Requirements: (DLS-AG, OPHLS-AG), (ICE-IL), (SDS-AS)
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4140 - Applied Design (4 Credits)
Crosslisted with BTRY 4140, ILRST 4140
This course begins with a discussion of some general principles of experimental design. Classical designs are covered in detail, motivated by real data applications. These include completely randomized, randomized block, balanced incomplete block, split-plot, repeated measures and fractional factorial designs. If time permits rank-based nonparametric versions of the classical designs will also be covered.
Prerequisites: STSCI 3200 or equivalent.
Distribution Requirements: (ICE-IL), (OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Spring 2022, Spring 2021, Spring 2020, Spring 2019
Learning Outcomes:
- Students will be able to explain the basic design principles such as randomization, blocking and stratification.
- Students will be able to determine an appropriate design based on design principles.
- Students will be able to apply standard designs to date using modern statistical software and interpret the results.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4270 - Introduction to Survival Analysis and Loss Models (3 Credits)
Crosslisted with BTRY 4270
Develops and uses statistical methods appropriate for analyzing right-censored (i.e., incomplete) time-to-event data. Topics covered include nonparametric estimation (e.g., life table methods, Kaplan Meier estimator), nonparametric methods for comparing the survival experience of two or more populations, and semiparametric and parametric methods of regression for censored outcome data. Emphasis is given to applications in medicine and actuarial studies. Substantial use is made of the R statistical software package.
Prerequisites: BTRY 3080 or equivalent, MATH 1120 or equivalent, and MATH 2210 or equivalent.
Distribution Requirements: (DLS-AG, OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Spring 2023, Fall 2019, Fall 2018, Fall 2017
Learning Outcomes:
- Students will be able to conduct appropriate nonparametric and parametric analyses of right-censored survival data using the R software language, including tabular and graphical methods (i.e., life tables and Kaplan Meier plots), hypothesis testing (e.g., logrank tests and Wald tests) and likelihood-based methods of regression (i.e., proportional hazards and accelerated failure time regression models).
- Students will be able to interpret the results of a statistical analysis involving right censored survival data as well as articulate the associated limitations of such analyses.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4520 - Statistical Computing (4 Credits)
This course is designed to provide students with an introduction to statistical computing. The class will cover the basics of programming; numerical methods for optimization and linear algebra and their application to statistical estimation, generating random variables, bootstrap, jackknife and permutation methods, Markov Chain Monte Carlo methods, Bayesian inference and computing with latent variables.
Prerequisites: BTRY 3080 or MATH 4710, enrollment in MATH 2220 and MATH 2240 or equivalents. Previous programming experience is recommended.
Distribution Requirements: (OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Learning Outcomes:
- Students will be able to enter, manipulate and plot data and run basic statistical analyses in R.
- Students will be able to implement estimators for non-standard statistical problems in R.
- Students will be able to simulate random variables and random experiments in R.
- Students will be able to design and implement Monte Carlo methods to evaluate integrals and perform simulations.
- Students will be able to design and conduct appropriate resampling methods to estimate sampling variance for statistical estimates.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4550 - Applied Time Series Analysis (4 Credits)
Crosslisted with ILRST 4550
Introduces statistical tools for the analysis of time-dependent data. Data analysis and application will be an integral part of this course. Topics include linear, nonlinear, seasonal, multivariate modeling, and financial time series.
Prerequisites: BTRY 3080 or equivalent, STSCI 4030 or ECON 3140, or permission of instructor.
Distribution Requirements: (DLS-AG, OPHLS-AG), (ICE-IL), (SDS-AS)
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4610 - Data Science in Risk Modeling (2 Credits)
The course teaches statistical methods used in modeling risk in asset returns. Students in this course will be able to: identify time series dependency in selected financial data, analyze trade-off between risk and return of a portfolio, analyze tail risk in context of asset returns, and apply factor analysis in context of asset returns.
Prerequisites: STSCI 3080.
Last Four Terms Offered: Spring 2024
Learning Outcomes:
- Identify time series dependency in selected financial data.
- Analyze trade-offs between risk and return of a portfolio.
- Analyze tail risk in the context of asset returns.
- Apply factor analysis in the context of asset returns.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4630 - Operations Research Tools for Financial Engineering (4 Credits)
Crosslisted with ORIE 4630
Introduction to the applications of OR techniques, e.g., probability, statistics, and optimization, to finance and financial engineering. The course reviews probability and statistics and surveys assets returns, ARIMA time series models, portfolio selection using quadratic programming, regression, CAPM and factor models, option pricing, GARCH models, fixed-income securities, and resampling techniques. Covers the use of R for statistical calculations, simulation, and optimization.
Prerequisites: engineering math through MATH 2940, ENGRD 2700 and ORIE 3500, and knowledge of R and multiple linear regression equivalent to ORIE 3120. No previous knowledge of finance required.
Distribution Requirements: (DLS-AG, OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4750 - Understanding Machine Learning (4 Credits)
The goal of this course is to teach you why machine learning works and how to implement it. We will cover the essentials of learning theory, including the probably approximately correct (PAC) framework and the bias-complexity tradeoff. We will then see how these concepts shed light on the mathematics behind linear regression, logistic regression, boosting (and AdaBoost), support vector machines and neural networks. We cover clustering algorithms and how to implement them. Data will be analyzed using modern software packages with the above algorithms, with the aim of reinforcing the mathematics behind them.
