NOVA COLLEGE-WIDE COURSE CONTENT SUMMARY
ITD 145 - APPLIED DATA SCIENCE TECHNIQUES (3 CR.)

Course Description

Reviews the fundamentals of descriptive and inferential statistics, probability, and distributions, as well as basic dataset manipulation and plotting techniques. Focuses on application to real datasets using graphical user interface (GUI) software tools as well as Python. Lecture 3 hours. Total 3 hours per week.

General Course Purpose

ITD 145 introduces students to the basic data science concepts using statistics to characterize and explore datasets by utilizing visual tools as well as coding in Python.

Course Prerequisites/Corequisites

None.

Course Objectives

- Describe, explain the purpose of, and use basic statistics on data
    - Use a statistical software package
    - Define and explain the purpose of a metric
    - Measure Central Tendency
    - Measure Dispersion
- Describe, work with, and manipulate datasets from various sources and formats.
    - Define and explain the purpose of evaluation metrics
    - Define and calculate precision
    - Define, explain and calculate correlations
    - Define and classify dependent and independent variables
    - Define and explain random variables
    - Explain the distribution of random variables
- Extract, transfer and clean up data from raw data sources, transforming them into usable forms.
    - Formulate null and alternative hypotheses
    - Compare and contrast z-tests and t-tests
    - Identify Type I (false positive) and Type II (false negative) errors
    - Use Python tools, such as Numpy, Pandas, and Matplotlib
- Describe and generate various visualizations from raw and derived data.
    - Explain the purpose of visualization plots and generate basic plots
    - Define and examine sample distributions, focusing on Gaussian distributions (normal distributions)

Student Learning Outcomes

- Explain the purpose of statistics
- Define and Explain:
    - qualitative variables
    - quantitative variables
    - continuous quantitative variables
    - discrete quantitative variables
- Define, obtain and use a data set
- Define and examine sample distributions, focusing on Gaussian distributions *(normal dist.)*
- Define and explain the purpose of a metric
- *Measuring Central Tendency*
    1. Define and calculate the mean, median and mode
- *Measuring Dispersion*

- o Define and calculate the range
- o Define and assess a skew
- o Define and calculate variability
- o Define and classify outliers
- o Define and calculate variance (σ^2)
- o Define and calculate standard deviation (σ)
- o Define and express the formula for standard error
- Define and explain the purpose of evaluation metrics
- Define and calculate accuracy, precision and recall
- Define, explain and calculate correlations
- Define and classify independent variables
- Define and classify dependent variables
- Explain the purpose of visualization plots and generate basic plots
- Define and explain random variables
- Explain the distribution of random variables
- Formulate null and alternative hypotheses
- Compare and contrast z-tests and t-tests
- Identify Type I *(false positive)* and Type II *(false negative)* errors
- Use a statistical software package to
  - o calculate sample means, standard deviation, and confidence intervals
  - o create appropriate graphs
- Use Python tools, such as NumPy, Pandas, and Matplotlib to
  - o load datasets into memory
  - o manipulate table data
  - o generate statistical information about a dataset, including metrics on data quality
  - o perform common statistical calculations
  - o locate missing data and determine whether to drop or impute such data
  - o generate various plots for visualizing raw datasets

Major Topic to be Included

- Basic descriptive statistics
- Statistical distributions
- Data manipulation and cleaning/'wrangling'
- Data visualization
- GUI applications and Python to describe datasets with statistics and plots

Required Time Allocation per Topic
To standardize the core topics of this course, the following student contact hours per topic are required. Each syllabus should be created to adhere as closely as possible to these allocations. Topics are not necessarily to be taught in the order shown.
There are normally 45 student contact-hours per semester for a three-credit course (14 weeks of instruction, excluding final exam week: 14*3.2 = 45 hours). Sections of the course offered in alternative formats (i.e., not standard 15-week) still meet for the same number of contact hours. The final exam is not included in the timetable.
The quickly evolving nature of data analytics means that some content noted in this document may be superseded or made obsolete. As such, it is important to include such changes in individual syllabi. Additionally, time is allocated for additional and optional topics in order to provide instructors flexibility in tailoring the course to special needs or resources.

Topics	Hours	Percentage
Basic descriptive statistics	6	13%
Statistical distributions	6	13%
Data manipulation and cleaning/'wrangling'	9	20%
Data visualization	6	13%
GUI applications and Python to describe datasets with statistics, plots	9	20%
Testing to include quizzes, tests and exams (excluding final exam)	3	8%
Other optional topics	6	13%
Total	**45**	**100%**