STATISTICS AND DATA SCIENCE

Degree Types: PhD in Statistics and Data Science, MS in Statistics and Data Science, Ad Hoc MS in Applied Statistics, BA/MS Combined Degree

The Doctoral Program in Statistics and Data Science (https://catalogs.northwestern.edu/tgs/statistics/statistics-phd/) provides students with comprehensive training in statistics and data science theory, methodology, and the application of these methods to problems in a wide range of fields.

Faculty have specialties in diverse areas including statistical machine learning, experimental design, linear models, sample surveys, statistical theory, computational Bayesian inference, bioinformatics and computational biology, and the analysis of qualitative data. They have interest and broad experience in the application of statistics to the biomedical sciences, social sciences, law, and public policy.

Students can tailor the program to meet academic interests and career goals, and cross-disciplinary work is encouraged. The program prepares students for careers as university teachers and researchers and as research statisticians in industry, government, and the non-profit sector.

The Department also has a Master's Program in Statistics and Data Science (https://catalogs.northwestern.edu/tgs/statistics/statistics-ms/), for students interested in earning a professional degree or who want to prepare for doctoral study in statistics, data science, or another field with a connection to statistics.

The Ad Hoc MS in Applied Statistics (https://catalogs.northwestern.edu/tgs/statistics/applied-statistics-ms/) program is for students in Northwestern PhD programs who would like to earn a MS in Applied Statistics, along with their PhD. This MS is awarded simultaneously with the PhD.

Additional resources:

- Department website (https://www.statistics.northwestern.edu/)
- Program handbook(s)

Degrees Offered

- Statistics Bachelor's/MS (https://catalogs.northwestern.edu/tgs/statistics/statistics-bach-mast/)
- Applied Statistics MS (Ad Hoc Combined Degree) (https://catalogs.northwestern.edu/tgs/statistics/applied-statistics-ms/)
- Statistics and Data Science MS (https://catalogs.northwestern.edu/tgs/statistics/statistics-ms/)
- Statistics and Data Science PhD (https://catalogs.northwestern.edu/tgs/statistics/statistics-phd/)

Statistics and Data Science: MS

Learning objective(s)/Students should be able to…

- Understand fundamental, key concepts of statistics and data science theory and methodology.
- Develop technical skills in statistical modeling, inference, and data science for practical applications.
- Prepare for a career as a professional statistician in industry, government, or the non-profit sector, or for doctoral study in statistics, data science, or related field.

Statistics and Data Science: PHD

Learning objective(s)/Students should be able to…

- Understand fundamental, key concepts of statistical theory, methods, and data science. Be able to apply them in practice.
- Understand a variety of advanced statistical topics and their applications.
- Contribute original research to scholarly community. Articulate research results and their impacts.
- Teach statistics to undergraduates.
- Create and communicate professional development plan.

Statistics Courses

STAT 301-1 Data Science 1 with R (1 Unit)
First course in Data Science with a focus on data management, manipulation, and visualization skills and techniques for exploratory data analysis. The course also introduces the R programming language. Students may not receive credit for both this course and STAT 303-1.
Prerequisite: STAT 201-0 or GEN_ENG 150-0 or GEN_ENG 151-0 and STAT 202-0 or STAT 210-0 or STAT 232-0 or PSYCH 201-0 or IEMS 201-0 or IEMS 303-0 or equivalent.
SoC Mathematics, Science, and Technology Distro

STAT 301-2 Data Science 2 with R (1 Unit)
Introduction to supervised machine/statistical learning with a focus on application using R. Course covers essential concepts in machine learning while surveying standard machine learning models such as linear and logistic regression. Course provides a foundation for learning more machine learning methods. Students may not receive credit for both this course and STAT 303-2.
Prerequisite: STAT 301-1.
SoC Mathematics, Science, and Technology Distro

STAT 301-3 Data Science 3 with R (1 Unit)
An intermediate course that covers machine learning methods in R, including supervised and unsupervised learning. It provides the knowledge and skills necessary to tackle real world problems with machine learning. Students may not receive credit for both this course and STAT 303-3.
Prerequisite: STAT 301-2.
SoC Mathematics, Science, and Technology Distro

