STATS 607 – Programming and Numerical Methods in Statistics

Fall 2018

Class Information

- **Days & Time:** Mondays & Wednesdays, 4pm –5:30pm.

- **Location:** 1372 EH.

- **Description:** This course is the first part to an advanced introduction to modern programming techniques used in statistics, modeling and data analysis. Course topics include: basic data structures, structured data formats, iteration and recursion, functional programming, classes and object-oriented programming, memory management, strategies for documenting and debugging code. This part of the course will cover programming fundamentals relevant for research on statistical methodology, and for working with large and complex data sets.

- **Textbook:** There's no official textbook. I will be adding resources for each lecture in the material's webpage.

- **Canvas:** You should access canvas frequently. There you will find important announcements.

- **Materials:** Class materials can be accessed at https://github.com/marcio-mourao/Stats607-Fall2018/wiki.

- **Course end date:** This is a half-semester course and will end on October 22, 2018.

Instructor Information

- **Name:** Márcio Duarte Albasini Mourão.
- **Office:** 3550 Rackham, 915 E. Washington St. (@CSCAR).
- **Office Hours:** Mondays and Wednesdays 10-11am.
- **Email:** mdam@umich.edu - please include 'Stats 607' in the subject!

GSI Information

- **Name:** Laura Niss.
- **Office:** Science Learning Center's (SLC) satellite location, 2165 USB.
- **Office Hours:** Tuesday 3:00 – 4:30pm.
- **Email:** lniss@umich.edu - please include 'Stats 607' in the subject!

Grading

Your scores in 3 assignments, with the tentative dates below, will determine your final grade in the course:

- **Assignment 1:**

    - **Out:** Sept 19, **Due:** Sept 26 (20%)

- **Assignment 2:**

    - **Out:** Sept 26, **Due:** Oct 5 (35%)

- **Assignment 3:**

    - **Out:** Oct 10, **Due:** Oct 22 (45%)

Python Environment

I'll be using JupyterLab to develop the lectures. The JupyterLab is the next generation of the Jupyter Notebook and allows the creation and sharing of documents with text, code, equations and images. For additional details, see here: https://jupyter.org. For code development and assignment resolution, I recommend using Spyder.

Both JupyterLab and Spyder are included in the Anaconda distribution. See here: https://www.anaconda.com.

Reading

Textbook: "Python for Data Analysis" by Wes McKinney published by O'Reilly Media. Python Online Tutorial: https://docs.python.org/3/tutorial/index.html.

Schedule

Week 1 (Sept 5)

>**Lecture 1:** Introduction to Stats 607 and Python

Week 2 (Sept 10 & 12)

>**Lecture 2:** Control Flow and Function Arguments

>**Lecture 3:** Data Structures

Week 3 (Sept 17 & 19)

>**Lecture 4:** Standard Library

>**Lecture 5:** Reading and Writing Files, Errors and Exceptions

Week 4 (Sept 24 & 26)

>**Lecture 6:** Python Internals, Memory Management, Serialization

>**Lecture 7:** Python Classes, Object Oriented Paradigm

Week 5 (Oct 1 & 3)

>**Lecture 8:** Numpy

>**Lecture 9:** Numpy

Week 6 (Oct 8 & 10)

>**Lecture 10:** Pandas Dataframes

>**Lecture 11:** Pandas Dataframes

Week 7 (Oct 15 & 17)

    Lecture 12: Profiling and Optimizing Python Code

    Lecture 13: Strategies for Documenting and Debugging Code

Week 8 (Oct 22)

    Lecture 14: To be determined