Data management and preparation using macros in *Stata*

Date: October 30th, 2014

Workshop instructor: Claudia Masferrer, Msc
PhD Candidate, McGill University

Are you tired of doing *copy/paste* for performing a series of repeated procedures and finding mistakes in your Stata do-files? Do you use routines and macros to prepare clear and efficient syntaxes? Data preparation is a very important step in quantitative analysis. Unfortunately, data preparation can be very time consuming if data is not managed efficiently and correctly. The main objective of this hands-on data workshop is to provide users with knowledge of the use of macros, as well as tricks and advice on programming essentials for data management.

As an example, we will prepare and manage longitudinal data. This workshop will also provide training in the use of a basic package that allows the renaming of time-varying variables, and thus facilitate the creation and use of databases either in person-level format or in person-period format depending on the software or the kind of longitudinal model used in the analysis. We will use data from the National Population Health Survey (NPHS) that cannot always be directly used for longitudinal analysis. The original data file is in a person-level format (that is, a "wide" matrix, where each respondent has only one record, with as many variables as the number of questions times the number of measurement occasions). Thus, one of the first necessary steps for longitudinal analysis is to organize the data in a person-period format (that is, a "long" file, where each person has one record for each measurement occasion). The ease of conversion of data files from one format to another hinges on a convention around variable names.

Outline:

SECTION 1: Introduction to the use of macros in *Stata*. Hands-on use of the commands *local, global, foreach, forvalues, while, if-else*.

SECTION 2: Review of commands needed to prepare longitudinal data. Discussion of the main characteristics of the National Population Health Survey (NPHS) and presentation of the syntax used for preparing NPHS files and an interactive implementation of the program.

Note: This workshop will be of interest to all researchers working, or planning to work, with longitudinal data, regardless of whether they use or intend to use NPHS or other longitudinal surveys. Given Statistics Canada regulations, the synthetic ("dummy") files from NPHS will be used.

Participants: The course is open to graduate students and postdoctoral fellows as well as to professors and practicing researchers. Priority will be given to researchers using datasets available at QICSS.

A minimum knowledge of the use of Stata is required, but participants do not need an advanced level of knowledge. Participants who are not Stata users should however be active users of other statistical software like SPSS or SAS. Participants with an advanced level of Stata would benefit from this workshop by learning programming tips and techniques.