STAT 302-0 Data Visualization (1 Unit)
Introduction to the knowledge, skills, and tools required to visualize data of various formats across statistical domains and to create quality visualizations for both data exploration and presentation.
Prerequisite: STAT 201-0 or GEN_ENG 150-0 or GEN_ENG 151-0 and STAT 202-0 or STAT 210-0 or STAT 232-0 or PSYCH 201-0 or IEMS 201-0 or IEMS 303-0 or equivalent.
SoC Humanities and Fine Arts Distro SoC Mathematics, Science, and Technology Distro

STAT 303-1 Data Science 1 with Python (1 Unit)
First course in Data Science, with focus on data management, manipulation, and visualization skills and techniques for exploratory data analysis. The course also introduces the Python programming language in the context of Data Science. Students may not receive credit for both this course and STAT 301-1.

Prerequisite: STAT 201-0 or GEN_ENG 150-0 or GEN_ENG 151-0 and STAT 202-0 or STAT 210-0 or STAT 232-0 or PSYCH 201-0 or IEMS 201-0 or IEMS 303-0 or equivalent.
SoC Mathematics, Science, and Technology Distro

STAT 303-2 Data Science 2 with Python (1 Unit)
Introduction to supervised machine/statistical learning with a focus on application using Python. Course covers essential concepts in machine learning while surveying standard machine learning models such as linear and logistic regression. Course provides a foundation for learning more machine learning methods. Students may not receive credit for both this course and STAT 301-2.
Prerequisite: STAT 303-1.
SoC Mathematics, Science, and Technology Distro

STAT 303-3 Data Science 3 with Python (1 Unit)
An intermediate course that covers machine learning methods in Python, including supervised and unsupervised learning. It provides the knowledge and skills necessary to tackle real world problems with machine learning. Students may not receive credit for both this course and STAT 301-3.
Prerequisite: STAT 303-2.
SoC Mathematics, Science, and Technology Distro

STAT 304-0 Introduction to Data Structures and Algorithms (1 Unit)
This course will introduce students to the design, implementation, analysis, and proper application of abstract data types, data structures, and their algorithms. Python will be used to implement and explore various algorithms and data structures. Students should be prepared for a significant amount of hands-on programming.
Prerequisite: STAT 201-0 or GEN_ENG 150-0 or GEN_ENG 151-0 and STAT 202-0 or STAT 210-0 or STAT 232-0 or PSYCH 201-0 or IEMS 201-0 or IEMS 303-0 or equivalent.
SoC Mathematics, Science, and Technology Distro

STAT 305-0 Information Management for Data Science (1 Unit)
This course aims to give students an extensive data processing and visualization skillset using various Python libraries. It will also focus on relational databases and queries in SQL. Students will learn data scraping from online sources and mobile applications as well as a brief introduction to statistical and predictive analysis after the data is clean and ready to use.
Prerequisite: STAT 201-0 or GEN_ENG 150-0 or GEN_ENG 151-0 and STAT 202-0 or STAT 210-0 or STAT 232-0 or PSYCH 201-0 or IEMS 201-0 or IEMS 303-0 or equivalent.
SoC Mathematics, Science, and Technology Distro

STAT 320-1 Statistical Theory & Methods 1 (1 Unit)
Sample spaces, computing probabilities, random variables, distribution functions, expected values, variance, correlation, limit theory. May not receive credit for both STAT 320-1 and any of STAT 383-0, MATH 310-1, MATH 311-1, MATH 314-0, MATH 385-0, ELEC_ENG 302-0, or IEMS 302-0.
Co-requisites: STAT 202-0 or STAT 210-0 or STAT 232-0 or PSYCH 201-0 or IEMS 201-0 or IEMS 303-0 or equivalent, STAT 228-0 or MATH 235-0 or both MATH 226-0 and MATH 230-2 or MATH 228-2 or MATH 281-2 or MATH 285-3 or MATH 290-3 or MATH 291-3 or ES_APPM 252-2.
SoC Mathematics, Science, and Technology Distro

