Subject: Re: Weighting for the Pakistan Special 2019 DHS on Stata
Posted by Bridgette-DHS on Mon, 15 Aug 2022 20:20:18 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

Your svyset command is correct (except that we usually omit "vce(linear)"). Unfortunately, the raw data files from this survey do not include an identifier for the strata. In standard recode files, the stratum variable is hv022 or v022 (which is usually repeated with hv023 or v023). Here is a comment from Ruilin Ren, the DHS sampling expert who worked on this survey. He refers to hv022/v022, which are NOT in the raw data (qregion and qtype are in the data), but his description of how the strata are constructed is valid. He is referring to a "recode data file" that was used for the preparation of the report but otherwise is not available.

"There is a stratification variable in the recode data file, HV022/V022 with 14 codes. Actually the sample was designed for 8 provinces stratified by urban rural which totals the number of strata to 16. Punjab and Islamabad were combined together, KPK and FATA were combined together to form survey domains, GB and AJK were stand alone. But the Pakistan NOS selected the sample and did not provide an identifier for FATA, which reduced the number of indefinable provinces in the data file to be 7, KPK and FATA were all together and labeled as KPK, so we can only code 14 strata. That is why HV022 and V002 have 14 codes. This is not a problem for any purpose of use because FATA was indeed a very small part of KPK before. A key point for Pakistan surveys is the fact that they made GB and AJK stand alone, any attempt to combine different surveys together, or to analysis the full data together may introduce serious bias (bias toward the two small provinces GB and AJK) without a proper treatment of the sampling weights. This is an important advice for all data users using Pakistan survey data."