

Following is a response from Senior DHS staff member, Tom Pullum:

Thanks for including DHS data in your courses! Unfortunately, weights are always required to produce unbiased estimates. You are right that much--maybe most--of the variation in weights is across strata, because large strata tend to be under-sampled and small strata tend to be over-sampled, to increase the efficiency (that is, to reduce the standard errors) of the national estimates. However, they also vary considerably across clusters in the same stratum. Variation in weights across clusters arises because (a) the observed number of households within a cluster typically differs from the number in the sampling frame and (b) an adjustment is made for nonresponse.

Within a cluster, the weights (hv005, v005, mv005, or d005) are the same for all cases (households, women, men, or respondents to the DV module). If you wanted to calculate, say, the mean number of children ever born (v201) or the proportion of women with no schooling (v106=0) for a specific cluster, you would not need to use weights. However, cluster-level analysis is almost never done. It IS done if you want to develop cluster-level covariates as contextual variables, but that's not a good example for an introduction to data analysis.

I suggest that you go ahead and introduce the students to data analysis just ignoring the weights and tell them that this is often done for exploratory work. Personally, I often start a new analysis without weights or any other adjustments, just to see whether there appears to be something going on in the data. And for data quality analysis it is normal to give equal weight to each observation, which is equivalent to ignoring the weights. Some researchers, mainly economists, prefer not to use weights at all, because it is more important for them to minimize the standard errors than to minimize the bias. For teaching purposes, I'd say it's ok to omit weights, but please do tell the students that their results are approximate/