STAT 320-2 Statistical Theory & Methods 2 (1 Unit)
Moment generating functions, parameter estimation, confidence intervals, hypothesis tests.
Prerequisite: STAT 320-1 or MATH 310-1 or MATH 310-1 or MATH 311-1 or MATH 314-0 or MATH 385-0 or ELEC_ENG 302-0 or IEMS 302-0 or STAT 383-0.
SoC Mathematics, Science, and Technology Distro

STAT 320-3 Statistical Theory & Methods 3 (1 Unit)
Comparison of parameters, goodness-of-fit tests, regression analysis, analysis of variance, and nonparametric methods.
Prerequisites: STAT 320-2 and MATH 240-0 or MATH 285-1 or MATH 281-3 or MATH 290-1or MATH 291-1 or GEN_ENG 205-1 or GEN_ENG 206-1.
SoC Mathematics, Science, and Technology Distro

STAT 325-0 Survey Sampling (1 Unit)
Probability sampling, simple random sampling, error estimation, sample size, stratification, systematic sampling, replication methods, ratio and regression estimation, cluster sampling.
Prerequisites: MATH 230-1 and 2 quarters of statistics, or consent of instructor.
SoC Individual and Social Behavior Distro SoC Mathematics, Science, and Technology Distro

STAT 328-0 Causal Inference (1 Unit)
Introduction to modern statistical thinking about causal inference. Topics include completely randomized experiments, confounding, ignorability of assignment mechanisms, matching, observational studies, noncompliance, and Bayesian methods.
Prerequisites: STAT 320-2 and STAT 350-0 or ECON 381-2 or MATH 386-2.
SoC Individual and Social Behavior Distro SoC Mathematics, Science, and Technology Distro

STAT 330-1 Applied Statistics for Research 1 (1 Unit)
First Quarter: Design of experiments and surveys, numerical summaries of data, graphical summaries of data, correlation and regression, probability, sample mean, sample proportion, confidence intervals and tests of significance, one and two sample problems, ANOVA. Second Quarter: Simple linear regression, inference, diagnostics, multiple regression diagnostics, autocorrelation, 1-way ANOVA, power and sample size determination, 2-way ANOVA, ANCOVA, randomized block designs.
SoC Mathematics, Science, and Technology Distro

STAT 344-0 Statistical Computing (1 Unit)
Exploration of theory and practice of computational statistics with emphasis on statistical programming in R.
Prerequisite: STAT 320-2 or equivalent. Some R programming experience is desired.
SoC Mathematics, Science, and Technology Distro

STAT 348-0 Applied Multivariate Analysis (1 Unit)
Statistical methods for describing and analyzing multivariate data. Principal component analysis, factor analysis, canonical correlation, clustering. Emphasis on statistical and geometric motivation, practical application, and interpretation of results.
Prerequisites: STAT 320-2 and MATH 240-0 or MATH 285-1 or MATH 281-3 or MATH 290-1 or MATH 291-1 or GEN_ENG 205-1 or GEN_ENG 206-1. Co-requisite STAT 350-0. The course uses R extensively; hence, some experience with R will be useful.
SoC Mathematics, Science, and Technology Distro

STAT 350-0 Regression Analysis (1 Unit)
Simple linear regression and correlation, multiple regression, residual analysis, model building, variable selection, multi-collinearity and shrinkage estimation, nonlinear regression.
Prerequisites: STAT 201-0 or GEN_ENG 150-0 or GEN_ENG 151-0 and STAT 202-0 or STAT 210-0 or STAT 232-0 or PSYCH 201-0 or IEMS 201-0 or IEMS 303-0. Co-requisite: STAT 320-1 or STAT 383-0 or MATH 310-1 or MATH 311-1 or MATH 314-0 or MATH 385-0 or ELEC_ENG 302-0 or IEMS 302-0.
SoC Mathematics, Science, and Technology Distro

STAT 351-0 Design and Analysis of Experiments (1 Unit)

