DATA 690 STATISTICAL ANALYSIS & VISUALIZATION WITH PYTHON

Course Description:

This course aims to provide

(i) an introduction to Python programming,

(ii) the fundamental statistical concepts and methods, which are frequently used in exploratory data analysis, and

(iii) how to carry out these analyses and visualize the findings in Python.

Prerequisites: Enrollment in the Data Science program.

Recommended References:

- Python Programming: An Introduction to Computer Science by John Zelle
- Programming in Python 3: A Complete Introduction to the Python Language by M. Summerfield
- Allen B. Downey, Think Stats, 2nd edition, O'Reilly
- Statistics for Business & Economics by Anderson et al.
- Python package* user guides (i.e. NumPy User Guide, Pandas Reference Guide, etc.)

* Python packages (tentative)

Analysis:	NumPy and SciPy
Handling Data:	Pandas
Plotting:	Matplotlib and Seaborn

Course Content (tentative):

Week	Description
1	Introduction to Python and Notebooks, basic data types
2	Python Statements, Loops, and Functions
3	Python: IO, Files, and Numerical Calculations with Numpy
4	Python: Basic plotting with Matplotlib and data handling with Pandas
5	Python: Advanced Plotting and Time Series
6	Introduction to Statistics
7	Descriptive Statistics-1
8	Descriptive Statistics-2
9	Introduction to Probability
10	Discrete and Continuous Probability Distributions
11	Sampling, Distribution, and Interval Estimation
12	Hypothesis Testing
13	Inference
14	Review

Recommended Software

The course will be using Python 3 with the following libraries: numpy, pandas, matplotlib, seaborn, and stats. It is the student's responsibility to have a working environment. If you'd like to have the environment installed locally, Anaconda is a Python distribution that has all required libraries. The recommended option is to use Goolge's Colab which is available from your UMBC account.

Course Format and Assignments

Students will complete weekly and/or bi-weekly homework assignments and a project. The assignments will give students an opportunity to gain practical insights with Python programming, explanatory statistical analysis and visualization. The project will give students the opportunity to practice various subjects studied in class.

Grading Criteria

Course work	Grade distribution
Attendance	7%
Homework Assignments	60%
Project	33%

Final grade might be computed as follows:

85% - 100%	A
70% - 84%	B
55% - 69%	C
40% - 54%	D
<40%	F

However, someone with a final score of 83 might also get an A, depending on the overall success of the students. (Of course, same thing might happen in the opposite direction).

Course Policies

UMBC provides a range of writing assistance, which can be found in the following:

- The Writing Center https://lrc.umbc.edu/tutor/writing-center/
- Research Guides & Tutorials https://lib.guides.umbc.edu/tutorial

Failure to follow guidelines for each assignment, including the required format, style, length, submission, etc., may result in at least one-letter-grade reduction on the paper depending on the type and/or number of transgressions.

Late/Incomplete assignments will not be accepted unless an extension has been agreed to in advance. Emergency situations will be handled on a case by case basis with appropriate justification and/or documentation.

Incomplete grades will not be entertained unless extenuating circumstances warrant and your request is made before the last week of class.

Academic Integrity

By enrolling in this course, each student assumes the responsibilities of an active participant in UMBC's scholarly community in which everyone's academic work and behavior are held to the highest standards of honesty. Cheating, fabrication, plagiarism, and helping others to commit these acts are all forms of academic dishonesty, and they are wrong. Academic misconduct could result in disciplinary action that may include, but is not limited to, failure, suspension or dismissal.

Refer to the UMBC policy at:
https://catalog.umbc.edu/content.php?catoid=14&navoid=718#academic-integrity