Prerequisites: CS 1110 or equivalent, MATH 4710, STSCI 3080, STSCI 4030 or STSCI 5030. Recommended prerequisite: STSCI 3740.
Last Four Terms Offered: Spring 2023, Spring 2022
Learning Outcomes:
  • Students will be able to demonstrate an understanding of how concepts in learning theory quantify the performance of the learning algorithms in the course description.
  • Students will be able to indicate a competency of how and in which circumstances to apply modern machine learning algorithms to real and simulated data.
  • Students will be able to verify theoretical results-such as the Fundamental Theorem of Statistical Learning-in practice using the software packages introduced and taught in the course.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4780 - Bayesian Data Analysis: Principles and Practice (4 Credits)
Bayesian data analysis uses probability theory as a kind of calculus of inference, specifying how to quantify and propagate uncertainty in data-based chains of reasoning. Students will learn the fundamental principles of Bayesian data analysis, and how to apply them to varied data analysis problems across science and engineering. Topics include: basic probability theory, Bayes's theorem, linear and nonlinear models, hierarchical and graphical models, basic decision theory, and experimental design. There will be a strong computational component, using a high-level language such as R or Python, and a probabilistic language such as BUGS or Stan.
Prerequisites: BTRY 3080 and BTRY 3020/STSCI 3200, or equivalent.
Distribution Requirements: (DLG-AG, OPHLS-AG), (SDS-AS)
Last Four Terms Offered: Spring 2022, Spring 2020, Spring 2018, Spring 2015
Learning Outcomes:
  • A basic understanding of the principles and foundations underlying the Bayesian approach.
  • Practical experience using basic/intermediate Bayesian methods.
  • Experience with widely-used tools and software development practices for producing and sharing collaborative, reproducible statistical research.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4850 - Data Science Consulting (2 Credits)
In this course students will learn about the consulting process using data science. They will understand how statistics and data science knowledge can be applied to real world questions starting with exploratory analysis and then applying specific modeling tools. In addition to analyzing a problem empirically, they will also learn core professional soft skills required to work with a client. These skills include designing and implementing a work plan, learning to communicate efficiently, presentation of a final deliverable to the client, as well as learning good practices to conduct data analysis, documentation, quality control and collaboration. Through hands-on experience and real-world examples, students will develop a basic understanding of consulting, and be familiar with the professional standard expected in the industry.
Prerequisites: STSCI 2100, STSCI 2150, ENGRD 2700.
Last Four Terms Offered: Spring 2025
Learning Outcomes:
  • Demonstrate the consulting process and working with a client.
  • Analyze example problems with data science tools and techniques.
  • Demonstrate good practice to conduct data analysis and documentation.
  • Recognize how to manage a project, organize team structure, communicate and collaborate.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4940 - Undergraduate Special Topics in Statistics (1-3 Credits)
Course of lectures selected by the faculty. Because topics usually change from year to year, this course may be repeated for credit.
Last Four Terms Offered: Fall 2020, Spring 2020, Fall 2019, Spring 2019
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4950 - Statistical Consulting (2 Credits)
Crosslisted with BTRY 4950
This course will give students the opportunity to apply the statistical knowledge gained in their courses to real-life problems. Students will be integrated in the Cornell Statistical Consulting Unit (CSCU) and be exposed to various areas in which statistical methods are applied. Students will gain experience in choosing appropriate statistical procedures and their implementations in various statistical software packages. They will also learn how to communicate effectively to understand the client's problem and to explain methods and results to non-statisticians.
Prerequisites: BTRY 3010 and BTRY 3020, or BTRY 6010 and BTRY 6020.
Last Four Terms Offered: Fall 2023, Fall 2022, Fall 2021, Fall 2020
Learning Outcomes:
- Integrate the statistical knowledge gained in courses and apply them to real-life problems.
- Learn to communicate effectively with clients to gather the information needed to make the link between the research questions to be addressed and the statistical methods.
- Research the application of statistical methodologies that are useful to clients and explain them to an audience of non-statisticians.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4970 - Undergraduate Supervised Teaching (1-4 Credits)
Students assist in teaching a course appropriate to their previous training. Students meet with a discussion or laboratory section and regularly discuss objectives with the course instructor.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4980 - Tutorial in Actuarial Statistics (2 Credits)
Problem solving sessions to prepare students for the first four actuarial examinations (probability, financial mathematics, statistical modeling, and risk management).
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4990 - Undergraduate Individual Study in Statistics (1-4 Credits)
Course consists of individual tutorial study selected by faculty. Because topics usually change year to year this course may be repeated for credit.
Exploratory Studies: (CU-UG)

Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 4995 - Internship in Data Science (1 Credit)
Students planning internships related to Statistics and Data Science are encouraged to enroll in the departmental internship course.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Learning Outcomes:
- Demonstrate professional skills that pertain directly to the internship experience.
- Demonstrate verbal and written communication skills. Participate well as a team member and build professional network.
- Demonstrate effect management of personal behavior, ethics and attitudes.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 4999 - Undergraduate Dissertation Research (1-4 Credits)
Research at the undergraduate level.
