

I am working with the continuous DHS survey from Peru, 2004-2016. I read in the link below that the sampling design differed in 2004-2008 (sampling framework from 2000 census) vs. 2009-2014 (sampling frame from 2007 census).

However, (A) the document seems to state that Peru's 24 departments were used as strata in 2004-2008. But v023 (which supposedly gives design strata) gives 50 values. I think these 50 values relate to rural/urban within each of the 24 departments + the province of Callao... but if so, it's not clear if the sampling strata is actually represented by v023 as suggested by the data or by region (v024) as suggested by the document.

(B) That strata are described in that document as being much more complex in 2009-2014 (4 categories of rural/urban within each department), but v023 again numbers in 2009-2014. Perhaps v023 must be combined with some other variable in 2009-2014, to create design strata? Or perhaps v024 (region) is meant to be combined with some other variable to create design strata? But I don't see any variable defining 4 levels of urban-rural stratification. So again, I'm not sure how to create the strata here.

(C) I'm skeptical that the sampling frame is really the same in 2013 and 2014 as in 2009-2012, because sample size rises dramatically from ~10K (2009-2012) to ~45K (2013, 2014). Did the strata change too? No idea.

(D) The document below doesn't address the sampling frame for 2015 and 2016, but something must have changed because the sample size again increases dramatically in those last 2 years, and strata go from 1-50 in 2004-2014 to 1-3 in 2015 and 2016. Does v023, numbering 1-3, truly represent strata in these 2 years?

(E) While the document below explains how PSU were drawn in 2004-2008 (different PSU each year, always 1/5 of the 2000 PSU for each 2000 department, very clear), it does NOT clearly explain how PSU were drawn in subsequent years. PSU IDs do repeat in those later years. Looking at the PSU IDs over the years, and Table 1 in the document, I think PSU were drawn for 2009-2011 with some actual overlap across the years: half of the 2009 PSU were truly re-visited in 2010 and half of the 2010 PSU were truly re-visited in 2011. Then I think the PSU strategy changed in 2012-2014, but I'm not sure how... perhaps PSU were again overlapped between 2012 and 2013, and between 2013 and 2014. And then it looks like PSU are newly and uniquely assigned in 2015 and in 2016, each separately. But this is all a lot of guesswork on my part, because it's not clearly explained in the document, and I don't speak Spanish and so can't read the year-specific documentation.

Again, I am sure that some of the year-specific documentation addresses many of these questions, but they are in Spanish, and I don't speak Spanish. So, here are my questions about sampling strategy, weights, strata and PSU in the continuous Peru survey. My goal is to be able to properly weight the observations for pooled analysis.

(1) Do weights after 2008 reflect population change, such that they can be used without

adjustment? The document clearly states that this adjustment IS done in 2004-2008, but it doesn't discuss how population change after the 2007 census is accounted for (or not) by the weights in the 2009-2014 survey. And I have no idea how weights were made in 2015 and 2016.

(2) How do I make the sampling strata for each regime? It looks to me like sampling strata are probably consistent within 2004-2008, within 2009-2011, and within 2012-2014, and within 2015-2016. And probably those strata IDs can be made out of some combination of v023, v024, maybe v025, maybe some other variable for the more complex rural-urban strata periods. But I'm really not sure how to do it.

(3) In which periods are PSUs actually being re-visited? Am I right that PSU IDs are consistent (like, a repetition really means they are truly being revisited) within the 2004-2008, within 2009-2011, and within 2012-2014? But each of the next 2 years go to completely PSUs?

Thank you!
Leah

Link: <https://dhsprogram.com/pubs/pdf/OP8/OP8.pdf>
