Stacked Data Matrices Workshop

Stata Script Tutorial

Giuseppe Carteny

01.06.2021

Contents

Overview

This document consists in a step by step walkthrough of the Stata script dedicated to the creation of a stacked data matrix (SDM) of the European Election Study (EES), in particular of the Italian voter study of 2019. After a brief introduction to the Data, the document continues with a presentation of the exemplary Stata script provided for the workshop ("EES2019_Stata_stacking_example.do").

Data

Data can be downloaded through the following link: https://github.com/giucarny/StackMat/archive/refs/heads/master.zip. After clicking or browsing the link, a .zip folder named 'StackMat-master' will be automatically downloaded. Unzip it and browse it at your wish. In sum, this folder represents the copy of a public GitHub repository named ' StackMat ' which contains three main subdirectories:

- '**Data**': In which are stored the (main and auxiliary) datasets for the stacking procedure (e.g., the dataset of the EES voter study of 2019) as well as the SDM generated by the Stata script explained in the following lines;

- '**Scripts**': In which are stored the (main and auxiliary) scripts for stacking country-specific data frames of the EES 2019 voter study , including by said Stata script;

- '**Documentation**': A folder containing relevant documents for the stacking process (e.g., the EES 2019 voter study questionnaire and codebook) and the tutorial (e.g., the codebook of the SDM generated by the Stata script presented below).

Stata Script

Browse the "~/Data/" subirectory and then double-click on the "EES2019_Stata_stacking_example.do" file. If ".do" files are linked to your current Stata version[1] this will open a new Stata working session. If not, then open Stata, and open the already mentioned ".do" file from the software.

Many workflows can be followed for creating a SDM. These workflows can be summarized, however, in two main strategies:

A. Mutate the relevant variables => Stack the original matrix
B. Stack the original matrix => Mutate the relevant variables

In Stata, both strategies are feasible. However, the script discussed below follows the first strategy (A), thus representing *one* of the possible workflows to create a SDM in general, and in particular in Stata.

---

[1]The script presented in this tutorial has been tested on Stata versions 15 and 16. If you need to purchase, download, or update Stata please browse to the "~/Documentation/" subdirectory and check the "StackMat---Software-Setup.pdf" file.

Setup Lines

The first lines of the script are dedicated to the software setup, in particular to the installation/udpate of the StackMe package. This package is available on a GitHub repository, but not on the Boston College Statistical Software Components (SSC) archive. Consequently, StackMe installation requires the prior installation of another package (github) that allows to download and install packages directly from GitHub.

If you already installed both github and StackMe packages, then you can skip the code lines below. If you did not, you can run the installation of both packages using the following commands. If you already installed both packages and re-run the installation Stata will automatically check the version and consistency of the packages.

```
net install github, from("https://haghish.github.io/github/")
github install ldesio/stackme
```

If you get any error during the installation/update of both packages, or problem in the script chunks that involve StackMe commands, then you might uninstall previous versions of the packages with the following commands, and then rerun the previous steps.

```
ado uninstall github
ado uninstall stackme
```

Admin

In this passage of the script you need to set up the working directory. For compiling properly, the following script requires that you set the working directory in the "Data" subdirectory of the 'StackMat-master' folder ("~/StackMat-master/Data/"). Thus, check the working directory of the current session of Stata and then set the working directory in the "Data" folder.

```
pwd
cd " ~ \StackMat\data"
```

Load (and Merge) Data

After setting up the relevant packages and the working directory, then you can load the data for the stacking procedure.

First, we load on Stata the main dataset (the EES 2019 voter study) for the stacking procedure.

```
use "ZA7581_v1-0-0.dta", clear
```

Then, we merge this dataset with an auxiliary one, containing party-specific data from the Chapell Hill Experts Survey (CHES) of 2019[2], such as experts average scores of party positions on the Left-Right *continuum* or about the European Union (EU) integration process, that will return later on in the script. Then, drop the _merge variable for allowing the following merging operation.

```
merge 1:m respid countrycode Q7 using EES_CHES_2019_aux
drop _merge
```

Finally the dataset is merged with a second auxiliary data frame, containing the EES 2019 party identification variable (Q25; See the EES2019 Codebook) recoded according to the 2019 European Parliament (EP) Elections vote choice variable values (Q7; See the EES2019 Codebook).

Select Party System and Relevant Parties

---

[2]These data are derived by the 1999-2019 CHES trend file, available in the "Data" subdirectory ("~/StackMat-master/Data/"). The dataset derived from the CHES file is created with an R script, available in the "Scripts" subdirectory ("~/StackMat-master/Scripts/"), named "StackMat-aux_data_script.R". This script can be easily implemented to include further party-specific variables or, in perspective, country-specific variables for comparative analyses.