Exploratory Studies: (CU-UG)

Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5010 - Applied Statistical Computation with SAS (4 Credits)
This course teaches the basics of SAS (Statistical Analysis System) programming and the SAS Enterprise Miner software. This course is composed two modules. The first module, in the first 12 weeks, covers the objectives tested on the SAS Base Programming for SAS 9 Exam, including basic SAS programming concepts, producing reports, creating and modifying SAS data sets, reading various types of raw data and other data handling techniques. At the end of module 1, all the students will take the SAS Base Programming for SAS 9 Exam, which is administered by the MPS Program in Applied Statistics on the Cornell campus in conjunction with the SAS Institute, Inc. The second module, in the last three weeks, introduces the SAS Enterprise Miner software and cluster analysis. Students will learn how to use the SAS Enterprise Miner software and SAS procedures to do cluster analysis.
Last Four Terms Offered: Spring 2025, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5030 - Linear Models with Matrices (4 Credits)
The focus of this course is the theory and application of the general linear model expressed in its matrix form. Topics will include: least squares estimation, multiple linear regression, coding for categorical predictors, residual diagnostics, anova decomposition, polynomial regression, model selection techniques, random effects and mixed models, maximum likelihood estimation and distributional theory assuming normal errors. Homework assignments will involve computation using the R statistical package.
Prerequisites: STSCI 2150 or STSCI 2200/BTRY 3010, BTRY 3080, MATH 1920, MATH 2210 or their equivalents, STSCI 3200/BTRY 3020 or BTRY 6020.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5040 - R Programming for Data Science (4 Credits)
Statistics courses usually use clean and well-behaved data, this leaves many unprepared for the messiness and chaos of data in the real world. This course aims to prepare students for dealing with data using the R programming language. The introduction will overview the basic R syntax, foundational R programming concepts such as data types, vectors arithmetic, and indexing, and importing data into R from different file formats. The data wrangling topics include how to tidy data using the tidy verse to better facilitate analysis, string processing with regular expressions and with dates and times as file formats, web scraping, and text mining. Data visualization topics will cover visualization principles, the use of ggplot2 to create custom plots, and how to communicate data-driven findings.
Prerequisites: STSCI 2200 or equivalent.
Forbidden Overlaps: AEM 2850, GDEV 4290, GDEV 5290, NTRES 6100, STSCI 3040, STSCI 5040
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Learning Outcomes:
- Learn basic R syntax, foundational R programming concepts such as data types, vectors arithmetic, and indexing, and importing data into R from different file formats.
- Learn data wrangling topics include how to tidy data using the tidy verse.
- Produce professional and informative data visualizations.
- Use R Markdown to create reports to document data analysis and communicate findings.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5045 - Python Programming and its Applications in Statistics (4 Credits)
The first part of the course teaches basic Python programming knowledge and skills, such as Python variables, data containers, language controls, functions, objects, class, data structures, regular expressions, graphics, GUI, Jupyter notebook, etc. The second part deals with Python application in statistics (e.g., 2D/3D data visualization and statistical analysis, using some important Python packages for statistical computing and machine learning, for example, Numpy, Scipy, Pandas, and Scikit-learn, etc.) Python-database integration (e.g., access, update and control an Oracle database), and Python web services (e.g., database-driven dynamic webpages using Python CGI scripts). These techniques are utilized in a comprehensive course project.
Prerequisites: STSCI 5060 (or basic SQL programming skill), and one intro statistics course.
Last Four Terms Offered: Fall 2024, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5050 - Modern Regression Models for Data Science (4 Credits)
Prerequisites: STSCI 3200, STSCI 3080 or equivalents.
Last Four Terms Offered: Spring 2024
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5060 - Database Management and SAS High Performance Computing with DBMS (4 Credits)
Using relational databases in statistical computing has become more and more important. The knowledge and skill of database management and the ability to combine this knowledge and skill with statistical analysis software tools, such as SAS, are a critical qualification of a statistical analyst. In this course we will study 1) the basics of modern relational database management systems, including database analysis, design and implementation, 2) database application in advanced SAS programming and, 3) SAS high performance computing using database-related techniques.
Corequisites: STSCI 5010 or Base SAS programming knowledge and skills.
Enrollment Information: Enrollment limited to: students in the MPS Program in Applied Statistics.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5065 - Big Data Management and Analysis (3 Credits)
Concepts, challenges, and industry trends of big data, with a focus on the Hadoop system. Topics include: basics of the Apache Hadoop platform and Hadoop ecosystem; the Hadoop distributed file system (HDFS); MapReduce or its alternative, a parallel programming model for distributed processing of large data sets; common big data tools, such as Pig (a procedural data processing language for Hadoop parallel computation), Hive (a declarative SQL-like language to handle Hadoop jobs), HBase (the most popular NoSQL database), and YARN; case studies; and integration of Hadoop with statistical software packages, e.g., SAS and R.
Prerequisites: knowledge of a general purpose computer programming language, such as JAVA, Python, Ruby, or C++, or at least taking STSCI 4060 in parallel with this course; STSCI 5060 or basic SQL knowledge; STSCI 5010 or basic knowledge of SAS programming; STSCI 4520 or STSCI 4030 or basic knowledge of R programming.
