How do I use the Stata survey (svy) commands?

Authored by

stats writer

July 1, 2024

ARABPSYCHOLOGY.COM

RECOMMENDED CITATION

stats writer (2024). *How do I use the Stata survey (svy) commands?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=164803

The Stata survey (svy) commands are a set of tools that allow users to analyze complex survey data in a statistically sound manner. These commands take into account the specific sample design and weighting of the survey data, ensuring accurate results. To use the svy commands, users must first specify the survey design and sampling weights using the appropriate syntax. Then, they can use the svy commands to perform various statistical analyses, such as means, proportions, and regression models. The results produced by the svy commands take into account the complex survey design and are therefore more reliable and accurate. Overall, the svy commands are an essential tool for researchers and analysts working with survey data, as they provide a robust and rigorous approach to analyzing complex survey data.

How do I use the Stata survey (svy) commands? | Stata FAQ

Here is a tiny example showing how to use the survey commands in Stata. Consider the data file we call svysmall shown below.

use https://stats.idre.ucla.edu/stat/stata/faq/svysmall, clear

list

house eth wt y x1 x2 x3
1 1 .4 3 4 5 3
1 1 .9 9 4 5 6
2 1 1.2 9 8 7 3
2 1 1 8 7 4 2
2 1 1.1 8 7 6 3
3 2 .8 8 7 3 2

ARABPSYCHOLOGY.COM

4 2 .4 8 2 0 3

4 2 .7 8 2 5 3

In this tiny example, house is the household, eth is the ethnicity, and wt is the weighting for the person. You can use the svyset commands to tell Stata about these things and it remembers them. If you save the data file, Stata remembers them with the data file and you don't even need to enter them the next time you use the data file. Below, we tell Stata that the psu (primary sampling unit) is the household (house). Further, the sampling scheme included stratified sampling (strata) based on ethnicity (eth). Finally, the weighting variable (pweight) is called wt.

The way the svyset command is constructed is different between Stata version 7, 8 and 9. If you are not using Stata 9 or later, the syntax below will not work. Please see this page for examples. An example is given below. Notice that the PSU variable is given before the pweight, which is given in square brackets.

svyset house , strata(eth)

ARABPSYCHOLOGY.COM

Once Stata knows about the survey via the svyset commands, you can use the svy: prefix using syntax which is quite similar to the non-survey versions of the commands. For example, the svy: regress command below looks just like a regular regress command, but it uses the information you have provided about the survey design and does the computations taking those into consideration.

svy: regress y x1 x2 x3

The output is below, and it tells you the pweight, strata, and psu variables so you can confirm the right variables have been chosen.

Survey: Linear regression

Number of strata = 2 Number of obs = 8
Number of PSUs = 4 Population size = 6.5000001
Design df = 2
F(2, 1) = .
Prob > F = .
R-squared = 0.2216

ARABPSYCHOLOGY.COM

| Linearized
y | Coef. Std. Err. t P>|t|
-------------+--
x1 | .3321757 .294268 1.13 0.376 -.9339573 1.598309
x2 | -.138397 .2335074 -0.59 0.613 -1.143098 .8663043
x3 | .5504173 .3170068 1.74 0.225 -.8135527 1.914387
_cons | 5.050307 2.040247 2.48 0.132 -3.728167 13.82878

ARABPSYCHOLOGY.COM