

Graduate Project for Biostatistics 2017. v1.0

Develop a statistics application using Python that would allow a non-expert to carry out a variety of statistical tests.

It is recommended you work in pairs given the size of the project.

Your application should support the following tests:

1. Plot a Q-Q plot for a set of x/y data
2. Compute mean confidence limits for large and small samples
3. Compute confidence limits for the difference between large and small samples
4. Compute mean confidence limits for a sample using simulation
5. Hypothesis testing:
 - (a) Large and small sample tests for population mean
 - (b) Large and small sample tests for the difference between two means.
 - (c) Chi-Squared Test
6. Correlation and linear Regression
 - (a) Compute the correlation coefficient for a set of x/y data
 - (b) Carry out hypothesis testing on the correlation coefficient
 - (c) Optionally allow user to supply error bars with the data
 - (d) Perform a linear regression on a set of x/y data
 - i. Display the plot with the data and fitted line and if available the error bars
 - ii. Allow the user to plot the residuals from a linear regression
 - iii. Provide in an output table the following information:
 - A. Best slope and y-intercept
 - B. Statistic test (H_0) that the correlation coefficient is zero (ie no relationship)
 - C. Compute the uncertainties in the slope and y-intercept
7. Propagation of error calculations

Users should be able to enter data either manually or via files.

Deliverables

Total points: 100

1. Points (15) A written summary of the software's capabilities, no more than a page, or if need be a page and a half.
2. Points (40) The application in Python, fully documented.
3. Points (25) A series of test data and runs that confirms the software works as expected
4. Points (20) A tutorial video that describes how to use the software. The video should go for at least 8 minutes if not more. Upload the video to YouTube and provide a link with your submission.