Methods of designing experiments and analyzing data obtained from them: one-way and two-way layouts, incomplete block designs, factorial designs, random effects, split-plot and nested designs.
Prerequisite: STAT 320-1 or STAT 383-0 or MATH 310-1 or MATH 311-1 or MATH 314-0 or MATH 385-0 or ELEC_ENG 302-0 or IEMS 302-0 or equivalent.
SoC Mathematics, Science, and Technology Distro

STAT 352-0 Nonparametric Statistical Methods (1 Unit)
Survey of nonparametric methods, with emphasis on understanding their application. Estimation of a distribution function, density estimation, and nonparametric regression.
Prerequisite: STAT 350-0 or ECON 381-2 or MATH 386-2.
SoC Mathematics, Science, and Technology Distro

STAT 353-0 Advanced Regression (1 Unit)
This course covers modern regression methods, including: (1) generalized linear models (binary, categorical, and count data), (2) random effects, mixed effects, and nonlinear models, and (3) model selection. The course emphasizes both the theoretical development of the methods, as well as their application, including the communication of models and results both verbally and in writing.
Prerequisites: STAT 350-0 or ECON 381-2 or MATH 386-2 and STAT 320-2 or STAT 420-2 or MATH 310-2.
SoC Humanities and Fine Arts Distro SoC Mathematics, Science, and Technology Distro

STAT 354-0 Time Series Modeling (1 Unit)
Introduction to modern time series analysis. Autocorrelation, time series regression and forecasting, ARIMA and GARCH models.
Prerequisites: STAT 320-1 or STAT 383-0 or MATH 310-1 or MATH 311-1 or MATH 314-0 or MATH 385-0 or ELEC_ENG 302-0 or IEMS 302-0.
Corequisite: STAT 350-0 or ECON 381-2 or MATH 386-2.
SoC Mathematics, Science, and Technology Distro

STAT 355-0 Analysis of Qualitative Data (1 Unit)
Introduction to the analysis of qualitative data. Measures of association, loglinear models, logits, and probits.
Prerequisite: STAT 320-2 or equivalent.
SoC Individual and Social Behavior Distro SoC Mathematics, Science, and Technology Distro

STAT 356-0 Hierarchical Linear Models (1 Unit)
Introduction to the theory and application of hierarchical linear models. Two and three level linear models, hierarchical generalized linear models, and application of hierarchical models to organizational research and growth models.
Prerequisites: STAT 320-2 and STAT 350-0 or ECON 381-2 or MATH 386-2.
SoC Individual and Social Behavior Distro SoC Mathematics, Science, and Technology Distro

STAT 357-0 Introduction to Bayesian Statistics (1 Unit)
Introduction to basic concepts and principles in Bayesian inference such as the prior, likelihood, posterior and predictive distributions, as well as an introduction to a variety of computational algorithms for Bayesian inference. Students learn how to develop, describe, implement and critique statistical models from a Bayesian perspective.
Prerequisites: STAT 320-2 and STAT 301-2 or STAT 350-0 or ECON 381-2 or MATH 386-2 or consent of instructor.
SoC Mathematics, Science, and Technology Distro

STAT 359-0 Topics in Statistics (1 Unit)
Topics in theoretical and applied statistics to be chosen by instructor.
Prerequisite: varies by course topic.
SOC Additional Distro

STAT 360-0 Introduction to Generative AI (1 Unit)
This course will provide an introduction to generative AI. In particular, we will cover large language models and diffusion models. By the end of the course, students should have a thorough understanding of all major components underpinning modern large language models. Students should be able to train their own large language models after taking this class.
Prerequisites: Linear algebra (MATH 240-0), STAT 320-2 or equivalent, some familiarity with deep learning, Python experience.
SoC Mathematics, Science, and Technology Distro

STAT 363-0 Deep Learning: Foundation, Methods, and Applications (1 Unit)
This course introduces state-of-the-art deep learning architectures and their applications across domains such as fraud detection, speech recognition, medical imaging, and autonomous systems. Students will learn the theoretical foundations of these models, understand their distinct purposes, and gain practical experience applying them to real-world problems, while also engaging with recent advances in deep learning research.
Prerequisites: STAT 362-0 or COMP_SCI 349-0 or STAT 415-0; and MATH 240-0 or MATH 281-3 or MATH 285-1 or MATH 290-1 or MATH 291-1 or GEN_ENG 205-1 or GEN_ENG 206-1; and STAT 320-1 or STAT 383-0 or MATH 310-1 or MATH 311-1 or MATH 314-0 or MATH 385-0 or ELEC_ENG 302-0 or IEMS 302-0. Programming experience in Python is expected. Students with comparable preparation may enroll with permission of the instructor.
SoC Mathematics, Science, and Technology Distro

