

Working with NCES Datasets

Will Doyle

2020-09-06

Intro

In this lesson we'll go over how to access and work with various NCES datasets in Stata. I diverge from the intended use from NCES, as there's a simpler way to get access to the data than the one that they lay out.

```
. capture log close

. log using "nces_datasets.log", replace
-----
      name:  <unnamed>
      log:   /Users/doylewr/lpo_prac/lessons/s1-03-nces_datasets/nces_datasets.log
log type:  text
opened on:  7 Sep 2020, 15:58:46
```

Directory structure

Data files (particularly large ones) should be stored in their own subdirectories. While it's possible to simply dump everything in one big directory, you may find that over time, as the folder grows, it becomes very difficult to find what you need and almost impossible to share your work with others. Yes, your computer can search really well. An organized directory structure is for you, the human. Get into the habit now, and you'll be thankful later.

Directory structures

In programming, we many times need to move around in directories on a computer. Sometimes we use fixed paths, which specify exactly where something is on the computer, other times we use relative paths. An example of a fixed path would be something like:

/Users/doylewr/lpo_prac/lessons/s1-03-nces_datasets

That path specifies the exact directory on my computer. In general, you really should avoid fixed paths, because everyone's computer is different. However, you might have something like a clone of a GitHub repository on your computer. Within that repository, you can specify relative paths to clarify where you want

the program to look. A standard directory structure for a statistical programming project is something like this:

```
—project_directory/
    ----data/
        -----source/
        -----analysis/
    ----scripts/
    ----output/
        -----tables/
        -----graphics/
    ----paper
```

Since our code exists in the `./scripts` directory, to access the source data names `source_data.dta` we would need to go up one level to the main project directory and then down into the source data directory. The command for this in Stata would be:

```
use ../data/source/source_data.dta
```

The `../` means to go up one level. Using `./` means to go into a subdirectory, or down one level.

In my github repository, I store large data files in the data directory. To access that directory from the current lesson I need to go up two levels and then into the data directory, so the relative path is: `../../data/`.

Working with globals

We're going to set the structure for ourselves using global macros. In Stata, a macro is a variable that can be set to some value. There are two types of macros in Stata— global and local. A global macro is persistent during a session (when Stata is open). A local macro is forgotten as soon as a script (do file) is run. In general, I will encourage you to use local macros, as their use enables better programming practices and replication. However, using global macros is a good idea for data management.

First we tell Stata what a macro will represent:

```
. global datadir "../../data/"
```

What the above means is that every time I call that macro, Stata will know I means the directory in question. We can test this by asking Stata to display the global ...

```
. display "$datadir"  
../../data/
```

... and there you have it.

Working with HSLS

The high school longitudinal study of 2009 tracks a set of students who began high school in 2009. It has been updated in 2012 and again in 2016. It's a great source of information about how students navigate high school and make the transition to college or the workforce (and in many cases both).

HSLS can be accessed using the nces electronic codebook: <https://nces.ed.gov/OnlineCodebook>
Once variables have been selected from the codebook, they can be accessed using the **use** . . **using** approach below:

```
. use ///  
    STU_ID ///  
    X1SES ///  
    using "${datadir}hsls_17_student_pets_sr_v1_0.dta", ///  
    clear  
  
. save "${ddir}hsls_analysis"      , replace  
file hsls_analysis.dta saved  
  
. exit
```