Subject: Clarification on applying sampling weights in Stata using India NFHS-5
Posted by Withadel on Thu, 17 Apr 2025 18:18:22 GMT
View Forum Message <> Reply to Message

I am currently working with the India NFHS-5 Individual Recode (IR) dataset for a project analyzing maternal health service utilization. While going through the guide on sampling weights, I still have some doubts and would appreciate any clarification.

Context of my analysis:

Dataset: IAIR7DFL.DTA

Population of interest: women aged 1549 who had a live birth in the past 5 years

Main outcomes: number of ANC visits (categorical), institutional delivery (binary)

Software: Stata 17

I understand from the DHS documentation that v005 should be used as the weight variable, and that it needs to be divided by 1,000,000 before applying. However, I am still confused about how to properly specify this in the svyset command in Stata, and how clustering and stratification should be handled for India, which has a complex sampling structure.

I came across some posts and papers suggesting using v023 as the stratification variable instead of v022, especially for India. Could someone confirm which one is appropriate here?

Given that NFHS-5 has state-specific sampling frames and uses oversampling in some states/urban areas, is there any adjustment needed beyond this basic svyset structure? Or does DHS already account for that in v005?