

# CS 499 Senior Project

## Project Assignment

# Statistical Analyzer

### Rationale

Anyone doing research, especially in the social sciences, uses statistics to analyze the resulting data. Usually a number of different statistical measures are performed on the data.

This class project will be to produce a software system which will take a set of data and perform user selected statistical analyses on the data then generate both an on-screen display of the results and a printable file of the results.

### Features

The features listed below shall be included in the software.

1. The proposed system will accept as input either (1) a comma delimited or tab delimited data file such as can be output from Microsoft Excel or (2) data input into an editable table or text fields on screen.
2. The user shall be able to select any or all of the following statistical measures to be calculated on the data if the measure is meaningful for the type of data (frequency, ordinal, interval): mean, median, mode, standard deviation, variance, coefficient of variance, percentiles, probability distribution, binomial distribution, least square line,  $X^2$  (Chi Square), correlation coefficient, sign test, rank sum test, and Spearman rank correlation coefficient.
3. When appropriate for the statistical measure the following graphs shall be selectable by the user for display: Horizontal bar chart, vertical bar chart, pie chart, normal distribution curve, X-Y graph, table
4. The system shall be designed in such a way that new statistical measures can be added to the system easily and with a minimum of code revision (hint: study Design Patterns by Gamma, Helm, Johnson, and Vlissides)
5. User interaction with the system shall be via a Graphical User Interface.
6. For statistical analyses the user may specify certain rows and columns within the rows, or columns and rows within the columns to perform the analyses on. The selection must be appropriate for the type of statistic to be calculated.
7. Users may select more than one statistical test for any specified set of data.
8. The system shall be capable of producing comma or tab delimited data files for appropriate statistical results which can be read and edited using Microsoft Excel, or a simple text based output if tabular output is not appropriate.
9. The system shall be capable of producing images in JPEG format of any graph displays.
10. The system shall be capable of producing text files summarizing the results of all statistical analyses performed on the data.

### Constraints

The choice of programming language is left up to the design team.

Users must be able to run the program on a PC under Windows 7 or 8, on a Macintosh under OS X, or on a Linux system.

The system will be considered to be open source and in the public domain, therefore all code must be original and may not include any copyrighted material.