Enrollment Information: Enrollment preference given to: MPS Applied Statistics students.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5080 - Probability Models and Inference (4 Credits)
This course provides an introduction to probability and parametric inference. Topics include: random variables, standard distributions, the law of large numbers, the central limit theorem, likelihood-based estimation, sampling distributions and hypothesis testing, as well as an introduction to Bayesian methods. Some assignments may involve computation using the R programming language.
Prerequisites: STSCI 2150 or STSCI 2200/BTRY 3010 or equivalent.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Learning Outcomes:
- Students will be able to manipulate random variables and their distributions using differential and integral calculus.
- Students will be able to derive properties of standard probability.
- Students will be able to derive maximum likelihood estimators for standard probability distributions and discuss their properties.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5090 - Theory of Statistics (4 Credits)
Crosslisted with BTRY 5090
Introduction to classical theory of parametric statistical inference
that builds on the material covered in BTRY 3080. Topics include
sampling distributions, principles of data reduction, likelihood, parameter
estimation, hypothesis testing, interval estimation, and basic asymptotic
theory.
Prerequisites: BTRY 3080 or MATH 4710, or equivalent and STSCI 2200 or
equivalent.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5100 - Statistical Sampling (4 Credits)
Theory and application of statistical sampling, especially in regard
to sample design, cost, estimation of population quantities, and
error estimation. Assessment of nonsampling errors. Discussion of
applications to social and biological sciences and to business problems.
Prerequisites: STSCI 2150 or STSCI 2200/BTRY 3010 or equivalent,
STSCI 3200/BTRY 3020 or BTRY 6020.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5110 - Statistical Methods for the Social Sciences II (4 Credits)
Crosslisted with ILRST 5110
Second course in statistics that emphasizes applications to the social
sciences. Topics include simple linear regression, multiple linear
regression (theory, model building, and model diagnostics), and the
analysis of variance. Computer packages are used extensively.
Prerequisites: STSCI 5200, BTRY 6010, ILRST 5100, or BTRY 5010.
Distribution Requirements: (ICE-IL)
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Spring
2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5111 - Multivariate Analysis (4 Credits)
This course is on the basics of multivariate statistical analysis. The
focus ison the applied side, and the students will learn by examples of
multiple real-life datasets. Studentswill learn to visualize the datasets and
conduct simple statistical analysis using linear/nonlinearmethods. We
will also cover web-scraping and data cleaning.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023
Learning Outcomes:
  • Prepares the students for real-life multivariate data analysis. The
    students will get their hands dirty in messy datasets and learn that
    each dataset calls for its own approach of analysis. They will get
    more familiar with manipulate datasets in R, collaborate with others,
    and enhance their skills in creative thinking and presentation.
  • Students will be able to analyze multivariate data using modern
    statistical software.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5120 - R Programming for Data Science (2 Credits)
The course will cover the basics of R programming for reading and
writing data, executing simple operations on data, performing basic
analyses, and producing visual graphics. An emphasis will be placed
on practical aspects and understanding of fundamental R objects and
functions.
Last Four Terms Offered: Fall 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5140 - Applied Design (4 Credits)
This course begins with a discussion of some general principles of
experimental design. Classical designs are covered in detail, motivated by
real data applications. These include completely randomized, randomized
block, balanced incomplete block, split-plot, repeated measures and
fractional factorial designs. If time permits rank-based nonparametric
versions of the classical designs will also be covered.
Prerequisites: BTRY 6020 or ILRST 5110 or equivalent.
Last Four Terms Offered: Spring 2022
Learning Outcomes:
  • Students will be able to explain the basic design principles such as
    randomization, blocking and stratification.
  • Students will be able to determine an appropriate design based on
    design principles.
  • Students will be able to apply standard designs to date using modern
    statistical software and interpret the results.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5150 - Introductory Statistics for Biology (4 Credits)
This course provides an introduction to data analysis and statistical
inference illustrated with biological applications. The computer labs
will teach graphical analysis and statistical computation using R.
Topics include graphical display, populations and sampling, probability
distributions, expectation and variance, estimation, testing, correlation,
regression, contingency tables, and the design of experiments. Emphasis
is on concepts and the careful modeling of biological data, so that
statistical methods are applied properly, pitfalls are avoided, and sound
conclusions are reached.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5160 - Categorical Data (3 Credits)
Categorical data analysis, including logistic regression, log-linear models,
stratified tables, matched pairs analysis, polytomous response, and
ordinal data. Applications in biological, biomedical and social sciences.
Prerequisites: BTRY 3020, BTRY 6020, or equivalent with BTRY 3080 or
MATH 4710 also highly recommended.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5200 - Statistics I (4 Credits)
Crosslisted with BTRY 5010
Students will be able to perform a variety of basic statistical analyses
including: t-tests, two-sample t-tests, tests for categorical data, and linear
regression.
Prerequisites: MATH 1110 or equivalent.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5201 - Statistics II (4 Credits)
Crosslisted with BTRY 5020
Applies linear statistical methods to quantitative problems addressed in biological and environmental research. Methods include linear regression, inference, model assumption evaluation, the likelihood approach, matrix formulation, generalized linear models, single-factor and multifactor analysis of variance (ANOVA), and a brief foray into nonlinear modeling. Carries out applied analysis in a statistical computing environment.
