Syllabus

MAT 165 - Introduction to Data Science

General Information

Date January 20th, 2023

Author Kimberly Wager

Department Mathematics

Course Prefix MAT

Course Number 165

Course Title Introduction to Data Science

Course Information

Catalog Description This course covers techniques for working with data, including getting and cleaning data, exploratory data analysis, data visualization, and statistical modeling and prediction. Students will learn how to ask good questions, apply data to practical problems, and communicate data analytic results. Statistical computing is integrated into the course.

Credit Hours 3

Lecture Contact Hours 4

Lab Contact Hours 0

Other Contact Hours 0

Grading Scheme Letter

Prerequisites

MAT 145 or placement into Math Level 3 or higher.

Co-requisites

None

First Year Experience/Capstone Designation

This course DOES NOT satisfy the outcomes applicable for status as a FYE or Capstone.

SUNY General Education

This course is designated as satisfying a requirement in the following SUNY Gen Ed category
Mathematics (and Quantitative Reasoning)

FLCC Values

Institutional Learning Outcomes Addressed by the Course
Inquiry and Interconnectedness

Course Learning Outcomes

Course Learning Outcomes

1. Import, organize and manipulate real-world data to support disciplinary inquiry.

2. Analyze and interpret descriptive statistics and graphical summaries of data regarding center, variation, and shape.

3. Apply the principles of inferential statistics to develop, evaluate, and interpret models based on data.

4. Communicate statistical results to technical and non-technical audiences.

Outline of Topics Covered

This is an introductory course on data science. Data science is a multi-disciplinary field incorporating elements of mathematics, statistics, computer programming, and data skills. In data science, statistical, mathematical, and data skills are applied to solve disciplinary problems in disciplinary data. Statistical concepts are taught while simultaneously learning the R programming language and environment to simulate results.

- **I.** Introduction to data science – Four pillars of inference
    - **a.** Strength of evidence
    - **b.** Breadth of results
    - **c.** Size of effect
    - **d.** Cause of effect
- **II.** Statistical inference using one group
    - **a.** Random processes and chance based models
    - **b.** Data visualization

    c.   Exploratory data analysis

    d.   Parameter estimation using bootstrapped distributions

    e.   Distribution of outcomes under the null hypothesis

    f.   Hypothesis tests for proportion and mean

    g.   Errors and significance

    h.   Confidence intervals for proportion and mean

    i.   Factors on width and precision of intervals

III.   Statistical inference using two groups

    a.   Contingency tables

    b.   Hypothesis tests for proportion and mean

    c.   Confidence intervals for proportion and mean

    d.   Effect size statistics

IV.   Modeling

    a.   Scatterplots and c**orrelation**

    b.   Ordinary least squares regression

    c.   Inference with correlation/regression coefficient

    d.   Prediction with least squares regression equation

    e.   Logistic regression coefficient: log-odds, odds ratio

    f.   Logistic regression using maximum likelihood

    g.   Prediction with logistic regression