STAT 364-0 Large Language Models: Foundations, Training, and Inference (1 Unit)
Introduction to the foundations and applications of large language models. Topics include tokenization, embeddings, attention, transformers, pretraining, fine-tuning, prompting, inference strategies, evaluation, and responsible use of LLMs. Students gain hands-on experience implementing and analyzing language models through coding assignments and a course project.
Prerequisites: STAT 362-0, COMP_SCI 349-0, or STAT 415-0; and MATH 240-0, MATH 281-3, MATH 285-1, MATH 290-1, MATH 291-1, GEN_ENG 205-1, or GEN_ENG 206-1; and STAT 320-1, STAT 383-0, MATH 310-1, MATH 311-1, MATH 314-0, MATH 385-0, ELEC_ENG 302-0, IEMS 302-0, or equivalent. Programming experience in Python is expected. Students with comparable preparation may enroll with permission of the instructor.
SoC Individual and Social Behavior Distro SoC Mathematics, Science, and Technology Distro

STAT 365-0 Introduction to the Analysis of Financial Data (1 Unit)
Statistical methods for analyzing financial data. Models for asset returns, portfolio theory, parameter estimation. The statistical software R is used.
Prerequisites: STAT 350-0 or ECON 381-2 and MATH 240-0 or MATH 285-1 or MATH 281-3 or MATH 290-1or MATH 291-1 or GEN_ENG 205-1 or GEN_ENG 206-1.
SoC Individual and Social Behavior Distro SoC Mathematics, Science, and Technology Distro

STAT 415-0 Introduction to Machine Learning (1 Unit)
This course is for students doing advanced studies in statistics and certain other fields will provide an introduction to modern machine learning methods. Topics include supervised learning, sparsity, logistic regression, SVM, kernel methods, deep learning, unsupervised learning, and real world problems including fairness and interpretability of black box models. Not for data science majors/minors - students studying data science should take STAT 362 instead.

Prerequisites: MATH 240-0 and MATH 230-2 and STAT 320-2 or Statistics and Data Science graduate standing.

STAT 420-1 Introduction to Statistical Theory & Methodology-1 (1 Unit)
Distribution theory, characteristic functions, moments and cumulants, random variables, sampling theory, and common statistical distributions.

STAT 420-2 Introduction to Statistical Theory & Methodology-2 (1 Unit)
Methods of estimation, hypothesis tests, confidence intervals, least squares, likelihood methods, and large-sample methods. Third Quarter: Normal linear models and its various extensions.

STAT 420-3 Introduction to Statistical Theory & Methodology-3 (1 Unit)
Normal linear models and its various extensions.

STAT 430-1 Probability for Statistical Inference 1 (1 Unit)
Foundations of measure theoretic probability, with applications to statistics.
Prerequisites: MATH 320-1 and STAT 420-1.

STAT 430-2 Probability for Statistical Inference 2 (1 Unit)
A second course in measure-theoretic probability, with an eye towards statistics. Topics include Markov chains, conditional expectation, martingales, Poisson processes, Brownian motion, and selected advanced topics, together with statistical applications.
Prerequisite: STAT 430-1 or permission of instructor.

STAT 435-0 Mathematical Foundations of Machine Learning (1 Unit)
In this course, students are expected to explore some mathematical foundations of modern machine learning under a problem-solving framework. Topics include probability theory, frequentist statistics, Bayesian statistics, tensor algebra, vector calculus, convex and stochastic optimization, stochastic processes and sampling, sequential optimization and dynamic programming. This class strongly emphasizes on developing problem-solving skills.
Prerequisite: 420-1(recommended but not required).

