

Rikard E. Grafstroem (reg)

### **Introduction, motivation**

This is an attempt at creating requirements for a friendlier interface for Stata, the statistics package for Windows. I used this package in an Economics course recently, and was amazed how poorly designed the interface was. I spent hours trying to figure out the simplest command, which in some cases required me to use very advanced constructs which at times resembled C more than anything else. How a non-CS major dealt with it is somewhat mysterious to me.

Stata has plenty of powerful tools, but they are completely inaccessible to the average user. It lacks good help functions, easy ways to analyze, graph, and input data. All input is made through a command line, and corresponding output comes in a separate window, which even lacks a scrollbar – a separate log file must be created before there exist a command history.

Each command comes with a set of optional variables, all which are poorly explained in the help files. The help gives you several examples of the commands without explaining to the user what the particular example syntaxes do. Hence, they are in many cases completely useless. It is very difficult to browse between the different commands or to search for related ones. Each command has so many optional variables that one easily focuses too much on a set of unnecessary items, which one doesn't even know if they are needed or not.

### **Requirements**

One should remember that Stata is designed for statisticians and one can assume that the user knows statistical terminology. S/he doesn't necessarily have extensive experience with data sets in a computer environment, hence it must be easy to input data, create variables, and display graphs, tables and other output.

Help files are key. It must be easy to find what functions are available, and they must be categorized in a logical manner, and it must be easy to find related commands. Optional parameters should be well explained and under what circumstances they are necessary and useful.

Command input should be very easy. It should be very clear as to what is required for the command to generate the correct output. A wizard-style interface would be helpful. Something that could be modeled on how MS Excel does it would be a good start. Commands must be well-categorized in some logical fashion, and the most frequently used commands should be quickly accessible. There must be multiple ways to input commands, so the more experienced user can skip the Wizard and input it directly on a command line. One should be able to use either way interchangeably. A toolbar with commonly used commands should be visible, and customizable.

Manipulation of variables should be easy. It should be easy to add, remove and change variables.

Input should be accepted from a variety of sources, and the program should be able to interpret raw data correctly. Input files could plain text files organized in a wide variety of ways, ideally it should also be able to deal with input files from other programs.

Output should come in a wide variety of graphs, tables, and charts. They should be highly customizable (color, shapes, labels, etc.), and they should be easy to publish – either on a hard copy or on the web. All this should of course be possible without having to use the command line – wizards, menus, and toolbars should provide this functionality.

It must be possible to log all actions done, and provide a printable document which shows what you have done to manipulate the data.