Prerequisites: BTRY 3010 or equivalent.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Learning Outcomes:
- Students will be able to design a statistical experiment using randomization techniques.
- Students will be able to analyze multivariate linear and nonlinear data that include quantitative and qualitative variables.
- Students will be able to apply generalized linear model, generalized additive models, and mixed effects models to appropriately collected data.
- Students will be able to formulate and evaluate parametric and nonparametric methods for determining model uncertainty.
- Students will be able to employ matrix methods to effectively design and implement linear models.
- Students will be able to assess the quality of a statistical analysis.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5220 - R Programming for Data Sci II (2 Credits)
Statistics courses usually use clean and well-behaved data, this leaves many unprepared for the messiness and chaos of data in the real world. This course will follow on from STSCI 5120 and cover more advanced data wrangling topics including how to tidy data using the tidyverse R packages to better facilitate data analysis. This includes string processing with regular expressions, manipulating date and time data, web scraping, and text mining. Data visualization topics will cover visualization principles, the use of ggplot2 to create custom plots, and how to communicate data-driven findings.
Prerequisites: STSCI 5120 or equivalent.
Learning Outcomes:
- Demonstrate ability to combine and tidy data using the tidyverse R package.
- Produce professional and informative data visualizations using the ggplot2 R package.
- Create reports to document data analysis and communicate findings using RMarkdown.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5270 - Introduction to Survival Analysis and Loss Models (3 Credits)
Develops and uses statistical methods appropriate for analyzing right-censored (i.e., incomplete) time-to-event data. Topics covered include nonparametric estimation (e.g., life table methods, Kaplan Meier estimator), nonparametric methods for comparing the survival experience of two or more populations, and semiparametric and parametric methods of regression for censored outcome data. Emphasis is given to applications in medicine and actuarial studies. Substantial use is made of the R statistical software package.
Prerequisites: BTRY 3080 or equivalent, MATH 1120 or equivalent, and MATH 2210 or equivalent.
Last Four Terms Offered: Spring 2023
Learning Outcomes:
- Students will be able to conduct appropriate nonparametric and parametric analyses of right-censored survival data using the R software language, including tabular and graphical methods (i.e., life tables and Kaplan Meier plots), hypothesis testing (e.g., logrank tests and Wald tests) and likelihood-based methods of regression (i.e., proportional hazards and accelerated failure time regression models).
- Students will be able to interpret the results of a statistical analysis involving right censored survival data as well as articulate the associated limitations of such analyses.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5520 - Statistical Computing (4 Credits)
This course is designed to provide students with an introduction to statistical computing. The class will cover the basics of programming; numerical methods for optimization and linear algebra and their application to statistical estimation, generating random variables, bootstrap, jackknife and permutation methods, Markov Chain Monte Carlo methods, Bayesian inference and computing with latent variables.
Prerequisites: BTRY 3080 or MATH 4710, enrollment in MATH 2220 and MATH 2240 or equivalents. Previous programming experience is recommended.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5550 - Applied Time Series Analysis (4 Credits)
Crosslisted with ORIE 5550
Introduces statistical tools for the analysis of time-dependent data. Data analysis and application will be an integral part of this course. Topics include linear, nonlinear, seasonal, multivariate modeling, and financial time series.
Prerequisites: BTRY 3080 or equivalent, STSCI 4030 or ECON 3140, or permission of instructor.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5600 - Integrated Ethics in Data Science (2 Credits)
Integrated Ethics in Data Science examines your responsibilities in data analysis. Our investigation starts with the aggregated impacts of data science on fairness, privacy, and justice outcomes for groups and individuals. Use of supplied data and applications are analyzed using agency, moral imagination, and virtue ethics. Responsible practices in data science, codes of conduct, and current regulations will be applied. Evaluation of the act of speaking up, supporting others and working with an ethics committee will develop professional skills. Case studies from legal issues, policy concerns, and industry practices provide problems to evaluate the individual choices that led to the results. Course success depends on your frequent written work, engagement in small group discussions, and planning for professional practice.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Fall 2022
Learning Outcomes:
  • Identify ethical conflicts in data science practices.
  • Analyze the connection between individual ethical choices and aggregated outcome impacts.
  • Create a plan for individual moral awareness, habits, and virtue development.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5610 - Data Science in Risk Modeling (2 Credits)
The course teaches statistical methods used in modeling risk in asset returns. Students in this course will be able to: identify time series dependency in selected financial data, analyze trade-off between risk and return of a portfolio, analyze tail risk in context of asset returns, and apply factor analysis in context of asset returns.
Prerequisites: STSCI 3080.
Last Four Terms Offered: Spring 2024
Learning Outcomes:
  • Identify time series dependency in selected financial data.
  • Analyze trade-offs between risk and return of a portfolio.
  • Analyze tail risk in the context of asset returns.
  • Apply factor analysis in the context of asset returns.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5630 - Operations Research Tools for Financial Engineering (4 Credits)
Crosslisted with ORIE 5630
Introduction to the applications of OR techniques, e.g., probability, statistics, and optimization, to finance and financial engineering. The course reviews probability and statistics and surveys assets returns, ARIMA time series models, portfolio selection using quadratic programming, regression, CAPM and factor models, option pricing, GARCH models, fixed-income securities, and resampling techniques. Covers the use of R for statistical calculations, simulation, and optimization.
