# Final Project Proposal

## Course

CCSU Stat 476. Spring 2022.

## Student

Tim Brockway. Student ID: 30259316 Email: BrockwayTim@My.CCSU.edu

## Professor

Roger L. Bilisoly. bilisolyr@ccsu.edu https://www2.ccsu.edu/faculty/bilisolyr

## Basic Idea

It seems that medical screening tests vaunting very high "general accuracy" can give staggering levels of false results when the prevalence of a disease is low. I'd like to understand this by writing a program that explores the effect the prevalence of an infection in a population on the usefulness of screening tests.

## Python Exploration

I'd like to use the project to explore Python's capabilities and weaknesses.

In particular I wish to learn about.

- Program modularization using functions driven by a main line.

- Handling global variables.

- Creating global variables using the 'global' statement is deprecated.

- These statements are in any case counterintuitive and complex to use.

- I want to try python classes to create global, self-initializing variables to allow program modularization.

- Panda based data handling. Pandas seem to be the standard data table for python.

- Some sort of GUI to permit the user to enter the medical tests parameters.

- A somewhat interactive plot to allow the user to choose which variables are displayed.

- Use Github to get a feel for how most folks are storing and publishing their work.

- Storing the code, project documentation and all project files in one place will be useful in any case.

- Some other Python features I've played around with this and would like to work it into the project.

- Nested Python dictionaries and pretty printing Python dictionaries.

- Enumeration of non iterable objects (Enum).

- Etc.

## Inputs

- A GUI of some kind, will invite the user to enter the following medical test statistics.

- Inputs will have to be validated by the code.

- Population.

- Test sensitivity.

- Test specificity.

- Start of population disease prevalence range of interest.

- End of population disease prevalence range of interest.

- Prevalence of particular interest.

- A way of indicating which graphs are required. All at once would be a bit cluttered.

- Some gratuitous file handing to explore Pythons i/o methods. (E.g., Download the pandas dataframe generated).

## Outputs

- A report on the statistics resulting from the users test specification.

- An interactive graph with:

x axis. The required range of disease prevalences.

y axis: Any or all of the following:

- False positive percentage.

- False negative percentage.

- Positive predictive value.

- Negative predictive value.

- False Positives (NPV)

- False Negatives (NPV)

- The general accuracy (The suspect statistic).

- Prevalence of interest (A vertical line).

- Error messages indicating errors in user's input.

- Code to allow.

- Download of a CSV file with the plot data. (Optional).

- Download the plot itself. (Optional).

- Some hyperlinks to (say) documentation and code at github.

- Anything else that the usual input/calculate/plot statistics program might use.

## Statistics

I will have to explore the epidemiological issues and terminology to be sure I generate the statistics correctly. I do not know if the various python libraries have the statistical functions needed to do screening test calculations per se. But I want to do the calculations myself in raw python. Prima facia once the variables' purpose and derivation are understood calculating them is easy. I should disclose that at the time of writing I am not fully on top of the epidemiology but there is a ton of help online.