Stata Documentation: SVY Example

Stata Documentation

- Stata has excellent help files accessible via the `help` command

- These files are a great asset for recalling commands, syntax, and brief examples

- These curated help files are also excellent for graphing details (which are very difficult to memorize due to the sheer number of options)

Help File Links

- At the top of many help files is also a link to **more documentation** in the *Stata Reference Manual* and *Stata User's Guide*. Let's take a look at `help graph` and `help regress`

Extended Documentation

- In addition to the in-program help files and the *Stata Reference Manual*, there are topic-specific guides provided by StataCorp: http://www..stata.com/features/documentation/

- Here you can find detailed information on the commands and tools required to perform more complex analyses or write your own programs

SVY Tools Disclaimer

- Survey design and analysis are not areas of expertise for me statistically

- Please treat this section as an example of how to explore Stata's rich documentation — interrupt and correct me often!

- We can go slowly and explore the methods / information that is most relevant to your needs

SVY Prefix

- The `svy` prefix allows for the proper analysis of survey data

- Stata's survey data analysis tools help the user take into account:

  - Sampling weights

  - Cluster sampling

  - Stratification

svyset

- In order to properly adjust analyses for survey data, Stata needs to know the type of survey data loaded

- The first step when dealing with survey analysis in Stata is to define the survey settings of the dataset using the `svyset` command

- Some data sets (.dta files) will already have survey settings (a blank `svyset` call will show current settings)

- Lets look at the `help svyset` to learn more about how to set up a survey dataset

svyset commands

svy : tabulate v1 [v2]

svy: mean

svy : regress

- With a correctly designated **svyset** statement at the beginning of analysis, these commands will return accurately adjusted results based on complex survey data

SVY `subpop` <u>**NOT**</u> `if/in`

Survey data concepts

The variance estimation methods that Stata uses are discussed in [SVY] **variance estimation**.

Subpopulation estimation involves computing point and variance estimates for part of the population. This method is not the same as restricting the estimation sample to the collection of observations within the subpopulation because variance estimation for survey data measures sample-to-sample variability, assuming that the same survey design is used to collect the data. Use the `subpop()` option of the `svy` prefix to perform subpopulation estimation, and use `if` and `in` only when you need to make restrictions on the estimation sample; see [SVY] **subpopulation estimation**.

- Use `subpop()` option rather than `if` or `in` qualifiers when performing subpopulation estimation

SVY `subpop`

- Let's follow along with Example 11 (*page 21)* from the SVY Documentation (www.stata.com/manuals14/svy.pdf)

Documentation Wrap-up

- Stata scales well into more statistically complex realms

- Unlike open-source languages, StataCorp provides and maintains excellent documentation for the user to learn Stata's specialized commands or check how statistical methods have been implemented

- With each new release, Stata adds functionality for more advanced types of analysis: the documentation is where to check out these developments