Prerequisites: MATH 2940, ENGRD 2700, ORIE 3500, ORIE 3120.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5640 - Statistics for Financial Engineering (4 Credits)
Crosslisted with ORIE 5640
Regression, ARIMA, GARCH, stochastic volatility, and factor models. Calibration of financial engineering models, estimation of diffusion models, estimation of risk measures, multivariate models and copulas, bayesian statistics. Students are instructed in the use of R software.
Prerequisites: ORIE 3500/ORIE 5500, ORIE 4600 or ORIE 4630 or ORIE 5600 (highly recommended).
Enrollment Information: Primarily for: M.Eng students in financial engineering.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5740 - Data Mining and Machine Learning (4 Credits)
We start off with a detailed refresher for Linear Regression. We then turn to popular methods for classification including Logistic Regression and Discriminant Analysis. Finally, we consider more advanced topics which may include - depending on the audience - Resampling Methods, Tree-based Methods, or Support Vector Machines. The statistics software R is introduced and used for applications.
Prerequisites: CS 1112 or equivalent, MATH 2220, STSCI 3200, STSCI 3080 or MATH 4710.
Forbidden Overlaps: CS 3780, CS 5780, ECE 3200, ECE 5420, ORIE 3741, ORIE 5741, STSCI 3740, STSCI 5740
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5750 - Understanding Machine Learning (4 Credits)
The goal of this course is to teach you why machine learning works and how to implement it. We will cover the essentials of learning theory, including the probably approximately correct (PAC) framework and the bias-complexity tradeoff. We will then see how these concepts shed light on the mathematics behind linear regression, logistic regression, boosting (and AdaBoost), support vector machines and neural networks. We cover clustering algorithms and how to implement them. Data will be analyzed using modern software packages with the above algorithms, with the aim of reinforcing the mathematics behind them.
Prerequisites: CS 1110 or equivalent, MATH 4710, STSCI 3080, STSCI 4030 or STSCI 5030. Recommended prerequisite: STSCI 3740.
Last Four Terms Offered: Spring 2023, Spring 2022
Learning Outcomes:
  • Students will be able to demonstrate an understanding of how concepts in learning theory quantify the performance of the learning algorithms in the course description.
  • Students will be able to indicate a competency of how and in which circumstances to apply modern machine learning algorithms to real and simulated data.
  • Students will be able to verify theoretical results-such as the Fundamental Theorem of Statistical Learning-in practice using the software packages introduced and taught in the course.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5780 - Bayesian Data Analysis: Principles and Practice (4 Credits)
Bayesian data analysis uses probability theory as a kind of calculus of inference, specifying how to quantify and propagate uncertainty in data-based chains of reasoning. Students will learn the fundamental principles of Bayesian data analysis, and how to apply them to varied data analysis problems across science and engineering. Topics include: basic probability theory, Bayes's theorem, linear and nonlinear models, hierarchical and graphical models, basic decision theory, and experimental design. There will be a strong computational component, using a high-level language such as R or Python, and a probabilistic language such as BUGS or Stan.
Prerequisites: BTRY 3080 and BTRY 3020/STSCI 3200, or equivalent.
Last Four Terms Offered: Spring 2022
Learning Outcomes:
- A basic understanding of the principles and foundations underlying the Bayesian approach.
- Practical experience using basic/intermediate Bayesian methods.
- Experience with widely-used tools and software development practices for producing and sharing collaborative, reproducible statistical research.
- Exposure to the Bayesian academic research literature.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5953 - MPS Career Management (0.5 Credits)
This course will focus on specific career and professional development topics for Master of Professional Studies in Applied Statistics students. Through lectures, workshops and seminars from MPS Statistics and Cornell Career Services staff and alumni, you will learn job search strategies, effective resume writing, professional dining etiquette, the art of effective presentation, as well as the opportunity for you to choose a minimum of three workshops to further develop career related strategies according to your personal needs.
Enrollment Information: Enrollment limited to: MPS Applied Statistics students.
Last Four Terms Offered: Fall 2023, Fall 2022, Fall 2021, Spring 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5954 - Project Development and Professional Communication (2 Credits)
Students will learn core professional skills required to work with a client. These skills include designing and implementing a work plan, learning to communicate efficiently, presentation of a final deliverable to the client, as well as learning good practices to conduct data analysis, documentation, quality control and collaboration. Through hands-on experience and real-world examples, students will develop a basic understanding of consulting, and be familiar with the professional standard expected in the industry.
Enrollment Information: Enrollment limited to: MPS Applied Statistics students.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022
Learning Outcomes:
- This course will equip MPS students with the tools and techniques to work with a client in their industry project.
- Students will learn how to manage a project, organize team structure, communicate and collaborate.
- Students will learn good practice to conduct data analysis and documentation.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5955 - Realtime Project Management (1 Credit)
This course is designed to run in parallel to STSCI 5999 (Applied Statistics MPS Data Analysis Project) and is a required part of the final-semester client project for MPS students. In this course students from each project team meet at regular intervals and briefly present their project milestones such as initiation of the project, data cleaning, analysis, etc. as well as raise any concern regarding any roadblock they are facing.
