

# Data Science (DATA)

---

## Data Science (DATA) Courses

### DATA 1517 [0.5 credit]

#### Data Modelling I

Introduction to formulating statistical problems and analyzing data using open-source software. Graphical and numerical descriptives. combinatorial formulae, Bayes' Theorem, probability, Discrete and continuous distributions, means and variances. Point and interval estimates, and hypothesis tests for one- and two-samples using Central Limit Theorem, and permutation tests.

Precludes additional credit for BIT 2000, BIT 2009, BIT 2100 (no longer offered), BIT 2300 (no longer offered), ECON 2201 (no longer offered), ECON 2210, ENST 2006, GEOG 2006, STAT 2507, STAT 2601, STAT 2606 (no longer offered), and STAT 3502. May not be counted for credit in any program if taken after successful completion of STAT 2655.

Prerequisite(s): An Ontario Grade 12 university-preparation Mathematics or equivalent, or permission of the Institute for Data Science.

Lectures three hours a week, laboratory one hour a week.

### DATA 1519 [0.5 credit]

#### Data Modelling II

Introduction to modelling real phenomena from planning data collection or gathering observational data to analyzing and providing insights. Topics include experimental design from first principles and simulating the data generating process, linear regression and correlation, one- and two-way Analysis of Variance using open-source statistical software.

Precludes additional credit for ECON 2202, ECON 2220, STAT 2509, STAT 2559, STAT 2602.

Prerequisite(s): DATA 1517 or (STAT 1500 and STAT 2507) or (STAT 1500 and STAT 2655) or (STAT 1500 and STAT 3502); or permission of the Institute for Data Science.

Lectures three hours a week, laboratory one hour a week.

### DATA 2200 [0.5 credit]

#### Communication Skills for Data Scientists

Technical communication and data visualization skills for data science majors, concentrating on writing and orally presenting scientific papers and technical reports. Principles of clarity and precision in writing and oral communication. Practical exercises and readings from recent technical publications will be used.

Prerequisite(s): DATA 1517 or STAT 2507 or STAT 2559.

Lectures three hours a week.

### DATA 2500 [0.5 credit]

#### Data Wrangling in R

Reproducible workflows from acquisition, to cleaning, manipulation, and visualization. Data are acquired from databases, APIs, and web scraping. Cleaning and manipulating Numeric, categorical, date, and text data are introduced including regular expressions. Data visualization and report generation using dynamic tools are emphasized.

Prerequisite(s): (COMP 1005 or COMP 1405), STAT 1500, and one of (DATA 1517, STAT 2507, or STAT 2655).

Lectures three hours a week, laboratory one hour a week.

### DATA 3500 [0.5 credit]

#### Statistical Programming in R

Modern coding practices in R including running simulations, workflows for common statistical models, retrieving diagnostics and model estimates, and presenting and visualizing results. Emphasis on modern, reproducible workflows and version control.

Prerequisite(s): DATA 2500.

Lectures three hours a week, laboratory one hour a week.

### DATA 3999 [0.0 credit]

#### Co-operative Work Term

On completion of each work term, the student must submit to the Institute for Data Science a written report on the work performed. Graded SAT or UNS.

Includes: Experiential Learning Activity

Prerequisite(s): registration in the Co-operative Education Option, and permission of the Institute for Data Science.

### DATA 4848 [1.0 credit]

#### Consulting Project

This course is designed to give students some practical experience as a data science consultant through classroom discussion of issues in consulting and participation in real consulting projects.

Includes: Experiential Learning Activity

Prerequisite(s): fourth-year standing in the Bachelor of Data Science program.