STAT 436-0 Reinforcement Learning (1 Unit)
The first half of the course will cover classical reinforcement learning concepts: MDPs, tree search, trajectory optimization, value iteration, policy iteration, and SARSA. The second half will cover modern deep reinforcement learning: deep Q learning, policy gradients, Monte Carlo tree search, inverse reinforcement learning, noise contrastive estimation. We will conclude by discussing philosophical questions related to AI and superhuman intelligence.
Prerequisite: Linear algebra (MATH 240-0), some familiarity with Stochastic Gradient Descent, some Python experience, prior experience with PyTorch helpful but not required.

STAT 439-0 Meta-Analysis (1 Unit)
Statistical methods for combining results of multiple studies. Effect size indexes and their estimators, combined estimation and test of heterogeneity, modeling between-study variation in effect sizes, models for publication selection.
Prerequisite: STAT 350-0 or a similar course in regression.

STAT 440-0 Applied Probability for Statistics (1 Unit)
In this course, students are expected to learn typical statistical applications of probability and stochastic processes. Topics may include iid processes (e.g., with applications in asymptotic statistics), Markov chains (e.g. with applications to MCMC convergence theory), martingales (e.g. with applications to survival analysis), Brownian motion (e.g. with applications to clinical trials / optimal stopping; Black-Scholes), and other statistical applications. An integral part will be the student presentations on related topics.

Pre-Requisites: Real analysis, Probability with measure (Stat 430-1,2), Statistics 420-1,2,3, some knowledge of functional analysis may also be helpful.

STAT 450-1 Advanced Statistical Theory I (1 Unit)
This is a first part of a Ph.D. course in theoretical statistics. We will cover a selection of modern topics in mathematical statistics, with a focus on high-dimensional statistical models and non- parametric statistical models. One of the main goals of this course is to provide the student with some theoretical background and mathematical tools to read and understand the current statistical literature on high-dimensional models.
Prerequisite: Linear algebra, real analysis, probability theory, statistical inference.

STAT 450-2 Advanced Statistical Theory II (1 Unit)
This is a second part of a Ph.D. course in theoretical statistics. We will cover a selection of modern topics in mathematical statistics, with a focus on high-dimensional statistical models and non-parametric statistical models. One of the main goals of this course is to provide you with some theoretical background and mathematical tools to read and understand the current statistical literature on high-dimensional models.
Prerequisites: STAT 450-1 Advanced Statistical Theory 1.

STAT 455-0 Advanced Qualitative Data Analysis (1 Unit)
Probit, logit, log-linear, and latent-class models. Multi-dimensional contingency tables; polytomous responses with continuous independent variables.
Prerequisites: STAT 350-0 and STAT 420-3.

STAT 456-0 Generalized Linear Models (1 Unit)
Inference and fitting of generalized linear models with application to classical linear models, binomial and multinomial logit models, log-linear models, Cox's proportional hazards model and GEE's for longitudinal data.
Prerequisites: STAT 350-0 and STAT 420-3.

STAT 457-0 Applied Bayesian Inference (1 Unit)
Introduction to computational algorithms for Bayesian inference. Observed data and data augmentation methods are considered in detail. Methods are illustrated with real examples.
Prerequisites: STAT 350-0 and STAT 420-3 or equivalent or students who have earned a Master's degree in Statistics or permission of the instructor.

STAT 461-0 Advanced Topics in Statistics (1 Unit)

STAT 465-0 Statistical Methods for Bioinformatics and Computational Biology (1 Unit)
An introduction of statistical methodologies in cutting-edge fields of computational biology and bioinformatics topics including gene expression data analysis; biological sequence analysis.

STAT 499-0 Independent Study (1-3 Units)
SEE DEPT FOR SECTION AND PERMISSION NUMBERS.

STAT 519-0 Responsible Conduct of Research Training (0 Unit)

STAT 590-0 Research (1-3 Units)
SEE DEPT FOR SECTION AND PERMISSION NUMBERS.

STAT 595-0 Internship (0 Unit)
It is an internship program under Curricular Practical Training (CPT), where students would do an unpaid or paid internship on campus or in a non-NU company.