Prerequisites: STSCI 5954. Corequisite: STSCI 5999.
Enrollment Information: Primarily for: STSCI MPS students.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023
Learning Outcomes:
- Implementation - Learn to implement various phases of a project.
- Communication - Learn to provide quick team updates as done in industry.
- Time management - Learn to meet internal deadlines and ensure final project delivery.
- Learn to keep all project stakeholders informed about project progress so that any roadblock can be quickly resolved.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5980 - Tutorial in Actuarial Statistics (2 Credits)
Problem solving sessions to prepare students for the first four actuarial examinations (probability, financial mathematics, statistical modeling, and risk management).
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5990 - Directed Studies in Applied Statistics (1-4 Credits)
For individual or group research projects conducted under the direction of a member of STSCI faculty or instructors in a special area of statistical science that is not covered by regular course offerings.
Prerequisites: multicalculus, linear algebra, and basic statistics.
Last Four Terms Offered: Summer 2025, Spring 2025, Fall 2024, Summer 2024
Schedule of Classes (https://classes.cornell.edu/)

STSCI 5995 - Internship in Data Science (1 Credit)
Students planning internships related to Statistics and Data Science are encouraged to enroll in the departmental internship course.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Learning Outcomes:
- Demonstrate professional skills that pertain directly to the internship experience.
- Demonstrate verbal and written communication skills. Participate well as a team member and build professional network.
- Demonstrate effect management of personal behavior, ethics and attitudes.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 5999 - Applied Statistics MPS Data Analysis Project (4 Credits)
It is a long-term, in-depth statistical analysis of real-world dataset using various statistical methods and computer packages (such as SAS, R, SPSS, etc.). Students work in teams to solve business, managerial or scientific problems for clients. Projects are assigned by Department of Statistics and Data Science to teams. Each team has a dedicated faculty adviser, who supervises the projects and assigns grades. Grades are assigned individually to team members. In special cases teams may collaborate with groups in other Departments, e.g., M.Eng. project teams.
Enrollment Information: Enrollment limited to: MPS students in Applied Statistics.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 6020 - Statistical Methods II (4 Credits)
Crosslisted with BTRY 6020
Continuation of BTRY 6010. Emphasizes the use of multiple regression analysis, analysis of variance, and related techniques to analyze data in a variety of situations. Topics include an introduction to data collection techniques; least squares estimation; multiple regression; model selection techniques; detection of influential points, goodness-of-fit criteria; principles of experimental design; analysis of variance for a number of designs, including multi-way factorial, nested, and split plot designs; comparing two or more regression lines; and analysis of covariance. Emphasizes appropriate design of studies before data collection, and the appropriate application and interpretation of statistical techniques. Practical applications are implemented using a modern, widely available statistical package.
Prerequisites: BTRY 6010 or equivalent.
Enrollment Information: Enrollment limited to: graduate students or permission of instructor.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 6520 - Statistical Computing I (4 Credits)
Modern statistical methods routinely used in practice for analyzing large, complex datasets require intensive computation. This course covers topics in statistical computing, including numerical optimization and finding zeros (likelihood and related techniques), random number generation and Monte Carlo methods, bootstrap and subsampling, dimension reduction, nonlinear predictive modeling and parallel computing. Programming will be done in R.
Corequisites: ORIE 6700 or MATH 6730, or equivalent and at least one course in probability, or permission of instructor.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Spring 2021
Learning Outcomes:
  • Demonstrate knowledge of specialized statistical methods.
  • Use modern statistical software and a programming language.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 6521 - Statistical Computing II (3 Credits)
Last Four Terms Offered: Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 6610 - Spatial and Spatial-Temporal Data Analysis (3 Credits)
The course covers the analysis of spatial and/or spatial-temporal data, defined as data that include references to spatial locations or regions. Topics may include spatial interpolation, regression with spatial data, Gaussian process models, construction of covariance functions, computational techniques for large datasets, hierarchical non-Gaussian models, analysis of point patterns, forecasting, and Bayesian methods.
Prerequisites: STSCI 4030/5030 or equivalent plus STSCI 3080 or equivalent plus STSCI 4520/5520 or proficiency in R programming.
Last Four Terms Offered: Fall 2022
Learning Outcomes:
  • Identify different types of spatial and spatial-temporal data.
  • Select an appropriate model for the data.
  • Estimate model parameters using statistical software.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 6730 - Mathematical Statistics I (3 Credits)
Crosslisted with MATH 6730
This class will cover fundamental concepts in mathematical statistics, including both finite sample and asymptotic theory. Specific topics include: elements of risk optimality, Cramer-Rao-type bounds; M-estimation with an emphasis on Maximum Likelihood Estimation, asymptotic efficiency, asymptotic testing under fixed and local alternatives; multiple testing under FDR control; estimation in high dimensions and adaptation to sparsity, the analysis of Lasso-type estimators; elements of concentration inequalities.
