

DATA ANALYTICS -DAT

DAT 111 Introduction to Reporting and Analysis 3 Credits

An introduction to the methods and tools for reporting quantitative data for decision support in a wide range of fields. This course is meant as an introductory course in the Data Science program, and for students in other disciplines preparing for decision support roles in a range of commercial, educational or research roles. Both the general theories and approaches to the presentation of data for decision support in tabular and graphic forms, and practical technical methods will be covered in the course. Most of the course time will be spent using Excel for these tasks, but Tableau and/or PowerBI as well as some basic SQL queries will also be covered. Whenever possible, "real-world" data drawn from a wide range of fields and disciplines will be used to illustrate problems and approaches to reporting of data.

Fulfills College Core: Field 7 (Mathematical Sciences)

Offered: every spring.

DAT 211 Advanced Statistics with R 3 Credits

This course is designed to introduce students to the programming language R. We will begin by talking about the benefits of R from a practical to an ethical level. Students will learn to install R and load packages. Students will then identify a data set they want to work with over the semester and preregister their project rationale, hypotheses, and analytic plan with OSF. Students will spend the majority of their time learning to execute their analytic plan in R. Students will present their project at Ignatian Scholarship Day. After their ISD presentation, students will archive their materials on OSF and update their preregistration to reflect any modifications made to the plan as they conducted their research, changes they would make if they were going to do the project again, and future analyses they would like to conduct with the data set.

Offered: once a year.

DAT 411 Econometrics 3 Credits

Econometrics is the science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena. Econometric modeling is an important research tool in Economics, Finance, and many other academic disciplines. The goal of this course is to provide you with a basic understanding of Econometric theory and practice. We will focus on model specification, estimation, and testing, using a "hands on" approach. Throughout the course, we will use EXCEL, R, and SAS. We will cover most of Chapters 1-10 of the textbook, followed by some selected special topics as time permits. You should read through each chapter as we cover it. Special emphasis will be placed on conceptual understanding and application of econometric methods. For those who are interested in more involved discussions of the theoretical framework and/or the statistical or mathematical derivation behind any of the ideas discussed in class, feel free to meet with me outside of class.

Prerequisite: MAT 111 and CSC 111.

Offered: once a year.

DAT 412 Machine Learning 3 Credits

A foundational development of the core ideas and concepts in machine learning, with emphasis on the statistical foundations of machine learning but also applied work in Python, or a comparable language. Topics covered will include feature engineering and basis sets, gradient descent model fitting, kernel methods, Model selection methods, bootstrapping and other permutation methods, model inference and averaging, tree based methods with boosting and bagging, neural nets and deep learning and graph based methods.

Prerequisite: MAT 219 and CSC 112.

Offered: once a year.

DAT 417 Machine Learning for Natural Language Processing 3 Credits

This course is on constructing, training and using Machine Learning tools (neural networks) for Natural Language Processing, covering the fundamentals of operation of ChatGPT and other tools for generative language applications, translation, theme detection, text summarize, question answering and a range of other applications. This is a programming driven course, in which students will construct and evaluate a number of machine learning applications. Students will construct NLP processing models (neural networks) using the Pytorch and/or TensorFlow frameworks within the python programming language using the Jupyter notebook system. The course will also cover text encoding, tokenization, embedded and other reduced space representations, string and sentence transformations and related topics. Basic predictive models will be covered in the introduction to PyTorch and TensorFlow. Data storage in the Apache Arrow and HuggingFace datasets systems will also be discussed. Students may need to subscribe to the Google Colab Pro platform at a modest cost if they do not have regular access to a computer with an Nvidia GPU. Cost of the subscription is comparable to that of a typical electronic textbook.

Prerequisite: CSC 112 and CSC 112L.

Offered: once a year.

DAT 499 Independent Study Course in Data Science 1-3 Credits

Study and work with a faculty supervisor. Project to be determined by faculty agreement. Independent studies require an application and approval by the associate dean.

Prerequisite: DAT 211.

Offered: every fall & spring.