MATH/INTE 255 – Intro to Data Analytics – SYLLABUS

Department of Mathematics
Millersville University

Description

Introduction to data analysis techniques and programming that enables real-time decision making in IT organizations. Includes skills and applications in pre-processing, preparing, and reporting data for further analysis. (3.0 credits)

Cross-listed with INTE 255, credit may not be received for both courses.

Prerequisites

MATH 130 OR MATH 234 OR MATH 235

Prerequisite that may be taken as a corequisite: MATH 130 OR MATH 234 OR MATH 235

Course Objectives

Students will be able to

- Be proficient in the R statistical programming language and to be able to expand the knowledge learned from the class to another preferred programming language, if need be
- Describe sources of data, their format, and how they can be prepared for analysis
- Evaluate the quality and validity of data used to support a claim or argument
- Implement quantitative strategies to use data for forecasting and prediction
- Develop and deliver effective presentations of data-informed conclusions to a specific audience

Assessment

Assessment of student achievement of the course objectives will vary from one instructor to another. Typical assessment will be made through work in class, homework, projects, presentations, and examinations administered in a traditional face-to-face classroom environment.

Use of Technology

Students will be well served by a statistical computing package on their own devices.

Topics

1. Introduction

    a) Why use R?
    b) Installing and getting started with R and RStudio
    c) Installing Packages

2. Dataset Creation
    a) Understanding datasets
    b) Understanding the different data structures, such as vectors, matrices, arrays, data frames, etc.
    c) Data inputting and importation from various sources
    d) Annotating datasets

3. Data Management
    a) How to create variables
    b) How to recode and rename variables
    c) How to deal with missing values
    d) How to convert data values
    e) How to sort data
    f) How to merge and subset datasets

4. Advanced Data Management
    a) Numerical and character manipulation functions (statistical, probability, mathematical, regex's, etc.)
    b) Applying the functions to matrices and data frames
    c) Logical expressions
    d) Loops
    e) User-written functions
    f) Aggregation, reshaping, melting, and recasting of data

5. Graphical Techniques
    a) Creating graphs with base R
    b) Creating graphs with ggplot (if time permits)
    c) Manipulating parameters of graphs (text, legends, etc.)
    d) Creating different types of graphs (bar graphs, histograms, box plots, etc.)
    e) Nonparametric tests for group differences (if time permits)

6. Statistical Methods
    a) Descriptive statistics
    b) Frequency and Contingency Tables
    c) Sampling distributions
    d) Confidence intervals
    e) Hypothesis testing (t-tests, z-tests, and tests for proportion for one-sample and two-sample data)

7. Regression
    a) Ordinary least Squares (Simple linear regression, quadratic regression, multiple linear regression with and without interactions (if time permits))
    b) Regression diagnostics and how to correct a model
    c) How to identify and deal with unusual observations
    d) Selecting the `best' model

Recently Used Textbooks

- *R in Action: Data Analysis and Graphics with R*, 2nd or 3rd edition, Robert Kabacoff, Manning (2015/2022).