Prerequisites: STSCI 4090/BTRY 4090, MATH 6710 or permission of instructor.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Schedule of Classes (https://classes.cornell.edu/)

STSCI 6740 - Mathematical Statistics II (3 Credits)
Crosslisted with MATH 6740
Focuses on the foundations of statistical inference, with an emphasis on asymptotic methods and the minimax optimality criterion. In the first part, the solution of the classical problem of justifying Fisher's information bound in regular statistical models will be presented. This solution will be obtained applying the concepts of contiguity, local asymptotic normality and asymptotic minimaxity. The second part will be devoted to nonparametric estimation, taking a Gaussian regression model as a paradigmatic example. Key topics are kernel estimation and local polynomial approximation, optimal rates of convergence at a point and in global norms, and adaptive estimation. Optional topics may include irregular statistical models, estimation of functionals and nonparametric hypothesis testing.
Prerequisites: MATH 6710 (measure theoretic probability) and STSCI 6730/MATH 6730, or permission of instructor.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 6750 - Probability II for Statistics (3 Credits)
This course will discuss several probabilistic results and techniques that are useful in mathematical statistics. Limit theorems and stochastic inequalities are two key components of the course. Topics to be covered include complements of basic limit theorems (Skorohod representation theorem, Berry-Esseen bound, Delta method), empirical distribution and quantile functions, U-statistics, bootstrap (and subsampling if time permits), stochastic convergence in metric spaces, and some elements of modern empirical process theory.
Prerequisites: MATH 6710, ORIE 6510, or equivalent.
Last Four Terms Offered: Spring 2025, Spring 2023, Spring 2021, Spring 2020
Schedule of Classes (https://classes.cornell.edu/)

STSCI 6780 - Bayesian Statistics and Data Analysis (3 Credits)
Crosslisted with ORIE 6780
Priors, posteriors, Bayes estimators, decision theory, asymptotic theory, Bayes factors, credible regions, hierarchical models, nonparametric Bayes, computational methods, Bayesian robustness, and applications. This is a theory-oriented course intended for PhD students. For a more applied introduction to Bayesian statistics, students should take STSCI 4780, Bayesian Data Analysis: Principles and Practice, which is offered in the spring.
Prerequisites: ORIE 6700 or an equivalent course in mathematical statistics. A basic knowledge of R is assumed.
Last Four Terms Offered: Spring 2025, Spring 2023, Spring 2022, Spring 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 6940 - Graduate Special Topics in Statistics (1-4 Credits)
Topics are arranged at the beginning of the semester for individual study or for group discussions. Or, students may elect to undertake a project in statistics. The work is supervised by a professor in this subject area.
Prerequisites: MATH 6710.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Spring 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 6970 - Individual Graduate Study in Statistics (3 Credits)
Individual tutorial study selected by the faculty. Because topics usually change from year to year, this course may be repeated for credit.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 7170 - Theory of Linear Models (3 Credits)
Crosslisted with ORIE 7170, ILRST 7170
Properties of the multivariate normal distribution. Distribution theory for quadratic forms. Properties of least squares and maximum likelihood estimates. Methods for fixed-effect models of less than full rank. Analysis of balanced and unbalanced mixed-effects models. Restricted maximum likelihood estimation. Some use of software packages and illustrative examples.
Prerequisites: BTRY 4090 or MATH 4720, BTRY 6020, or equivalents.
Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021
Schedule of Classes (https://classes.cornell.edu/)

STSCI 7180 - Generalized Linear Models (3 Credits)
A theoretical development of generalized linear models and related topics including categorical data problems, generalized additive models, and generalized linear mixed models.
Prerequisites: BTRY 6020, BTRY 4090 or MATH 4720 or equivalent.
Enrollment Information: Enrollment limited to: Ph.D. students in statistics.
Last Four Terms Offered: Spring 2022
Learning Outcomes:
- A deep understanding of the generalized linear model framework.
- The ability to perform inference in GLMs.
- A facility with the algorithms involved in GLMs.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 7951 - Advanced Statistical Consulting (2 Credits)
This course focuses on practical data analysis involving advanced statistical methods applied to real datasets. Students will be required to attend meetings with Cornell Statistical Consulting Unit (CSCU) clients. Emphasis will be placed on report writing, communicating effectively and explaining methodology and results to non-statisticians. Project lengths will vary based on the scope of the problems and could range from several short projects to a single semester-long project. Students will be required to collaborate in teams on some projects.
Prerequisites: STSCI 7170, STSCI 7180, STSCI 6520 or equivalents.
Enrollment Information: Primarily for: PhD students.
Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022
Learning Outcomes:
- Students will be able to integrate the statistical knowledge gained in courses and apply them to real life problems.
- Students will be able to learn to communicate effectively with clients to gather information needed to make the link between research questions and statistical methods
- Students will be able to research the application of statistical methodologies that are useful to clients and explain them to an audience of non-statisticians.

Schedule of Classes (https://classes.cornell.edu/)

STSCI 7999 - Graduate Level Dissertation Research (1-9 Credits)
Research at the Ph.D. Level.
Enrollment Information: Enrollment limited to: Ph.D. candidates by permission of Graduate Field Member concerned.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)

STSCI 9999 - Doctoral Level Dissertation Research (1-9 Credits)
Doctoral Level Dissertation Research.
Last Four Terms Offered: Spring 2025, Fall 2024, Spring 2024, Fall 2023
Schedule of Classes (https://classes.cornell.edu/)