STATISTICS AND DATA SCIENCE

statistics.northwestern.edu

Statistics and Data Science are closely related scientific disciplines that deal with the collection, organization, analysis, interpretation, and reporting of data. As data becomes more abundant and readily accessible, the need for methods and techniques for extracting information from data has greatly increased. The wide range of applications of Statistics and Data Science methods include finance, engineering, medicine, sports, law, and biological, social, and physical sciences. Indeed, it is hard to think of any discipline nowadays that does not call upon the use of statistical methods and approaches.

Statistical methods are widely used in observational studies and for the design and analysis of experiments, sample surveys, and censuses. Such analysis involves diverse fields as clinical trials, political polling, actuarial science, and the design of financial instruments.

Data Science methods are widely used in settings with large amounts of data with a focus on computer analysis, efficiency in terms of both compute time and memory demands, and prediction in aid of decision-making. Entire new fields based on these methods have sprung up such as deep learning, artificial intelligence, and bioinformatics.

Programs of Study

- Data Science Major (https://catalogs.northwestern.edu/undergraduate/arts-sciences/statistics-data-science/data-science-major/)
- Data Science Minor (https://catalogs.northwestern.edu/undergraduate/arts-sciences/statistics-data-science/data-science-minor/)
- Statistics Major (https://catalogs.northwestern.edu/undergraduate/arts-sciences/statistics-data-science/statistics-major/)
- Statistics Minor (https://catalogs.northwestern.edu/undergraduate/arts-sciences/statistics-data-science/statistics-minor/)

STAT 101-7 College Seminar (1 Unit) Small, writing and discussion-oriented course exploring a specific topic or theme, and introducing skills necessary to thriving at Northwestern. Not eligible to be applied towards a WCAS major or minor except where specifically indicated.

STAT 101-8 First-Year Writing Seminar (1 Unit) Small, writing and discussion-oriented course exploring a specific topic or theme, and focused on the fundamentals of effective, college-level written communication. Not eligible to be applied towards a WCAS major or minor except where specifically indicated.

STAT 201-0 Introduction to Programming for Data Science (1 Unit) This course is an introduction programming for Data Science. It will prepare students to use essential programming methods as implemented in either Python or R as a tool in the subsequent data science courses including STAT 301-1, STAT 301-2, STAT 301-3, STAT 303-1, STAT 303-2, STAT 303-3, STAT 304-0, STAT 305-0, STAT 362-0, and STAT 390-0, etc. *Empirical and Deductive Reasoning Foundational Dis Formal Studies Distro Area*

STAT 202-0 Introduction to Statistics and Data Science (1 Unit) Data collection, summarization, correlation, regression, sampling, confidence intervals, tests of significance. Introduction to data analysis techniques using R programming, no prior programming experience required. Does not require calculus and makes minimal use of mathematics. May not receive credit for both STAT 202-0 and STAT 210-0. *Empirical and Deductive Reasoning Foundational Dis Formal Studies Distro Area*

STAT 202-SG Peer-Guided Study Group: Introduction to Statistics and Data Science (0 Unit) Peer-guided study group for students enrolled in STAT 202-0. Meets weekly in small groups, along with a peer facilitator, to collaboratively review material, work through practice problems, and clarify course concepts. Enrollment optional. Graded S/U.

STAT 210-0 Introduction to Probability and Statistics (1 Unit) A mathematical introduction to probability theory and statistical methods, including properties of probability distributions, sampling distributions, estimation, confidence intervals, and hypothesis testing. STAT 210-0 is primarily intended for economics majors. May not receive credit for both STAT 202-0 and STAT 210-0. Prerequisite: strong background in high school algebra (calculus is not required). *Empirical and Deductive Reasoning Foundational Dis Formal Studies Distro Area*

STAT 210-SG Peer-Guided Study Group: Introduction to Probability and Statistics (0 Unit) Peer-guided study group for students enrolled in STAT 210-0. Meets weekly in small groups, along with a peer facilitator, to collaboratively review material, work through practice problems, and clarify course concepts. Enrollment optional. Graded S/U.

STAT 228-0 Series and Multiple Integrals (1 Unit) Sequences and series, and convergence tests. Power series, Taylor polynomials and error. Double integrals, triple integrals, and change of variables. Students may receive credit for only one of MATH 235#0, MATH 226#0, or STAT 228#0. Prerequisite: MATH 218#3 or MATH 220#2, and MATH 228#1 or MATH 230#1 or MATH 281#1 or MATH 285#2 or MATH 290#2 or MATH 291#2 or ES_APPM 252#1. *Empirical and Deductive Reasoning Foundational Dis Formal Studies Distro Area*

STAT 232-0 Applied Statistics (1 Unit) Basic concepts of using statistical models to draw conclusions from experimental and survey data. Topics include simple linear regression, multiple regression, analysis of variance, and analysis of covariance. Practical application of the methods and the interpretation of the results will be emphasized. Prerequisites: STAT 202-0, STAT 210-0, or equivalent; MATH 220-1. *Formal Studies Distro Area*

STAT 301-1 Data Science 1 with R (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 R programming language in the context of Data Science. Students may not receive credit for both this course and STAT 303-1.
Prerequisite: STAT 202-0 or STAT 210-0 or consent of the instructor.
Formal Studies Distro Area

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 or consent of instructor.
Formal Studies Distro Area

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 or consent of the instructor.
Formal Studies Distro Area

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 202-0 or equivalent.
Formal Studies Distro Area

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 202-0 or STAT 210-0 or consent of the instructor.
Formal Studies Distro Area

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 or consent of the instructor.
Formal Studies Distro Area

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 or consent of the instructor.
Formal Studies Distro Area

STAT 304-0 Data Structures and Algorithms for Data Science (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.
Prerequisites: STAT 202-0 or STAT 210-0 or STAT 232-0, and COMP_SCI 110-0 or COMP_SCI 111-0.
Formal Studies Distro Area

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.
Prerequisites: STAT 202-0 or STAT 210-0 or STAT 232-0, and COMP_SCI 110-0 or COMP_SCI 111-0.
Formal Studies Distro Area

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 202-0.
Co-requisites: STAT 202-0 or STAT 210-0, and STAT 228-0 or MATH 235-0 or both MATH 226-0 and MATH 230-2.
Formal Studies Distro Area

STAT 320-2 Statistical Theory & Methods 2 (1 Unit)
Parameter estimation, confidence intervals, hypothesis tests.
Prerequisite: STAT 320-1 or MATH 310-1.
Formal Studies Distro Area

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, MATH 240-0.
Formal Studies Distro Area

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.
Formal Studies Distro Area

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, STAT 350-0.
Formal Studies Distro Area

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.

STAT 330-2 Applied Statistics for Research 2 (1 Unit)
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.

STAT 332-0 Statistics for Life Sciences (1 Unit) Application of statistical methods and data analysis techniques to the life sciences. Parametric statistics, nonparametric approaches, resampling-based approaches.
Prerequisite: 1 introductory statistics course. *Formal Studies Distro Area*

STAT 342-0 Statistical Data Mining (1 Unit)
Methods for modeling binary responses with multiple explanatory variables. Potential topics include statistical decision theory, binary regression models, cluster analysis, probabilistic conditional independence, and graphical models.
Prerequisites: courses in probability and statistics comparable to STAT 320-1, STAT 320-2; a course in multiple regression comparable to STAT 350-0; familiarity with statistical computing software such as MINITAB or SPSS.
Formal Studies Distro Area

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.
Formal Studies Distro Area

STAT 345-0 Statistical Demography (1 Unit)

Introduction to statistical theory of demographic rates (births, deaths, migration) in multistate setting; statistical models underlying formal demography; analysis of error in demographic forecasting.
Prerequisite: STAT 350-0, MATH 240-0, or equivalent.
Formal Studies Distro Area

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, MATH 240-0, and STAT 350-0.
Formal Studies Distro Area

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. Prerequisite STAT 202-0 or STAT 210-0 or STAT 232-0 or PSYCH 201 or IEMS 201 or IEMS 303. 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 202-0.
Formal Studies Distro Area

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 equivalent.
Formal Studies Distro Area

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.
Formal Studies Distro Area

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 320-2 or 420-2 or MATH 310-2 and a first course in regression is required at the level of STAT 350-0.
Formal Studies Distro Area

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. Corequisite: STAT 350-0.
Formal Studies Distro Area

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.
Formal Studies Distro Area

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, STAT 350-0.
Formal Studies Distro Area

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 320-1, STAT 320-2, STAT 301-2 or 350-0, or consent of instructor.
Formal Studies Distro Area

STAT 359-0 Topics in Statistics (1 Unit)
Topics in theoretical and applied statistics to be chosen by instructor.
Prerequisite: consent of instructor.
Formal Studies Distro Area

STAT 362-0 Advanced Machine Learning for Data Science (1 Unit)
This course aims to focus on the theory and applications of advanced Machine Learning (ML) and Deep Learning (DL) topics. It also includes an introduction to Bayesian Modeling and Reinforcement Learning (RL). The students are expected to have a basic understanding of ML from STAT 301-1-2-3/303-1-2-3. The coding language for the homework projects is Python. Prerequisites: STAT 301-3 or STAT 303-3. Co-requisite MATH 240-0. *Formal Studies Distro Area*

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.
Prerequisites: STAT 320-3, MATH 240-0.
Formal Studies Distro Area

STAT 370-0 Human Rights Statistics (1 Unit)
Development, analysis, interpretation, use, and misuse of statistical data and methods for description, evaluation, and political action regarding war, disappearances, justice, violence against women, trafficking, profiling, elections, hunger, refugees, discrimination, etc.
Prerequisites: Two of STAT 325-0, STAT 350-0, STAT 320-2, STAT 320-3; or ECON 381-1, ECON 381-2; or MATH 386-1, MATH 386-2; or IEMS 303-0, IEMS 304-0.
Formal Studies Distro Area

STAT 383-0 Probability and Statistics for ISP (1 Unit) Probability and statistics. Ordinarily taken only by students in ISP; permission required otherwise. May not receive credit for both STAT 383-0 and any of STAT 320-1; MATH 310-1, MATH 311-1, MATH 314-0, MATH 385-0; ELEC_ENG 302-0; or IEMS 202-0. Prerequisites: MATH 281-1, MATH 281-2, MATH 281-3; PHYSICS 125-1, PHYSICS 125-2, PHYSICS 125-3. *Formal Studies Distro Area*

STAT 390-0 Data Science Project (1 Unit) An opportunity to develop and create solutions for stakeholders with data needs. Students will work in teams to appropriately scope and solve data problems. Students should expect to spend significant amounts of time coordinating and working with team mates outside of class. Prerequisites: STAT 301-3 or STAT 303-3 or consent of instructor.

STAT 398-0 Undergraduate Seminar (1 Unit)

STAT 399-0 Independent Study (1-3 Units) Independent work under the guidance of a faculty member. Consent of department required.